TRANSLATIONAL TWISTS AND TURNS: SCIENCE AS A SOCIO-ECONOMIC ENDEAVOR

PROCEEDINGS OF STI 2013 BERLIN

18TH INTERNATIONAL CONFERENCE ON SCIENCE AND TECHNOLOGY INDICATORS

Institute for
Research Information
and Quality Assurance

European Network of
Indicators Designers

Published for the 18th International Conference on Science and Technology Indicators
“Translational twists and turns: Science as a socio-economic endeavor”

organized under the auspice of the European Network of Indicator Designers (ENID)
hosted by the Institute for Research Information and Quality Assurance (iFQ)
Schützenstraße 6a
10117 Berlin
Germany
Phone +49(0)30-206 41 77-0
Fax +49(0)30-206 41 77-99
E-Mail info@forschungsinfo.de
Web www.forschungsinfo.de (German)
www.research-information.de (English)

ISSN 1864-2799

September 2013

Translational twists and turns: Science as a socio-economic endeavor

Proceedings of STI 2013 Berlin

18th International Conference on Science and Technology Indicators
Berlin, Germany | September 4 – 6, 2013

Edited by Sybille Hinze / André Lottmann

Index

ENID Conference Committee

Anthony F. J. van Raan	Centre for Science and Technology Studies (CWTS), Leiden University / President ENID
Sybille Hinze	Institute for Research Information and Quality Assurance (iFQ), Berlin
Benedetto Lepori	University of Lugano
Emanuela Reale	Institute for Economic Research on Firms and Growth (CERIS-CNR), Rome
Robert Tijssen	Centre for Science and Technology Studies (CWTS), Leiden University

ENID Scientific Committee

Jonathan Adams / Per Ahlgren / Judit Bar-Ilan / Aparna Basu / María Bordons / Benjamin Bowman / Kevin W. Boyack / Linda Butler / Rodrigo Costas / Rickard Danell / Hans-Dieter Daniel / Koenraad Debackere / Gemma Derrick / Ying Ding / Tim Engels / Rainer Frietsch / Stephan Gauch / Wolfgang Glänzel / Jochen Gläser / Isabel Gomez / Juan Gorraiz / Stevan Harnad / Diana Hicks / Stefan Hornbostel / Masatsura Igami / Margriet Jansz / Ben Jongbloed / Koen Jonkers / Sylvan Katz / Richard Klavans / Vincent Larivière / Grit Laudel / Loet Leydesdorff / Liming Liang / Marc Luwel / Valentina Markusova / Göran Melin / Johann Mouton / Ulle Must / Anton Nederhof / Ed Noyons / Bluma C. Peritz / Olle Persson / Ismael Rafols / Ed Rinia / Ronald Rousseau / Elias Sanz-Casado / Christian Schloegl / Ulrich Schmoch / Jesper Schneider / Anne Sigogneau / Giorgio Sirilli / Daniel Sirtes / Gunnar Sivertsen / Stig Slipersæter / Henry Small / Mike Thelwall / Bart Thijs / Robert Tijssen / Peter van den Besselaar / Nees Jan Van Eck / Thed Van Leeuwen / Karen Vandevelde / Peter Vinkler / John Walsh / Ludo Waltman / Jos Winnink / Matthias Winterhager / Paul Wouters

Supporters

Stifterverband für die Deutsche Wissenschaft
Elsevier
Thomson Reuters

Preface

In recent years the demand for and the use of science, technology and innovation indicators increased significantly. On the one hand science policy is asking for 'objective' data to assess the performance, relative position, and interaction of the various players in the STI system. Among other things, this rise in demand can be explained by the widespread move towards new modes of governance which increasingly can be characterized as 'governing by competition' and new public management. On the other hand science itself is calling for a better understanding of its knowledge production, dissemination and transfer processes, cumulating in innovations for economic welfare and wider societal usage. Both sides serve as continuous driving forces for the development of quantitative studies of science, technology, and innovation.

New demands also emerged as a consequence of the growing complexity of science and innovation systems. This concerns, for instance, interactions at the regional, national and international level but also changes in the scientific communication due to the increasing relevance of new communication channels. Consequently, new data sources open up new analytical options and new performance indicators emerge addressing these aspects (e. g. collaboration indicators, web indicators, indicators on human resources, career advancement and mobility). Furthermore, the continuously growing computer power and memory enable increasingly extensive and large-scale data analysis. Existing indicators are put to test whether they are still adequate and address the needs of science policy as much as science studies.

The International Science and Technology Indicators Conference, which is organized under the auspice of ENID, the European Network of Indicator Designers (www.enid-europe.org), provides a European and worldwide forum for presenting and discussing advances in constructing, using, and interpreting science and technology indicators. The requirements for data generation, indicator production, also for specific domains, their design, methodology, experimental development, and application will be discussed. The STI conference series is devoted to bring together researchers, STI producers and users, as well as other stakeholders. The STI conference series thus contributes to gain a better understanding with regard to STI indicators applied in different contexts which range from understanding institutional structures, developmental processes, and contexts of science itself to their use as analytical tools in knowledge management and science policy decision making. It does so for exactly 25 years now since the first STI conference was held in Leiden in 1988.

Science is increasingly expected to contribute to solve the so called 'Grand Societal Challenges'. The expectation extends to indicators: such contributions should be traceable and its effects and impacts measured. These expectations pose challenges to the STI community which we address with the overarching theme of this year's conference: "Translational twists and turns: Science as socio-economic endeavor". In order to address this topic we invited contributions on eight key topics:

- Measuring and assessing benefits of science (including scientific outputs, innovation, social and environmental benefits)
- Knowledge transfer, knowledge exchange and specific issues relating to translational research
- Science push or pull factors for societal change
- Governance by competition: modes, instruments, effects and impacts of competition in research funding
- Impact of performance-based funding systems
- Production and reproduction of hierarchy: pros and cons of university rankings
- The geography of science and higher education systems: globalization, regionalization and localization processes
- Cross-sectorial research collaborations: fertilization or impediment?

This year's conference is organized by the Institute for Research Information and Quality Assurance (iFQ). It is held at the Berlin-Brandenburg Academy of Sciences and Humanities in Berlin on September 4–6, 2013.

In total we received 116 submissions. Each paper was reviewed by at least two members of the scientific committee. The program committee took the final decision about paper acceptance based on these votes. 57 papers were accepted for oral presentations. These as well as 28 posters are contained in the present book of proceedings. The contributions are organized in alphabetical order by surname of the first author.

We would like to sincerely thank all the authors for their submissions and all the members of the scientific committee for their participation in the review process. Furthermore we want to thank our partners who provide financial support.

Thank you all for being here with us and participating in STI 2013. We are looking forward to a successful and inspiring conference.

Anthony van Raan / Sybille Hinze / Benedetto Lepori / Emanuela Reale / Robert Tijssen

[Programme Committee]

I. Full Papers

Relationships between policy, funding and academic performance – Examination of a Danish success story

Kaare Aagaard ka@cfa.au.dk / The Danish Centre for Studies in Research and Research Policy / Aarhus University, Bartholins Allé 8, 8000 Aarhus C (Denmark)

Jesper W. Schneider jws@cfa.au.dk / The Danish Centre for Studies in Research and Research Policy / Aarhus University, Bartholins Allé 8, 8000 Aarhus C (Denmark)

Abstract

The relationship between research policy and academic performance has high policy relevance. The knowledge of how different systemic factors affect research performance is however limited and inconclusive. Although a number of comparative studies have been conducted, no clear patterns have emerged. One possible reason is that these questions are broad and highly complex with a number of intervening factors and considerable time-lags. Long timeframes and the inclusion of extensive quantitative and qualitative data that goes beyond simple key figures are therefore required when these questions are studied.

In this study a one country case study is accordingly presented where the relationship between research policy and research performance is examined in detail over a timespan of more than three decades within a Danish context. This approach reveals that Denmark which today stands out as one of the top-performing research nations of the world experienced a turning point in performance in the early 1990's and that this turning point correlates with a number of systemic policy changes which have been maintained at least up until 2006. The results raise the question of how the extensive Danish research policy reforms of the last 5–7 years will affect this strength position.

Introduction

Research policy and research funding have gone through substantial transformations in most OECD countries since the 1970s. The development has been covered with focus on the content of different periods, paradigms and policy agendas (Elzinga & Jamieson, 1995; Guston, 2000), and the nature and consequences of the related changes in research funding mechanisms (Geuna, 2001; Lepori et al, 2007; Auranen & Nieminen, 2010; Benner & Öquist, 2012). Central questions have been how research policy changes have affected research performance.

While there is some evidence of the short-term usefulness of incentives and competition, long term country-specific information on research performance in relation to the scope, scale and composition of public research funding seems to be largely missing (Geuna and Martin, 2003;

Liefner, 2003). In one of the few available studies Auranen & Nieminen (2010) shows that there are significant differences in the competitiveness of different policy- and funding systems, but no straightforward connection between financial incentives and the efficiency of university systems. There are accordingly no clear crosscutting conclusions; except maybe that we in all cases are dealing with highly complex questions and multiple possible answers and correlations. In other words it appears to be a complex mix of factors which together create the conditions for performance in national systems (Auranen & Nieminen 2010, Benner & Sandstrøm 2000). Furthermore, the literature emphasizes that research activity is related to a number of other factors than funding and policy objectives (Jongbloed 2007).

To go beyond these limitations in previous studies it is therefore necessary to conduct more detailed and thorough case studies to highlight the complex relationships. We do this by zooming in on just one strong performer and examine the development over a long time period. In terms of academic output measured by bibliometric indicators Denmark currently stands out among the top performing countries in the world, but the foundation for this performance has been the target of surprisingly little research so far. The Danish case is not least interesting because a rather significant decline in performance from 1981 to 1990 was replaced with a steadily increasing trajectory all the way to 2010. Due to newly created time-series covering different aspects of funding and research organisation supplemented with solid qualitative descriptions and bibliometric performance indicators it is now possible to explore this development in a longitudinal study within a Danish context. The research question of the paper is accordingly:

How can we explain the development in Danish research performance in the period 1981 to 2010?

The main question is, in other words, if it is possible to detect any systematic correlation related to the different links between research policy, funding decisions and academic performance when a 30-year time-span is applied to the analysis.

Theoretical frame

The study takes its departure from a theoretical framework combining recent contributions to Historical Institutionalism with HE-specific work on the effects of changing authority relationships within the research policy system.

Within historical institutionalism the question of change is one of the areas where the biggest recent advancements have been achieved (Streeck & Thelen 2005; Mahoney & Thelen 2009). This development has in recent years led to a convincing and theoretically innovative contribution to the understanding of evolutionary change. One of the strengths of this approach in relation to this study is that it points our attention to possible critical junctures and formative moments in the Danish research policy, while also maintaining the attention on evolutionary changes.

There is however a need to bring in additional field specific theories to steer the analytical attention towards the right factors in the attempt to explain change. An interesting field specific theory can here be found in a number of recent contributions to the HE literature focusing at changing authority relationships as a framework for analysing science policy transformations (Whitley, Gläser & Engwall 2010; Whitley 2011). Much of this recent research has focused on how intensified competition for resources and increasing demands for relevance and accountability have affected patterns of authority relations between academics and various stakeholders (Louvel 2010). The understanding of these changing authority relationships is crucial in the attempt to explain the Danish development and cannot be captured in quantitative time series alone.

Data and methods

A thorough analysis of the combined development in research policy, research funding and academic performance has so far been lacking in a Danish context, but these correlations can now be explored due to newly created time series of public research funding and academic performance covering more than three decades. These quantitative data are supplemented with detailed qualitative descriptions of changes in research policy within the same period which enrich the interpretation of the time series and allow for a longer and more detailed analysis on this subject than – to our knowledge – has been conducted in any other country.

The aim is to show that this mixed methods approach makes it possible, not only to reconstruct a detailed picture of three decades of evolution of research policy in a national context, but also to relate these changes to academic performance during the whole period – although the links by no means are straightforward.

To explain the development in performance we look at total public R&D funding, balance between floor funding and competitive funding, balance between academic orientation and external orientation in the funding system, organization of the sector, level of state-steering, leadership structures and ph.d. education/uptake. No clear causal explanations can be given based on this approach, but we can nevertheless present a set of empirically grounded hypotheses that can be tested in other countries and which also will be used in our coming work where we aim to disaggregate the results of this study. In a forthcoming study we accordingly aim to scrutinise this study of Danish research performance by breaking it down from the macro level to the meso levels of fields and institutions in order to examine in what way these units of analyses contribute to the "Danish success story".

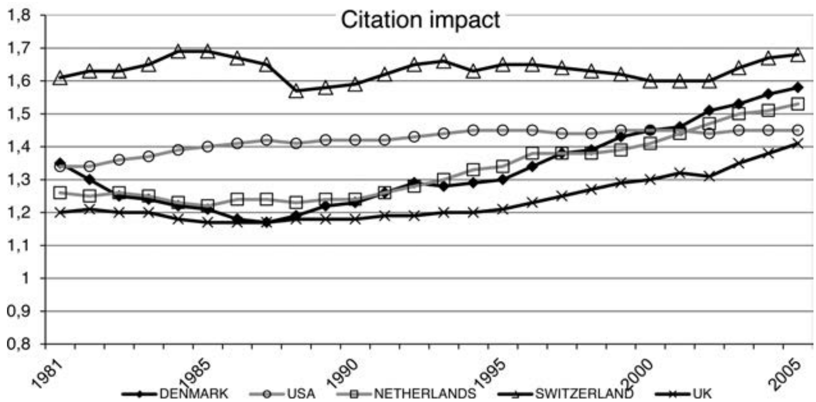
A main foundation for the present analysis is scientometric performance at the national level. In public discourse indicators of citation impact are often uncritically translated into indicators of research quality, signifying importance, and analyses at the national level are often presented as if they covered all fields of research in a country. Obviously, this is unwarranted. It is important to emphasize the restrictions of scientometric indicators, their constrained coverage, as well as

their validity and reliability (e.g., Gläser & Laudel, 2007). The indicators behind the "Danish success story" are at best a proxy for the performance of Danish research. It is based on journal article literature indexed in Thompson Reuters' Web of Science database (formerly ISI) and good field coverage herein is basically restricted to science and medical fields (Moed, 2005). The validity of citation indicators rests on the assumption that citations, at the meso and macro levels reflect short-term impact upon the scientific communication system and that this impact can be conceived of as a proxy for research performance (van Raan, 2004). The perceived direct relation between citations and research quality has never been firmly established and can therefore not be inferred. Nevertheless, citation impact matters; it is highly implausible to assume that citations are granted arbitrarily, i.e. independent of an article's actual impact (Nedehof & van Raan, 1987). We therefore find it valid to assume that citation impact at an aggregated level say something about the immediate perception and visibility of a country's publications in the scientific community, as delineated by Web of Science.

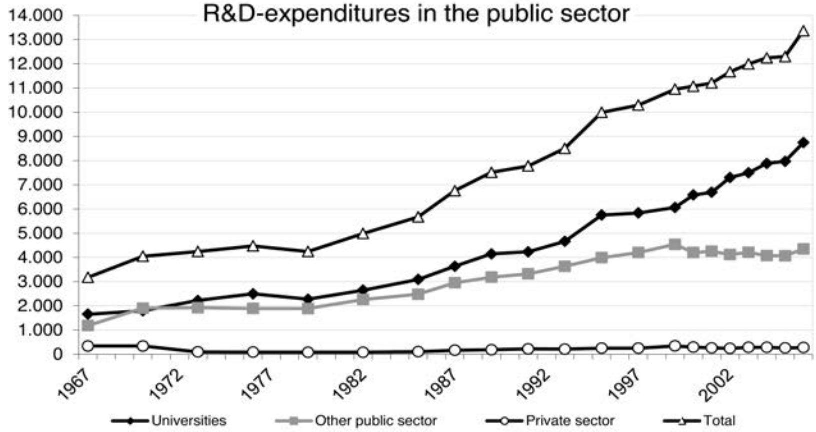
Results

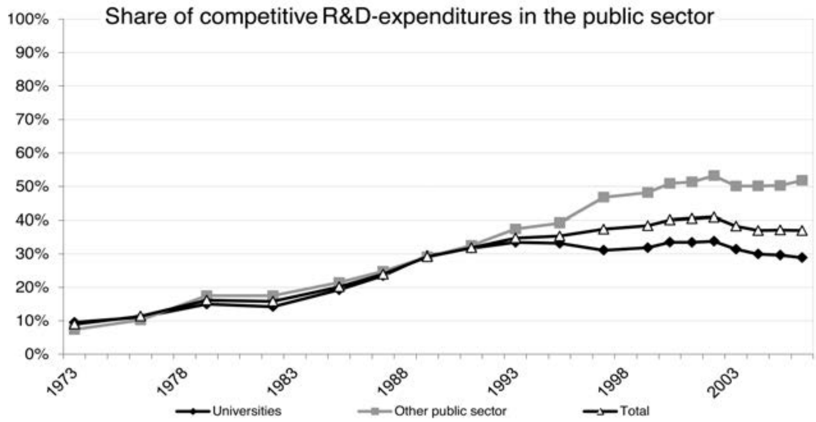
In the following selected results are briefly presented. Due to the limited space only a few of the timeseries are shown.

Figure 1. Citation impact, R&D expenditures, Share of funding in competition



... Continuation Figure 1





Sources: National Science Indicators 2010. Thompson Reuters ©; Lauridsen, P., Graversen, E. 2013.

Several studies have documented the initial decline and subsequent rise in Danish research performance shown above (e.g., Nørretranders, 1990; Schneider et al., 2011; Karlsson & Persson, 2012). Different citation indicators have been used to measure performance. Early studies focused on average relative citation impact based on whole publication counts. More recent studies, use fractional counting of publications and focus on the share of highly cited publications for a country. Given the variety of measures and different techniques applied in these studies, we find the evidence for the general claim of continuous rising impact of Danish research very reliable. It is apparent from these analyses that Danish science and medical research articles indexed in the Web of Science from 1990 onwards have experienced a constant rise in average impact, as well as continuous rise in the number of highly cited articles. Especially, the latter fact is interesting, as the share of highly cited articles to some extent is considered an indication of "excellence" (e.g., Zitt, Ramanana-Rahary & Bassecoulard, 2005).

Policy changes

This development in performance corresponds to a number of policy developments which are briefly highlighted in the following based on Aagaard (2011) divided into four decades:

The 1970s: National research policy objectives were largely missing throughout the 1970s. Decisions affecting Danish research policy up until the 1980s were side-effects of decisions taken in policy areas such as education, industry and culture. Funding of research was almost fully dominated by floor funding with very few strings attached. Funding to the universities was in general allocated based on input factors (student numbers as main criteria). The research council system was marginal, in general very academically oriented and closely linked to the universities. Concentration of research resources was not an issue; rather the opposite: the period as a whole was characterised by a steady growth in number of institutions with regard to both universities and public research institutions. In addition, the organisation of the public research sector was characterised by a strong functional division between academic research at the universities and applied research at the public research institutes with the two sectors constituting app. half each. The management structure at the Universities changed in 1970 from a classical meritocratic model to a democratic model with strong student and administrative representation. Critics argued that a consequence was a shift from quality to equality in the internal allocation of research funding. In general, there was no use of research evaluations; neither at national nor institutional level.

The 1980s: By the early 1980s the public research effort started to be perceived as lacking integration and cooperation with the outside world in general and the private sector in particular. In addition university research was seen to suffer from a lack of competition and a lack of mobility. This was perceived as a serious dysfunction of the system considering the emerging policy belief that renewed industrial growth should be based on key "generic" technologies such as information technology, biotechnology and materials science. The link between student numbers and floor funding was abandoned in 1982 and an activity-based funding system was introduced for education, but no alternative was presented with regard to research funding. This development led to a strong growth in earmarked program funding of university research; especially from 1985 and forward. A major part of the program funding was placed in special comities outside of the existing academically oriented research council structure. With regard to concentration of research resources it was a main objective to strengthen the best research environments within selected areas, but the impression was that the institutions and research councils reduced this selectivity by reallocating the remaining research funds in the system. Recruitment to the system through new ph.d's was in general limited and unformalized.

The 1990s: The 1990s saw a reorientation back towards academic research after the strong growth of strategic research in the previous decade. This reorientation, however, came with new instruments, new funding channels and stronger internal management at the universities. 1991 saw the establishment of the Danish National Research Foundation, which was created to support new centers of excellence solely based at academic quality criteria. Furthermore, it was by the early

1990s agreed that the balance between floor funding and external funding should be maintained at approximately 60/40, which in reality stopped the relative growth in program funding. The formulation of research programs changed from the large and long term allocations of the late 1980s towards a proliferation of small short term programs. Attempts were also made to reform the research council system, to formulate a national research strategy and to develop and output-based allocation system of floor funding. In reality, only limited changes were carried through. Finally socalled development contracts for the universities were introduced, but with no funds directly attached. The discussion of concentration of resources continued, but there was still no reduction in number of research units. However, the universities started to receive a larger part of the total public research funds at the expense of the institute sector. Ph.d education was highly prioritized and formalized from the beginning of this decade.

The 2000s: In 2001 a new right wing government took office and started a sweeping reform-process including a far reaching management reform of the universities, a new and more strategic oriented research council system and a new performance-based floor funding model to mention just a few. The University Act from 2003 introduced a board with an external majority as the superior authority of universities and prescribed employed leaders instead of elected. The objective was to sharpen up the profiles of individual institutions and to increase collaboration between the actors of the research and innovation system – the latter exemplified by the new claims put forward for universities to formulate goals and strategies for cooperation with trade and business and by the introduction of external members in the boards. The Act emphasised that the universities' new management should make strategic selections of research areas and give high priority to these areas. At the same time it was argued, that there was too little competition for research funding and that the funding was spread too thinly. This led to changes towards competition and output factors of the research council system as well as the allocation system of university floor funding. In 2006 the Government furthermore initiated a far reaching merger process which reduced the number of universities from twelve to eight and closed down the majority of the public research institute sector. There was, in other words, a large concentration of resources within few institutions, and also a clear break with the former division of academic research and more applied research. Finally, the 2000's saw a number of new funding channels: besides the division of the research council structure between a bottom up – and a strategic research council, a Council of Technology and Innovation and a new High Technology Fund was created. Once again ph.d.-education was highly prioritized with a 100% increase in uptake in the period 2004–2010.

Table 1. Key figures and characteristics

	1970–1979	1980–1989	1990–1999	2000–2010
Public R&D as share of GDP	0.54–0.45	0.45–0.67	0.67–0.76	0.75–1.00
Floor funding share (%)	91 to 84	84 to 71	68 to 62	60 to 56
External orientation	Very low	High	Medium	High
University share of total public R&D funding	50	50	60	85
Ph.d uptake	Medium/ unformalized	Low/ unformalized	High/ formalized	High/ formalized
Leadership	From meritocratic to democratic (1970)	Democratic	Democratic/ Hierarchical (1993)	Hierarchical (2003)
Bibliomtric performance	(High)	Decreasing	Increasing	Increasing

Conclusion and discussion

The data presented above points to a number of possible explanations of the Danish development in performance. The 1980's which saw a considerable decline in performance were characterized by instability in funding, a strong turn towards competition and external orientation, a weak institutional leadership, a fragmentation in the national institution structure with a very large institute sector and a weak and unformalized ph.d education and uptake. In changing authority relationship terms these developments can be characterized as shifts from academically oriented groups towards the state and external actors. All these factors were changed somewhat in the early 1990's where there was a reorientation back towards academic research, where the balance between floor funding and competitive funding found a stable level, where the institutional leadership was strengthened, where the universities started to receive a larger part of the total public R&D spending and where the ph.d.-education was strengthened and formalized. These policy changes corresponds with the reversal of the decreasing trend in performance in the 1990's and appear to have created the foundation for the steady increase that we have witnessed up until 2010. These changes can, however, not be characterized as elements in a large master plan, but rather as a number of decisions that pointed in the same direction – followed by 10–15 years of relative stability in funding and frame conditions.

The big question in Danish research policy today is however, how the recent major changes will affect academic performance in the coming years. Once again Danish research policy has seen a reorientation back towards a situation with more authority to external interests, more weight at competitive funding, and an organisation of the sector with larger, more complex, hybrid and more hierarchical institutions. The period has however also seen a strong growth in total research funding, a renewed focus on ph.d.-education and the introduction of an academically oriented performance based floor funding model. These developments are probably not yet visible in the performance indicators as most of the changes have taken place after 2006, but it will be highly interesting to see how they will affect academic performance.

References

Aagaard, K. (2011). *Kampen om basismidlerne*. Ph.d. dissertation. The Danish Centre for studies in research and research policy.

Aagaard, K., Mejlgaard, N. (ed.). (2012). *Dansk Forskningspolitik efter årtusindskiftet*. Aarhus Universitetsforlag.

Auranen, O. & Nieminen, N. (2010). University research funding and publication performance–An international comparison. *Research Policy* 39 (2010) 822–834.

Benner, M., Sandström, U. (2000). Institutionalizing the triple helix: research funding and norms in the academic system. *Research Policy* 29 (2), 291–301.

Benner, M., Öquist G. (2012). *Fostering breakthrough research: a comparative study*. Kungl. Vitenskabsakademien.

Geuna, A. (2001). The changing rationale for European university research funding: are there negative unintended consequences? *Journal of Economic Issues* XXXV, 607–632.

Geuna, A., Martin, B.R. (2003). University research evaluation and funding: an international comparison. *Minerva* 41 (4), 277–304.

Gläser, J. & Laudel, G. (2007). The social construction of bibliometric evaluations. In: R. Whitley & J. Gläser (eds.), *The changing governance of the sciences*, 101–123.

Jongbloed, B. (2007). Steering the Dutch academic research enterprise: Universities' responses to project funding and performance monitoring. A Paper Presented at the *20th CHER Conference, Dublin, 30 August–1 September*.

Karlsson, S. & Persson, O. (2012). *The Swedish production of highly cited papers*. The Swedish Research Council's report series 5:2012: http://www.cm.se/webbshop_vr/pdfer/2012_05L.pdf

Lauridsen, P., Graversen, E. (2013). Forskning og udviklingsarbejde i den offentlige sektor 1967–2006. Dansk Center for Forskningsanalyse. Aarhus Universitet.

Lepori, B., P. van den Besselaar, M. Dinges, B. Potì, E. Reale, S. Slipersæter, J. Theves and B. van der Meulen (2007). Comparing the evolution of national research policies: what patterns of change? *Science and Public Policy*, 34(6), July 2007, 372–388.

Liefner, I. (2003). Funding, resource allocation, and performance in higher education systems. *Higher Education* 46 (4), 469–489.

Louvel, S. (2010). Changing authority relations within French academic research units since the 1960s. In . R. Whitley, J. Gläser and L. Engwall (ed) *Reconfiguring Knowledge Production: Changing Authority Relations in the Sciences and Their Consequences for Intellectual Innovation*. Oxford University Press: 184–210.

Mahoney, J. & Thelen, K. (2009). *Explaining institutional change: ambiguity, agency, and power*. Cambridge University Press.

Moed, H. (2005). *Citation analysis in research evaluation*. Dordrecht: Springer.

Nederhof, A.J. & van Raan, A.F.J. (1987). Citation theory and the Ortega Hypothesis, *Scientometrics*, 12:325–328.

Nørretranders, T. (1990). *Dansk dynamit*. Forskningspolitik nr. 8.

Schneider, J.W. et al. (2010). *Bibliometric Research Performance Indicators for the Nordic Countries*, Norden, NordForsk, Oslo: www.nordforsk.org/_img/bibliometri_2.pdf

Streeck, W., Thelen, K. (2005). *Beyond Continuity: Institutional Change in Advanced Political Economies*, Oxford: Oxford University Press.

van Raan, A.F.J. (2004). Measuring science. In: H.F. Moed, W. Glänzel og U. Schmoch (eds.), *Handbook of quantitative science and technology research*, 19–50.

Whitley, R. (2011). Changing governance and authority relations in the public sciences. *Minerva* (2011) 49:359–385

Whitley, R., Gläser, J., Engwall, L. (2010). *Reconfiguring Knowledge Production: Changing Authority Relations in the Sciences and Their Consequences for Intellectual Innovation*. Oxford University Press.

Zitt, M., Ramanana-Rahary, S., Bassecoulard, E. (2005). Relativity of citation performance and excellence measures: from cross-field to cross-scale effects of field-Normalisation. *Scientometrics*, 63 (2), 373–401.

Size of web domains and interlinking behavior of higher education institutions in Europe

Isidro F. Aguillo isidro.aguillo@csic.es / Cybermetrics Lab, IPP – CSIC, Albasanz, 26–28, Madrid, 28037 (Spain)

Benedetto Lepori benedetto.lepori@usi.ch / Centre for Organisational Research (CORe), Universitá della Svizzera Italiana, Via Lambertenghi, 10, Lugano, 6900 (Switzerland)

Marco Seeber marco.seeber@usi.ch / Centre for Organisational Research (CORe), Universitá della Svizzera Italiana, Via Lambertenghi, 10, Lugano, 6900 (Switzerland)

Abstract

During the last years two different approaches have been used to describe the interlinking patterns inside the academic web space, one using document properties in a similar way to the citation analysis (webometrics) and an alternative one that takes into account organizational properties of the institutions (social network analysis). A data set including organizational data from the EUMIDA project, bibliometrics information from the Scimago Institutions Ranking and a matrix of web links from the MajesticSEO database was compiled for a group of 400 top universities from 20 European countries. For this university group evidence were checked for addressing which properties were more relevant. The results show that both organizational and document properties matter, with some signs indicating that the former might be more important for predicting web links. However, the inclusion of the web domain size improves the fit.

Introduction

Link analysis has been traditionally performed by information scientists under the disciplinary name of cybermetrics or webometrics (Aguillo et al., 2006; Thelwall, 2001). But there is also a different approach to the study of web links coming from social network analysis that is focused on matching social relationships with interlinking patterns. Most of those papers use the academic web space (Thelwall & Wilkinson, 2004; Vaughan & Thelwall, 2005) for answering policy relevant questions as both old and new datasets are becoming available for further analysis.

The aim of this contribution is to analyse to which extent the size of web domains of Higher Education Institutions (HEIs), measured by the number of webpages, influences the number of links they send and receive.

There are two possible interpretations of the web links. In the first one, the links are connecting documents, like in the case of bibliographic citations (Cronin, 2001), where the motivations of web authors are largely determined by the characteristics and contents of the webpages, including

personal or institutional authorship (deep linking). A second alternative option is that web links are markers of underlying relationships between organizations, as related to research and educational cooperation, access to resources, reputation, etc. (institutional linking).

Distinguishing between these two models is critical to interpret interlinking data and draws to very different predictions concerning their statistical properties. Document networks are expected to display much skewed distributions such as power laws (Adamic & Huberman, 2000), to be strongly asymmetric and reputation based and to be influenced by web domains content and structure. Organizational network are expected to display different patterns, to be less-skewed (lognormal distribution), closely related to resources and reputation and more sensible to social proximity, e.g. belonging to the same country (Seeber et al., 2012; Daraganova et al., 2012).

In this paper, we empirically address this issue by combining data on size of web domains (number of webpages), on interlinking (counts of web links between two Higher Education Institutions; HEIs) and organizational characteristics of HEIs (size, international reputation, geographical distance) for a sample of 400 HEIs in 20 European countries.

We thus test to which extent the size of web domains influences the number of web links sent/received between two HEIs when we control for organizational characteristics: How is the web domain size related to organizational characteristics (size, etc.). Has the size of the web domain an impact on number of interlinking? Are the web domain size and organizational variables effects independent, meaning that coefficients of size, reputation, etc are not influenced significantly by web domain size. To which extent is a model including web domain size better in predicting counts of web links?

Methodology

Datasets

Three different data sources were used in this study: EUMIDA project (2009) provided organizational data, including size of the institution or reputation (Lepori and Bonaccorsi 2013); Scimago Institutions ranking (http://www.scimagoir.com/) supplied the bibliometric information and the Cybermetrics Lab collected the web size and interlinking data by processing information from MajesticSEO (http://www.majesticseo.com/, February2013).

A set of 400 European universities from 20 different countries were built by the overlap of the three different data sources used in this analysis. The country distribution is mostly proportional to its performance in Scimago Institutions Ranking (2011 edition), except for several Polish universities that were excluded because their shared web domains make unreliable the web data obtained.

Interlinking data and data on size of web domains

Traditionally web data for very large populations has been extracted from the general search engines like Google, Yahoo or Bing. The huge size and global coverage of their databases made them especially suitable for webometric studies, but unfortunately their capabilities and operators for link analysis were very limited. Additional problems related to the opaque and irregular behaviour make advisable to search for alternative sources.

Since 2010 the Cybermetrics Lab has been evaluating three different link data commercial providers: the US SeoMoz (http://www.opensiteexplorer.org/) the British MajesticSEO (http://www.majesticseo.com/) and the Ukrainian ahrefs (http://ahrefs.com/) services.

The API provided by MajesticSEO was especially suited for obtaining both web size (number of different URLs indexed) and linking information among pages or complete web domains. Two main problems persisted in the data collection using MajesticSEO: The first one is related to duplicate domains as for 26 universities there were not a unique but two main web domains. The second problem is that the system provides erroneous results for universities with shared domains, a common situation for several Polish universities that uses city-related domains (for example Wrocław University of Technology / Politechnika Wrocławska uses the city domain: pwr.wroc.pl). In the first case the procedure were to combine (addition) the link numbers to both domains, but as it can be assumed a large overlap between the contents of both domains the web size was obtained from the maximum value between the two ones. In the second case, all the Polish universities in those circumstances were excluded from the analysis. However the finally included universities from this country are generally the most prestigious of productive ones.

Organizational data

For the 400 HEIs the following data were available: Size measured through the number of academic staff, kilometric distance between two HEIs, a dummy variables if two HEIs belong to the same country, a measure of subject similarity of two HEIs based on overlap in the distribution of students in educational fields and a dummy variable for mission similarity, which is 1 if both HEIs are doctorate awarding.

We use as a measure of international reputation the brute force indicator calculated as the product of total number of publications of an HEIs with their average impact factor, normalized by the number of academic staff (van Raan 2007); this indicator builds on the insight that the international visibility of a HEI is related both to quality and volume of output. Data are derived from the Scimago Institutions Rankings for the year 2011 (http://www.scimagoir.com/).

In a previous paper, it has been shown that these factors strongly impact on the presence of a web link between two HEIs, as well as on the number of links (Seeber et al. 2012).

Statistical methods

We first provide descriptive statistics on the size of web domains and of its relationships with HEI size and other organizational variables to test whether they are strongly correlated, to check for outliers and to investigate whether the distribution of web domain size is more skewed than the one of size, as it would be predicted by theory (Daraganova et al. 2012).

Second, we compare distributions of web domain size, organizational size and total degree of web links to test whether their properties are different and whether degree distribution is more similar to the one of size (in most cases lognormal) or of web size.

Third, we perform regressions between links count (dependent variable) on the one side, organizational characteristics and size of web domains on the other side. Since we deal with count data, we use a negative binomial regression which includes a parameter to model overdispersion as compared to a standard Poisson regression (Cameron and Trivedi 1998). This type of models is robust against non-normality of distributions and the presence of outliers and we successfully used to model web links (Seeber et al. 2012).

We compare the following models concerning their ability to predict observed counts of web link in our sample: model 1 is the benchmark model including all organizational variables; in model 2, we replace measures of HEI size with the one of their web domains, to test whether it better predicts web links; finally, in model 3, we include both size and size of the web domain (normalized by size) to test to which extent the level of fits improves from the benchmark model.

Model comparisons are made through standard fit statistics (loglikelihood, deviance, AIC); in a further step, we will more thoroughly compare the models against the ability of predicting the observed values. It is already know from previous works that the baseline model is quite efficient in this respect (Seeber et al. 2012).

Preliminary results

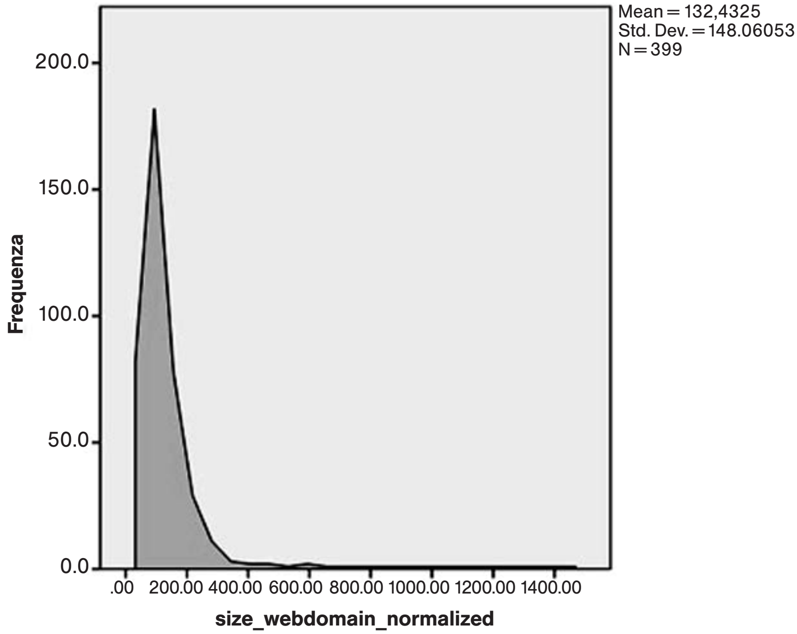
We provide in this section preliminary results not including data on reputation, which will be integrated in a later stage in the models.

Descriptive Statistics

The size of web domains displays as expected a highly skewed distribution with a minimum value of less than 4000 webpages and a maximum of around 2.5 mio; the average size is 178’000 pages with a standard deviation of 213’000 and a skewness of 5.465. Correlation with size is relatively high and significant (.418**), even if this result might be strongly affected by outliers.

The distribution of web domain size normalized by academic staff displays somewhat different patterns (Figure 1).

Figure 1. Size of the web domain normalized by HEI size (academic staff)



As a matter of fact, we observe the superposition of a normal distribution for the values below 200 pages per unit of staff (with a cut at this threshold the distribution passes the Kolmogorov-Smirnov test for normality) following by a long queue of HEIs for which the size of the web domain is much larger.

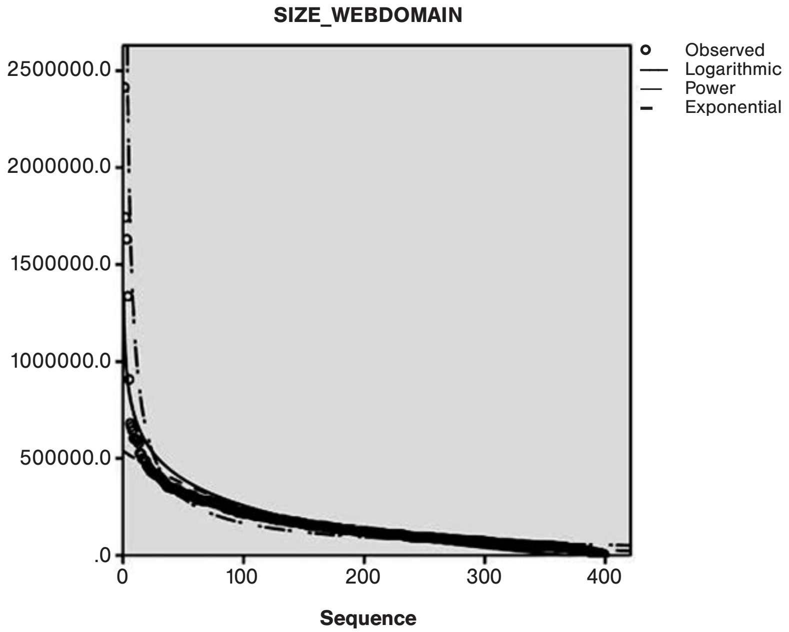
This group of HEIs with web domain size normalized above 200, which includes 46 HEIs, has an average size very similar to the full sample and does not show specific patterns in terms of country, of international reputation and research orientation. The reasons for these high values need to be further investigated, but our hypothesis is that it is related more to characteristics of the websites, like including personal pages or document repositories. A more in-depth analysis of web domains by document type will be performed to address this question.

In the full sample, the web domain size normalized by academic staff is slightly negatively correlated with academic staff (-.140), while differences between countries in the medians are relatively limited (considering that some countries have few cases). It will be further investigated to which extent research intensity and international reputation has an impact on the size of the website.

Web domain size and total degree

Data show that organizational size and degree display a lognormal distribution (Kolmogorov-Smirnov statistics is .488 for size and 1.403 for degree), whereas web domain size is approximated by an exponential distribution, which fits better the queue of HEIs with very large websites (Figure 2).

Figure 2. Distribution of web domain size



While it is well-known that, following Gibrat's law, organizational size tend to follow a lognormal distribution (Ijiri & Simon, 1964), these results support the assumption that, while mechanisms generating webpages are typical of document networks and thus display a more skewed distribution, those generating web links relate to organizational relationships, leading to a distribution of degree which has similar properties than the one of size.

Predicting counts of web links

Table 1 displays the results of the negative binomial regression with the three models selected, as well as with the null model as a reference.

Table 1. Results of the binomial regression

	Null model		Size model		Websize model		Combined model		
	Est.Std.	Error	Est.Std.	Error	Est.Std.	Error	Est.	Std. Error	Beta
(Intercept)	2.4776	0.0067***	2.606	7.021e-02***	1.645e+00	7.085e-02***	1.249e+00	7.015e-02***	
Size_receiver			6.576e-04	5.061e-06***			7.034e-04	5.025e-06***	0.0045
size_sender			5.760e-04	5.064e-06***			6.147e-04	5.028e-06***	0.0039
Websize_receiver					4.825e-04	2.389e-06***			
Websize_sender					4.018e-04	4.018e-04***			
Websize_receiver_ norm							2.433e-03	3.395e-05***	0.0022
Websize_sender_ norm							2.237e-03	3.398e-05***	0.0020
Log_distance			-1.320e+00	1.848e-02***	-9.891e-01	1.871e-02***	-1.142e+00	1.821e-02***	-0.0026
Country			1.604e+00	2.074e-02***	1.724e+00	2.099e-02***	1.628e+00	2.039e-02***	0.0033
subject			1.253e+00	2.179e-02***	1.399e+00	2.191e-02***	1.273e+00	2.153e-02***	0.0020
phd			1.512e-01	3.829e-02***	2.952e-01	3.910e-02***	1.435e-01	3.791e-02***	0.0001
Log(theta)	0.1409	0006***	0.25725	0.00122***	0.25278	0.00119	0.26838	0.00129	
-2Log-likelihood	-783945		-721150		-722098		-716876		
Residual deviance	138716		141584		140730		141631		
AIC	783949		721167		722114		716896		

*Signif. Codes 0*** 0.001** 0.01* 0.05'*

Results are as follows. The model including web domain size instead of organizational size provides a lower fit, showing that the counts of web links is better predicted by organizational size than by the number of webpages. The difference is highly statistically significant. This supports the hypothesis that the underlying mechanisms of web links are institutional relationships rather than document relationships.

Expectedly, the model including both organizational size and web domain size (normalized) provides a significantly better fit, showing that, nevertheless, the size of the web domain (relative to organizational size) has some impact on number of web links. This might be explained by the fact that, when the size of the web domain increases, an underlying institutional relationship is translated into a larger number of links. The fact that all other coefficients are not significantly affected by the introduction of the web domain size supports this interpretation: a larger website multiplies all count of web links sent and received by an HEI, but does not affect their distribution across the network. The only significant interaction effect is with distance, which will be investigated in a further step by introducing in the model cross-terms between web domain size and distance.

Finally, an analysis of the strength of the effects through beta coefficients shows that the impact of web domain size (normalized) is about half of the one of organizational size and smaller than the one of distance and of country; thus, despite large differences in the size of web domains, the main pattern of interlinking between HEIs remains determined by organizational attributes and spatial factor, consistently with their interpretation as institutional networks.

Discussion

As the starting hypothesis was that web links mostly follow a social relationships model and the introduction of page counts does not influence significantly this model, the results show that web domain size certainly improves the fit. So, the proposed way of addressing the problem is promising, as there is evidence that both organizational and document properties matter, with a strong evidence that the former might be more important for predicting web links.

The preliminary results have relevant implications for applying link analysis in the description of complex heterogeneous organizations like universities. The release of rankings based on link visibility indicators as suitable proxies of performance in global scenarios is supported. But interlinking is not the only way to evaluate web visibility and other indicators should be explored (PageRanks, HITS, co-links, etc.)

References

Buckland, M. & Gey, F. (1994). The relationship between recall and precision. *Journal of the American Society for Information Science*, 45, 12–19.

Adamic, Lada A. & Huberman, Bernardo A. (2000). Power-Law Distribution of the World Wide Web. Science, 287 (5461), 2115. DOI: 10.1126/science.287.5461.2115[a]

Aguillo, I.F., Granadino, B., Ortega, J.L., Prieto, J.A. (2006). Scientific research activity and communication measured with cybermetrics indicators. *Journal of the American Society for Information Science and Technology*, 57(10), 1296–1302. DOI: 10.1002/asi.20433

Cameron, A. C. & Trivedi, P. K. (1998). Regression *Analysis of Count Data*. Cambridge: Cambridge University Press

Cronin, Blaise (2001). Bibliometrics and beyond: some thoughts on web-based citation analysis. *Journal of Information Science*, 27(1), 1–7. DOI: 10.1177/016555150102700101

Daraganova, G., Pattison, P., Koskinen, J., Mitchell, B., Bill, A., Watts, M. & Baum, S. (2012). Networks and geography: Modelling community network structures as the outcome of both spatial and network processes. *Social Networks*, 34 (1), 6–17. DOI: 10.1016/j.socnet.2010.12.001

Ijiri, Yuji & Simon, Herbert A. (1964). Business Firm Growth and Size. *The American Economic Review*, 54 (2), 77–89

Lepori, B. & Bonaccorsi, A. (2013). Towards an European census of higher education institutions. Design, methodological and comparability issues. Minerva, forthcoming.

Seeber, M., Lepori, B., Lomi, A., Aguillo, I., Barberio, V. (2012). Factors affecting web links between European higher education institutions. *Journal of Informetrics*, 6 (3), 435–447. DOI: 10.1016/j.joi.2012.03.001

Thelwall, Mike (2001). Extracting macroscopic information from Web links. *Journal of the American Society for Information Science and Technology*, 52 (13), 1157–1168. DOI: 10.1002/asi.1182

Thelwall, Mike & Wilkinson, David (2004). Finding similar academic Web sites with links, bibliometric couplings and colinks. *Information Processing & Management*, 40 (3), 515–526. DOI: 10.1016/S0306-4573(03)00042-6

van Raan, A. F. J. (2007). Bibliometric statistical properties of the 100 largest European universities: prevalent scaling rules in the science system. *Eprint arXiv*:0704.0889

Vaughan, Liwen & Thelwall, Mike (2005). A modeling approach to uncover hyperlink patterns: the case of Canadian universities. *Information Processing & Management*, 41(2), 347–359. DOI: 10.1016/j.ipm.2003.10.001

Knowledge and technology transfer performance of European universities and research institutions: assessing the influence of institutional by-laws and practices[1]

Franz Barjak franz.barjak@fhnw.ch / School of Business, University of Applied Sciences and Arts Northwestern Switzerland (FHNW), Riggenbachstrasse 16, 4600 Olten (Switzerland)

Nordine Es-Sadki n.es-sadki@maastrichtuniversity.nl / Maastricht Economic Research Institute on Innovation and Technology (UNU-MERIT), Keizer Karelplein 19, 6211 TC Maastricht (the Netherlands)

Anthony Arundel anthony.arundel@utas.edu.au / a.arundel@maastrichtuniversity.nl / University of Tasmania, Hobart 7004, Australia and UNU-MERIT

Abstract

Several studies have shown that the knowledge and technology transfer (KTT) performance of universities and research institutions can be explained by institutional differences. Some of the factors that influence performance include providing staff incentives and the establishment of a knowledge transfer office (KTO). However, only a few empirical studies have looked at the influence of institutional by-laws and practices on KTT performance, even though they might have strong impacts. We use 2010 and 2011 data for 224 European universities and 48 public research organisations to investigate the effects of KTO characteristics, staff incentives, and policies on three measures of KTT: the number of invention disclosures, patent applications and licenses. Using a negative binomial count model, we find that a written IP policy is correlated with an increase in all three outcomes, although open publication of policies has no effect. Rules on managing conflict are positively correlated with the number of invention disclosures and patent applications. The only incentive for researchers that has a consistent positive effect on all outcome measures is to give inventors a share of the revenues. Non-monetary rewards such as career enhancement or social rewards have no effect.

Introduction

Since the early 1990s an increasing stream of publications has investigated the knowledge and technology transfer (KTT) performance of universities and public research institutions and the effect of different policies on this performance. Awareness among university administrators and

1 This paper draws on the data and results of the *Knowledge Transfer Study 2010-12* conducted for the European Commission, DG Research & Innovation, under contract RTD/DirC/C2/2010/SI2.569045 by empirica Communication and Technology Research, UNU-MERIT and FHNW. We gratefully acknowledge this support.

policy makers about the importance of regulations governing KTT activities has also risen. Yet there is little evidence on which policies are effective and which are not. This is of particular concern for Europe, which has a heterogeneous pattern of national laws and regulations on the ownership and transfer of academic research results, as well as different levels of enforcement of existing laws (Geuna & Rossi, 2011). To encourage good practice, the European Commission issued in 2008 a recommendation for a Code of Practice (COP) containing eighteen principles for the management of intellectual property and knowledge transfer activities by universities and public research institutes (European Commission, 2008). The COP includes seven principles for internal intellectual property (IP) policy, seven principles for knowledge transfer policy, and another four principles for collaborative and contract research. EU member states were encouraged to adapt their national guidelines and regulations accordingly and support their universities and public research organisations in the process of making guidelines and practices compatible with the COP.

This paper uses a recent dataset to explore the effect of KTO characteristics and institutional policies on KTT performance outcomes. The dataset provides information on both of these factors for up to 224 European universities and 48 European public research institutes in 32 European countries, permitting an evaluation of the effect of a greater range of policies than possible for single country studies. Most previous research is based on a single country, such as research using the US Association of University Technology Managers (AUTM) dataset. One of the few exceptions using a multi-country dataset is Conti (2009).

The analyses address several gaps in the academic literature in respect to the effect of IP policies on KTT performance. The policies cover staff incentives, how license revenues are shared, and rules for invention disclosure and managing conflicts of interest.

We first give a short overview of the relevant literature, then describe the sample and data collection methods, present key results and close with a discussion and conclusions.

Transfer performance – state-of-the-art

Previous studies have evaluated the effect of a large number of external and internal influences on KTT performance, including KTO characteristics, the characteristics of the affiliated university or research institute and institutional and national policies.

The most widely studied factors concern the *characteristics of the KTO*. These include the size and age of the KTO (Carlsson & Fridh, 2002; Conti, 2009; Friedman & Silberman, 2003; Link & Siegel, 2005; Siegel, Waldman, & Link, 2003), the industry background and experience of KTO staff (Conti, 2009; Siegel, Waldman, Atwater, & Link, 2003), and other KTO characteristics including the level of autonomy of the KTO from its affiliated university (Conti, 2009; Markman, Gianiodis, & Phan, 2009; Siegel, Waldman, & Link, 2003), the degree of centralization of services

(Bercovitz, Feldman, Feller, & Burton, 2001; Debackere & Veugelers, 2005) and transfer strategies (Belenzon & Schankerman, 2009; Friedman & Silberman, 2003; Litan, Mitchell, & Reedy, 2007).

The *characteristics of the affiliated institution include* ownership (public or private), size, the existence of engineering and natural sciences departments and hospitals, and research excellence and business orientation. These factors have been evaluated by Belenzon & Schankerman (2009) and Caldera & Debande (2010).

Of greatest interest to this study is the effect of IP/KTT policies on knowledge transfer. These can include *national laws and regulations* on IP ownership discussed for instance in Geuna & Rossi (2011), Goldfarb & Henrekson (2003), and Valentin & Jensen (2007).

Institutional policies and practices include incentives for researchers. A lack of such incentives is perceived as one of the main barriers against knowledge transfer (Siegel, Waldman, & Link, 2003; Siegel, Waldman, Atwater, & Link, 2004). Incentives for university inventors, such as a share of revenues, have been identified as an influential factor in transfer performance in many studies (see e.g. Lach & Schankerman, 2004; Lach & Schankerman, 2008).

In addition, university by-laws on patenting and IP can act as a stimulus to promote patenting (Baldini, 2010; Baldini, Grimaldi, & Sobrero, 2006), whereas an impact on licensing agreements could not be found (González-Pernía, Kuechle, & Peña-Legazkue, 2013). Internal rules on various third mission activities can influence transfer success. These include rules on conflicts of interest, which correlate with higher numbers of university-industry R&D contracts and income from such contracts, licences executed and income from licences and numbers of start-ups in Spanish universities (Caldera & Debande, 2010). In the same study, rules on R&D contracts correlate negatively with the number of contracts. The authors suggest that this is because the rules stipulate the participation of the institution in the resulting benefits from commercialisation. Lerner (2004) argues that too vigorous conflict-of-interest regulations can also have a chilling effect on entrepreneurial activity.

Methodology

Data were collected in two separate questionnaire surveys in 2011 and in 2012. Each survey referred to the preceding year. The EKTIS survey collected data on the characteristics of KTOs serving universities and public research institutes (size, year of establishment etc.) and data on performance outcomes for 2010 and 2011. It focused on the leading research-intensive universities and research institutes in each of the 27 EU member states, the EFTA countries, and EU candidate countries. The sample was weighted by each country's share of total European Government and Higher Education R&D expenditures, with data collected for approximately 500 universities and research institutes combined. The second survey used a sub-sample of EKTIS respondents to collect information on institutional and national policies for knowledge transfer in 2010 and 2011.

The two sets of survey data were then linked. The number of cases for analysis is increased by including all respondents for at least one of the two surveys. If a respondent replied to both surveys, the results for the most recent year (2011) are used. Since economic conditions, research funding and KT policies in Europe did not change notably between 2010 and 2011, any biases due to combining years should be minor. The linked dataset includes 170 cases for which 2011 data are available and 118 cases for which only 2010 data were available. Further information on the survey methodologies is available on request from the authors.

The effect of multiple policies and KTO characteristics on performance was examined in a series of regressions for each of the three outcome indicators: invention disclosures, patent applications and executed licenses. These three indicators cover one of the commercialisation pathways, from the identification of potentially valuable discoveries to the successful transfer of the IP to the private sector though a license. A patent is not required for all licenses, but as over 78% of license income earned by European universities and research institutions is based on a patented invention (Arundel & Bordoy, 2009), patents applications are an important intermediary step in the commercialisation process.

The independent variable in all regressions consists of the number of outcomes, using count data. As there are a significant number of zero responses (see Table 1), we use a negative binomial model appropriate for count data. Note that due to the cross-sectional nature of the data set it is not possible to unambiguously identify cause and effect, for example whether policies drive performance or if institutions alter their policies as a result of performance outcomes. However, the latter is more likely to occur in response to poor performance, which would reduce the ability to identify policies with a positive effect on outcomes.

Every model was first estimated with a set of control variables to account for the characteristics of the institution and KTO: the number of researchers at each institution (numb_res), if the institution is a research institute or a university (univers), whether it has a hospital (hospital), whether the KTO was established before the year 2000 (KTO_age) and the number of KTO staff in full time equivalents (KTO_size). In addition, another dummy was included for institutions where the inventors can own the IP resulting from their research (IPR_ownership). The control variables were kept only if they were significant in the baseline regression (as part of the sensitivity analysis we also included them in some of the regressions with the policy variables, but as their inclusion had no notable effect the results are not presented here).

The regressions were conducted including a dummy variable for year (2011 versus 2010). The variable had no effect on the number of invention disclosures or patent applications, but it did have a negative effect on the number of licenses, indicating that license activity declined in 2011. However, the variable for year had no effect on any of the policy variables and is consequently not included in the final regressions given below.

Table 5 in the appendix defines each of the institutional policies included in the regressions.

Descriptive results

Table 1 gives the number of outcomes for the three indicators that are the focus of this study: invention disclosures, patent applications, and licenses. On average, universities reported 33.0 invention disclosures, 14.4 patent applications, and 11.6 executed licenses.

Table 1. Key performance indicators for universities and other public research institutes

	Universities				Other research organisations			
	N[1]	Mean	Total reported	Perc. zero[2]	N[1]	Mean	Total reported	Perc. zero[2]
Invention disclosures	219	33.0	7,228	10.5%	44	32.8	1,441	11.4%
Patent applications	223	14.4	3,215	18.8%	44	13.7	601	11.4%
Licenses executed	196	11.6	2,279	26.5%	41	11.8	483	14.6%

Source: MERIT, European Knowledge Transfer Indicator Survey 2011 and 2012.

1: Number of KTOs reporting results for each performance measure (including zero outcomes).

2: Percent of respondents reporting 'zero' for each outcome. For example, 10.5% of 219 universities reported zero invention disclosures in 2010 and/or 2011.

Tables 2, 3 and 4 provide the number and percent of universities and other research organisations that reported specific types of policies. As shown in Table 2, the most common IP policy consists of rules on who owns the IP, with 85.2% of universities reporting such a rule. In contrast, only 6.5% of universities have published their license policy, although 25.2% have one in writing.

Table 2. IP and licence policies

	Universities			Other research organisations		
	N	Answer = Yes	in %	N	Answer = Yes	in %
Institution has a written IP policy	223	150	67.3	48	33	68.8
Institution has published the IP policy	223	60	26.9	48	12	25.0
IP policy includes rules on invention disclosures	237	188	79.3	49	42	85.7
IP policy includes rules on IP ownership	237	202	85.2	49	45	91.8
IP policy includes rules for conflict management	237	120	50.6	49	23	46.9
Institution has a written licence policy	214	54	25.2	46	15	32.6
Institution has published the licence policy	214	14	6.5	46	3	6.5

Source: FHNW, European Knowledge Transfer Practice Surveys 2011 and 2012.

Table 3 gives results for different types of incentives. The most common incentive, reported by 81.7% of universities, is to provide researchers with a percentage of the revenues. Salary upgrades are the least common, reported by only 6.1% of universities.

Table 3. Provision of incentives for researchers and students to protect and exploit IP

	Universities			Other research organisations		
	N[1]	Answer = Yes	in %	N[1]	Answer = Yes	in %
Percentage of the revenues	213	174	81.7	47	41	87.2
Lump-sum payments	213	48	22.5	47	18	38.3
Salary upgrades	214	13	6.1	47	4	8.5
Additional funds for R&D	214	72	33.6	47	16	34.0
Inclusion in promotion & career decisions	214	53	24.8	47	15	31.9
Social rewards (e.g. awards, publicity)	214	117	54.7	47	23	48.9

Source: FHNW, European Knowledge Transfer Practice Surveys 2011 and 2012.
1: Number of KTOs answering the question.

Table 4 provides the average distribution of knowledge transfer revenues by the type of recipient. On average, the inventor and other researchers obtain approximately 40% of the revenues, with the KTO receiving slightly over 7%.

Table 4. Average shares of revenues from IP by recipient

	Universities		Other research organisations	
	N[1]	in %	N[1]	in %
Department(s), institute(s) or other institutional subunits	176	20.2	41	15.9
Institution	176	30.3	41	36.7
Inventor(s), researcher(s) from the institution	176	40.7	41	38.7
KTO or other intermediaries	147	7.5	38	7.1
Other beneficiaries	176	2.6	41	2.1
Total		**100.0**		**100.0**

Source: FHNW, European Knowledge Transfer Practice Surveys 2011 and 2012.
1: Number of KTOs answering the question. KTO shares were collected separately only in 2012.

Regression results

The baseline model for *invention disclosures* finds that the number of researchers (numb_res), the number of staff at the KTO in full time equivalents (KTO_size), the presence of a hospital and when the case is a research organisation (the reference category is a university) are all positively correlated with the number of invention disclosures (see Table 6 in the appendix). Whether or not the inventor owns the IP (IPR_ownership) and whether or not the KTO was established before 2000 (KTO_age) have no effect on the number of invention disclosures.

For the policy variables, the existence of a written IP policy (IP_policy) is significantly correlated with an increase in invention disclosures in all models except for model 8. As the European COP promotes the publication of IP policy, we ran regression 1 with an additional dummy for this (results not shown). It had no effect. Three of the incentives for active staff involvement in knowledge transfer significantly increase the number of invention disclosures: lump sum payments (INC_lumpsum), higher salaries (INC_salary) and social rewards (INC_social). Giving inventors a percentage of the revenues (INC_percentage) is not significant ($p = 0.16$), possibly because this incentive is so common that the model is unable to detect a significant effect. When all incentives that are positive and significant (plus INC_percentage) are included simultaneously, only the percentage of the revenues and a higher salary correlate with higher invention disclosures (model 8). Based on this we would also expect that the average share of the revenues given to inventors (Share_invent) is significant. This is not the case, however. The share given to the department is positive and significant. The share given to the institution as a whole is negatively related to the number of invention disclosures.

All three of the rules governing IP are statistically significant and positive if included separately. However, none are significant when included together (see model 16).

The results are similar for *patent applications* (see Table 7), except that the presence of a hospital does not increase the number of applications. A written IP policy is positively correlated to the number of patent applications. Whether the policy is published or not (not shown) does not make any difference. As is the case for invention disclosures, only giving inventors a share of the revenues and paying higher salaries is significantly correlated with the number of patent applications. None of the variables for the distribution of revenues correlates with patent applications. Among the rules for IP-related issues, a rule for the management of conflicts (Rule_Conflictemp) is consistently significant (models 15 & 16).

For *licence agreements*, a written licence policy (Lic_policy) is the most consistently significant policy variable. The variable for IP policy also remains significant in most models (see Table 8). Whether the licence policy is published has no effect (results not shown). The strongest incentive is giving inventors a share of the revenues, whereas lump-sum payments have a negative effect on licensing (INC_lumpsum in models 3 and 8). As in the case for invention disclosures, the share of revenues given to the department (Share_depart) is positively related to the number of licence

agreements. The share of license income given to the KTO has a negative effect (Share_kto). Among the variables for IP rules, the existence of a rule for disclosure is negatively correlated to the number of licenses while a rule on IP ownership has a positive impact on the number of licence agreements.

Discussion and conclusions

Our results add to current knowledge on the relationship between institutional rules and practices and transfer performance.

First, a written IP policy is positively correlated to all three performance measures and a written licence policy is positively related to licence agreements. This matches with Baldini et al. (2006) who showed with panel data on Italian universities that the introduction of an IPR regulation increased their patent applications; however, it contradicts González-Pernía et al. (2013) who do not find an effect of such regulations on licensing agreements. The European COP promotes these policies in principles 1 and 9. It also stipulates that the policies should be publicised. We did not find any additional benefit from publication. All this suggests that the benefit of clarifying and codifying institutional practice in regard to IP management and KTT is more internal than external.

Second, providing rules on how to manage conflicts of interest – as put forth in principle 2 of the COP – is positively correlated with the number of invention disclosures and patent applications. This matches the findings of Caldera and Debande (2010). However, we do not know if many disclosures/applications create a need for formalising approaches to dealing with conflicts, or whether such rules are conducive to increasing disclosures and subsequent patent applications. Another result might help to get a better understanding of the direction of the statistical relationship: The provision of rules on the disclosure of inventions is positively correlated with invention disclosures and negatively with licence agreements. This latter finding makes us believe that policies are driven by performance rather than vice versa. Institutions with an unsatisfactory licensing performance might find that one approach to change this is by tightening the screws on their faculty's handling of inventions.

Third, the only incentive that has a consistently positive effect on the three outcome measures is to give inventors a share of the revenues. The causality seems to be clear in this case, as a university would find little reason to motivate their faculty to become involved in transfer activities if it already performs well. Hence we would conclude that this particular type of incentive has a positive effect on the willingness of faculty to disclose, support the patent application and help with finding licensees and concluding a licence agreement. Raising the salary as an incentive also has a positive effect, though notably not on licence agreements. Of note, non-monetary rewards such as career enhancements or social rewards have no effect. The relevant recommendation of the COP to support these types of rewards (in principle 4) does not seem to be justified.

Fourth, the share of revenues from transfer activities that is allocated to inventors is unrelated to the outcome measures, which might seem illogical and counter-intuitive. It is certainly not in line with American experience (see e.g. Link & Siegel, 2005). One explanation could be the large variety of IP ownership rules in Europe (Geuna & Rossi, 2011), possibly leading to poor enforcement.[i] Universities which either give the inventors the IP or which do not enforce university ownership rights will have fewer invention disclosures and patent applications because the inventor will either not need to or actively avoid disclosing potentially valuable inventions. Other universities operating in a legal framework where they own the IP generated by their faculty, e.g. based on national university laws, might raise invention disclosures by offering their faculty higher personal benefits such as revenue shares. Essentially, different logics apply. We can confirm this argument by comparing the revenue shares given to inventors between organisations where inventors also own (some of) the IP, and where they don't: in the former the inventors' share of revenues from IP is on average 48% and in the latter, at 36%, significantly lower (ANOVA, $p < .01$). This might also suggest that organisations are forced to provide higher revenue shares when the inventor has the option of commercialising the invention privately.

References

Arundel, A., & Bordoy, C. (2009) Final results for the ASTP survey for Fiscal Year 2008, UNU-MERIT report to the ASTP, UNU-MERIT, Maastricht.

Baldini, N. (2010). Do royalties really foster university patenting activity? An answer from Italy. *Technovation*, 30(2), 109–116.

Baldini, N., Grimaldi, R., & Sobrero, M. (2006). Institutional changes and the commercialization of academic knowledge: A study of Italian universities' patenting activities between 1965 and 2002. *Research Policy*, 35(4), 518–532.

Belenzon, S., & Schankerman, M. (2009). University Knowledge Transfer: Private Ownership, Incentives, and Local Development Objectives. *Journal of Law & Economics*, 52, 111–144.

Bercovitz, J., Feldman, M., Feller, I., & Burton, R. (2001). Organizational Structure as a Determinant of Academic Patent and Licensing Behavior: An Exploratory Study of Duke, Johns Hopkins, and Pennsylvania State Universities. *The Journal of Technology Transfer*, 26(1), 21–35.

Caldera, A., & Debande, O. (2010). Performance of Spanish universities in technology transfer: An empirical analysis. *Research Policy*, 39(9), 1160–1173.

Carlsson, B., & Fridh, A.-C. (2002). Technology transfer in United States universities. *Journal of Evolutionary Economics*, 12, 199–232.

Conti, A. (2009). *Three Essays on University-Industry Technology Transfer and the Economics of Science*.Doctoral degree, ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, Lausanne.

Debackere, K., & Veugelers, R. (2005). The role of academic technology transfer organizations in improving industry science links. *Research Policy*, 34(3), 321–342.

European Commission. (2008). *Recommendation on the management of intellectual property in knowledge transfer activities and Code of Practice for universities and other public research organisations*. Luxembourg: Office for Official Publications of the European Communities Retrieved from http://ec.europa.eu/invest-in-research/pdf/download_en/ip_recommendation.pdf.

Friedman, J., & Silberman, J. (2003). University Technology Transfer: Do Incentives, Management, and Location Matter? *The Journal of Technology Transfer*, 28(1), 17–30.

Geuna, A., & Rossi, F. (2011). Changes to university IPR regulations in Europe and the impact on academic patenting. *Research Policy*, 40(8), 1068–1076.

Goldfarb, B., & Henrekson, M. (2003). Bottom-up versus top-down policies towards the commercialization of university intellectual property. *Research Policy*, 32(4), 639–658.

González-Pernía, J. L., Kuechle, G., & Peña-Legazkue, I. (2013). An Assessment of the Determinants of University Technology Transfer. *Economic Development Quarterly*, 27(1), 6–17.

Lach, S., & Schankerman, M. (2004). Royalty Sharing and Technology Licensing in Universities. *Journal of the European Economic Association*, 2(2–3), 252–264.

Lach, S., & Schankerman, M. (2008). Incentives and invention in universities. *The RAND Journal of Economics*, 39(2), 403–433.

Lerner, J. (2004). The University and the Start-Up: Lessons from the Past Two Decades. *The Journal of Technology Transfer*, 30(1), 49–56.

Link, A. N., & Siegel, D. S. (2005). Generating science-based growth: an econometric analysis of the impact of organizational incentives on university–industry technology transfer. *The European Journal of Finance*, 11(3), 169–181.

Litan, R. E., Mitchell, L., & Reedy, E. J. (2007). Commercializing university innovations: A better way. NBER Working Paper.

Markman, G. D., Gianiodis, P. T., & Phan, P. H. (2009). Supply-Side Innovation and Technology Commercialization. *Journal of Management Studies*, 46(4), 625–649.

Siegel, D. S., Waldman, D., & Link, A. (2003). Assessing the impact of organizational practices on the relative productivity of university technology transfer offices: an exploratory study. *Research Policy*, 32(1), 27–48.

Siegel, D. S., Waldman, D. A., Atwater, L. E., & Link, A. N. (2003). Commercial knowledge transfers from universities to firms: improving the effectiveness of university-industry collaboration. *The Journal of High Technology Management Research*, 14(1), 111–133.

Siegel, D. S., Waldman, D. A., Atwater, L. E., & Link, A. N. (2004). Toward a model of the effective transfer of scientific knowledge from academicians to practitioners: qualitative evidence from the commercialization of university technologies. *Journal of Engineering and Technology Management*, 21(1–2), 115–142.

Valentin, F., & Jensen, R. (2007). Effects on academia-industry collaboration of extending university property rights. *The Journal of Technology Transfer*, 32(3), 251–276.

Appendix

Table 5. Overview of the variables in the regression analyses

IP_policy	Institution has a written IP policy (defined as principles implemented to identify, protect and manage the IP resulting from R&D activities in which faculty or staff from the institution is involved.)
Lic_policy	Institution has a written licence policy (defined as principles that rule the granting of licenses or similar rights to users of IP owned by the institution.)
Rule_empdisclo	Rules for employees for invention disclosures
Rule_IPemp	Rules for employees for IP ownership
Rule_Conflictemp	Rules for managing conflicts of interest
INC_percentage	Incentive based on the percentage of the revenues
INC_lumpsum	Incentive in the form of lump-sum payments
INC_promotion	Incentive from a salary upgrades
INC_Rdfunds	Incentive in the form of additional funds for R&D
INC_salary	Incentive via promotion & career decisions
INC_social	Incentive based on social rewards (awards, publicity)
Share_invent	Share of license revenue for inventor(s) and researcher(s)
Share_depart	Share of license revenue for department(s), institute(s) or other institutional subunits
Share_kto	Share of licence revenue for the KTO or other intermediaries
Share_inst	Share of licence revenue for the institution

Source: Authors.

Table 6. Negative binomial regressions on invention disclosures in fiscal year 2011 (or 2010, if 2011 missing)

	Basis	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
N	210	220	205	206	206	206	206	206	205	172	172	139	172	219	219	219	219
numb_res	0.202**	0.249**	0.221**	0.219**	0.232**	0.228**	0.222**	0.241**	0.219**	0.212**	0.187**	0.180**	0.195**	0.250**	0.257**	0.264**	0.265**
IPR_owner-ship	-0.069																
hospital	0.641**	0.696**	0.659**	0.733**	0.698**	0.688**	0.749**	0.723**	0.755**	0.689**	0.695**	0.705**	0.645**	0.746**	0.697**	0.739**	0.741**
inst	0.387*	0.288	0.24	0.164	0.283	0.269	0.277	0.270	0.168	0.279	0.380+	0.433*	0.297	0.250	0.270	0.347+	0.307+
KTO_age	0.045																
KTO_size	0.042**	0.032**	0.031**	0.033**	0.031**	0.032**	0.033**	0.029**	0.031**	0.038**	0.042**	0.036**	0.037**	0.031**	0.029**	0.029**	0.028**
IP_policy		0.571**	0.448*	0.447**	0.436*	0.490**	0.416*	0.435*	0.273	0.568**	0.426*	0.775*	0.523**	0.546**	0.517**	0.509**	0.480**
INC_percentage			0.293						0.419+								
INC_lumpsum				0.379*					0.264								
INC_promo-tion					0.195												
INC_Rdfunds						0.064											
INC_salary							0.495+		0.538*								
INC_social								0.252+	0.243								
Share_invent										0.05							
Share_depart											0.10*						
Share_kto												-0.002					
Share_inst													-0.008*				
Rule_empdisclo														0.437*			0.449
Rule_IPemp															0.607*		0.139
Rule_Conflictemp																0.310*	0.225

Source: Authors.

Table 7. Negative binomial regressions on patent applications in fiscal year 2011 (or 2010, if 2011 missing)

	Basis	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
N	214	204	191	191	192	192	192	192	191	163	163	133	163	203	203	203	203
numb_res	0.25**	0.283**	0.271**	0.267**	0.279**	0.278**	0.263**	0.287**	0.254**	0.250**	0.243**	0.216**	0.248**	0.290**	0.293**	0.302**	0.302**
IPR_owner-ship	0.248	0.235															
hospital	0.093	0.025															
inst	0.379*	0.312	0.254	0.190	0.267	0.267	0.239	0.297	0.180	0.459*	0.479*	0.699**	0.463*	0.275+	0.297	0.336+	0.307
KTO_age	0.242	0.271+	0.301+	0.348*	0.349*	0.295+	0.307+	0.335*	0.299*	0.339+	0.299+	0.173	0.309+	0.282+	0.273+	0.265+	0.253
KTO_size	0.023**	0.017*	0.018*	0.019*	0.016*	0.017*	0.019*	0.016+	0.019*	0.024**	0.024**	0.033**	0.024**	0.016*	0.015+	0.016*	0.014+
IP_policy		0.626**	0.557**	0.568**	0.551**	0.606**	0.506**	0.541**	0.356+	0.548**	0.490*	0.983**	0.537**	0.533**	0.534**	0.501**	0.458*
INC_percentage			0.311						0.541*								
INC_lumpsum				0.303+					0.126								
INC_promo-tion					0.280												
INC_Rdfunds						0.203											
INC_salary							0.688*		0.785*								
INC_social								0.229	0.213								
Share_invent										0.03							
Share_depart											0.03						
Share_kto												-0.06					
Share_inst													0.000				
Rule_empdisclo														0.404+			0.227
Rule_IPemp															0.413		0.106
Rule_Conflictemp																0.361*	0.296+

Source: Authors.

Table 8. Negative binomial regressions on licence agreements in fiscal year 2011 (or 2010, if 2011 missing)

	Basis	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
N	197	184	173	173	174	174	174	174	173	151	151	121	160	183	183	183	183
numb_res	0.171**	0.235**	0.179**	0.207**	0.207**	0.205**	0.209**	0.216**	0.185**	0.147**	0.138**	0.243**	0.149**	0.235**	0.234**	0.236**	0.259**
IPR_ownership	-0.025																
hospital	0.195																
inst	0.464*	0.384+	0.242	0.438*	0.350+	0.382+	0.342+	0.356+	0.392+	0.469*	0.548*	0.892**	0.468*	0.449*	0.369+	0.408*	0.494*
KTO_age	0.381*	0.271	0.275	0.293+	0.296+	0.309+	0.301+	0.287+	0.247	0.344+	0.410*	0.572**	0.380*	0.328*	0.263	0.275+	0.300+
KTO_size	0.035**	0.027**	0.028**	0.026**	0.027**	0.030**	0.028**	0.026**	0.024**	0.036**	0.037**	0.028**	0.034**	0.028**	0.026**	0.025**	0.022**
IP_policy		0.409*	0.190	0.385+	0.311	0.331	0.364+	0.347+	0.162	0.468*	0.351	0.661*	0.466*	0.498*	0.327	0.358+	0.374+
Lic_policy		0.735**	0.784**	0.783**	0.723**	0.753**	0.749**	0.678**	0.689**	0.555**	0.428*	0.665**	0.571**	0.714**	0.752**	0.670**	0.603**
INC_percentage			0.829**						1.059**								
INC_lumpsum				-0.355+					-0.446*								
INC_promotion					0.158												
INC_Rdfunds						-0.214											
INC_salary							-0.235		0.359								
INC_social								0.139	0.282								
Share_invent										-0.002							
Share_depart											0.013**						
Share_kto												-0.022*					
Share_inst													-0.003				
Rule_empdisclo														-0.389+			-0.808**
Rule_IPemp															0.245		0.748*
Rule_Conflictemp																0.174	0.275

Source: Authors.

i There are no data for enforcement, but several KTO managers interviewed by the authors reported reasons for low enforcement of IP ownership rules at their university.

An Indicator of Translational Capacity of Biomedical Researchers[1]

Kevin W. Boyack kboyack@mapofscience.com / SciTech Strategies, Inc., Albuquerque, NM 87122 (USA)

Richard Klavans rklavans@mapofscience.com / SciTech Strategies, Inc., Berwyn, PA 19312 (USA)

Michael Patek mpatek@mapofscience.com / SciTech Strategies, Inc., Berwyn, PA 19312 (USA)

Patrick Yoon yeyoon@cis.upenn.edu / CIS Department, University of Pennsylvania, Philadelphia, PA 19104 (USA)

Lyle H. Ungar ungar@cis.upenn.edu / CIS Department, University of Pennsylvania, Philadelphia, PA 19104 (USA)

Abstract

The recent focus on translational research is aimed at reducing the time required for basic discoveries to be translated into effective patient treatment. Much of the focus is on the linkages between basic biomedical research and clinical medicine. These linkages are likely to best improved by individual researchers and teams. We have thus developed an indicator to assess the translational capacity of medical researchers and teams. This is done using the system of research levels first established by CHI. A new model, based on words from titles, abstracts, and references, has been created to assign individual articles to research levels. To show the utility of the new model, these research levels are then used to characterize the output of top researchers in the field of Alzheimer's Disease. We propose an indicator of translational capacity based on the distribution of the output of a researcher across research levels. Although most top researchers tend to publish in a single research level, many publish sufficiently in two levels to suggest that they have high translational capacity.

Introduction

Translational research has become a topic of great interest in the biomedical community over the past several years. The US National Institutes of Health (NIH) has made it a recent priority as evidenced by its funding of the Clinical and Translational Science Awards (CTSA) program[i]. Translational research is also specifically promoted in the European Commission's Seventh

1 This research is supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D11PC20152. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government. We also acknowledge Aaron Sorensen for continually bugging us to make sure this work got done.

Framework Programme (FP7)[ii]. The UK Medical Research Council (MRC) names translational research as one of its research initiatives, and has dedicated translation funding programs[iii].

Translational research has different meanings in different contexts, but is now widely thought of as consisting of two main types (Woolf, 2008). The first type, often labelled as T1, is known by terms such as "bench-to-bedside" and describes the interface between basic science and clinical medicine. It is aimed at taking the information from basic discoveries and incorporating them into new approaches for prevention, treatments, etc. The second type, often labelled T2, is an extension of the first. It is "the translation of results from clinical studies into everyday clinical practice and health decision making" (Sung et al., 2003).

With this recent focus on translational research comes the need to be able to assess the capacity of researchers and teams to be effective at the translation process. This study focuses on assessing translation type T1 because the biomedical literature contains a wealth of information along the spectrum from basic science to clinical observation. One impedance to the speed of translation, and in many cases to any translation at all, is the fact that most researchers are so highly specialized that they cannot move their research to the next stage. Most researchers speak only a single language, whether it be the language of basic science or the language of clinical trials, or any of the possible languages in between. Translators are needed. This study uses the notion of research level, first introduced by Narin, Pinski & Gee (1976), to assess the translational capacity of individual researchers. We demonstrate this by 1) developing an improved method to accurately assign research levels to individual articles, and 2) applying this method to researchers in the field of Alzheimer's Disease.

Background

Narin et al. (1976) and CHI Research introduced the world to research levels (RL) when they classified 900 biomedical journals into four research levels. Journals were assigned to a RL based on a combination of expert knowledge and citation patterns. The citation pattern portion was based on the assumption that clinical research would cite basic research, but that the reverse would not be true. Thus, given the types in Table 1, journals in RL1 would cite journals in RL2, RL3, and RL4, but journals in RL4 would only cite other RL4 journals.

Table 1. CHI research levels and exemplars.

Research Level	Type	Example Journal
RL1	Clinical observation	British Medical Journal
RL2	Clinical mix	New England Journal of Medicine
RL3	Clinical investigation	Immunology
RL4	Basic research	Nature

The CHI research level classification system was expanded in the 1980s to include journals in the physical sciences (Carpenter et al., 1988). Although additional journals have been added to the research level list at various times, of the nearly 20,000 journals available in Scopus, only around 4,000 have assigned research levels (Boyack & Klavans, 2011).

Since their introduction, RL have been used to characterize research at academic institutions (Carpenter et al., 1988; McAllister & Narin, 1983), and have been correlated to the attraction of funding (with basic research favored) in Australia (Butler, Biglia, & Bourke, 1998). It has also been shown that the majority of the biomedical papers cited by industrial patents are from the basic science category (McMillan, Narin, & Deeds, 2000), while most papers cited by clinical guidelines are from the two most applied categories (Grant, Cottrell, Cluzeau, & Fawcett, 2000).

When using the CHI system, two implicit assumptions must be acknowledged. First, all papers in a journal are assumed to be of the same RL regardless of their actual level. Second, journals are not allowed to change RL over time, even if the actual level does change. Both assumptions suggest that some fraction of the papers will be misclassified using the CHI system. It should also be mentioned that most of the CHI RL assignments are quite old; many are likely outdated.

Lewison and Paraje (2004) took the first step in addressing these deficiencies by creating their own method for assigning RL based on sets of "basic" and "applied" words. Their word sets were created using a combination of processing of word count lists derived from titles of papers in RL1 (applied) and RL4 (basic) journals, followed by manual fine tuning of those lists to produce sets of words with high discriminatory power. This method discriminates very well between RL1 and RL4, and does a reasonable job of discriminating RL2 and RL3 (which are interpolated). It has the advantage of being very simple and reproducible from the word sets. The greatest advantage, however, of any word-based approach to assigning RL is that it provides a standardized method that can be used with any paper, whether it appears in a journal with an assigned RL or not.

Cambrosio et al. (2006) then used the Lewison method in a practical study to show the change in the cancer landscape over time. They showed an increased linkage between the research levels from 1980 to 2000. In 1980, cancer research was divided into two groupings – RL1+RL2 and RL3+RL4 – that had little connection. By 2000, the landscape showed a much more smooth transition from RL1 to RL2+RL3 to RL4 with a much higher level of connectivity. This is important in that it shows that the potential for translation between levels is much higher today than it was decades ago. Translation, however, is much more effective when it occurs in individuals and teams than simply through journal-level pathways. Our study is thus designed to characterize individual researchers and their potential to bridge two or more levels.

Research level model

The first step in the study was to determine if we could develop a word-based classification system for RL that would be even more discriminatory than that developed by Lewison & Paraje (2004), particularly with respect to RL2 and RL3. Given that the focus for the study was biomedical research and researchers, we chose a corpus of biomedical articles that could be used to develop the classification system. Using Scopus data from 2010, PubMed data from 2010, and the UCSD journal classification system (Börner et al., 2012), we designated those UCSD disciplines where at least half of the Scopus articles were also available in PubMed as medical disciplines. The corpus was limited to journals (and their papers) from those medical disciplines with an assigned CHI research level. This resulted in a set of 425,345 papers from 2,183 journals.

Two different research level models were created from these data. First, a multinomial logistic regression model was created using title and abstract words as features. The 8,384 words that appeared more than 100 times in titles, and the 30,362 words that appeared more than 100 times in abstracts were used in a paper/feature matrix. The model was trained against the CHI research levels using binary matrix entries rather than actual counts; only papers containing two or more features were used. The resulting *Word Feature Model* (WFM) contained weights for each feature – research level combination.

Since the original CHI research levels were partially based on citation patterns, we created a second logistic model with eight features and papers as samples. The first four features for each paper are the research level probabilities calculated by the WFM. The second set of four features is comprised of the relative proportions of references published by journals associated with the various research levels. So, if a paper cites 1 paper from each research level, then the last four input feature values would be (0.25, 0.25, 0.25, 0.25). If a paper does not cite anything from a journal associated with a research level, then the last four input features are zeros. The resulting *Combined Feature Model* (CFM) is represented as a set of coefficients and intercepts. For each research level, a model has an intercept and a vector of positive/negative coefficients of length equal to the number of features. The probability of each research level is computed as a dot product between the feature vector of the paper and the coefficients, with the intercept being added as a coefficient.

Results

Research level calculations

The probability of belonging to each RL was calculated for all biomedical papers in Scopus for 1996–2011 by first calculating their weights using both the WFM and CFM. For each model, each article was then assigned to the single RL for which it had the highest probability. Table 2 shows that of those articles that had CHI research levels (#Art-CHI), the WFM assigns over 59% of them to that level while the CFM assigns nearly 66% of them to that level. Using references

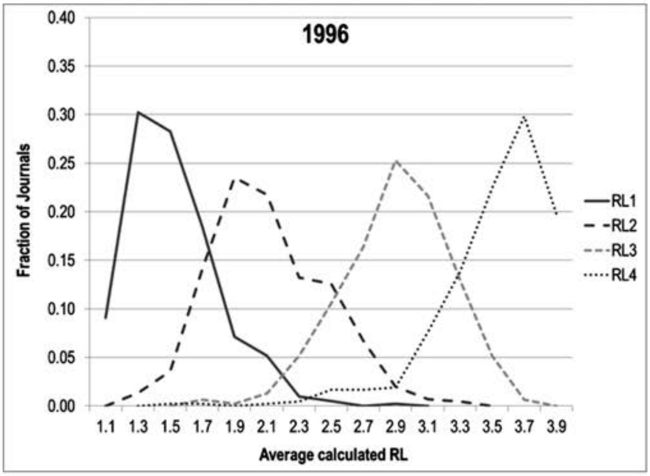
clearly has a positive effect on the match rate, increasing it by 6–7% in most years. The match rate based on words is roughly constant for most of the time period, with the exception of the year associated with the training data (2010), which is much higher. The positive effect of references is also diminished in early years since very little reference data from before 1996 was available to the CFM. Given its superiority over the WFM, the balance of the paper uses the CFM.

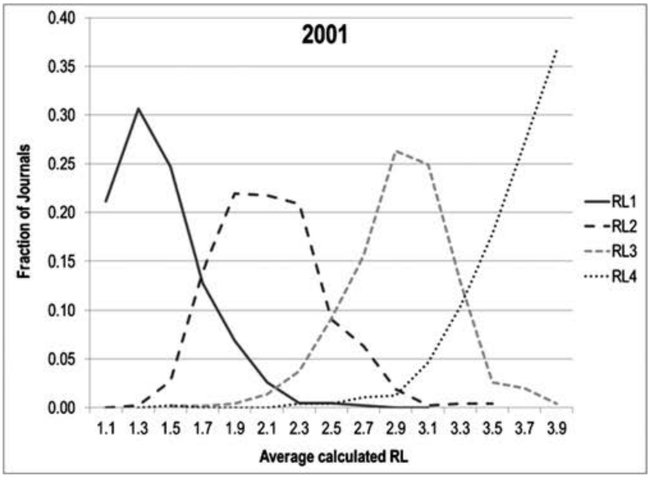
Table 2. Properties of the research level calculations at the article level.

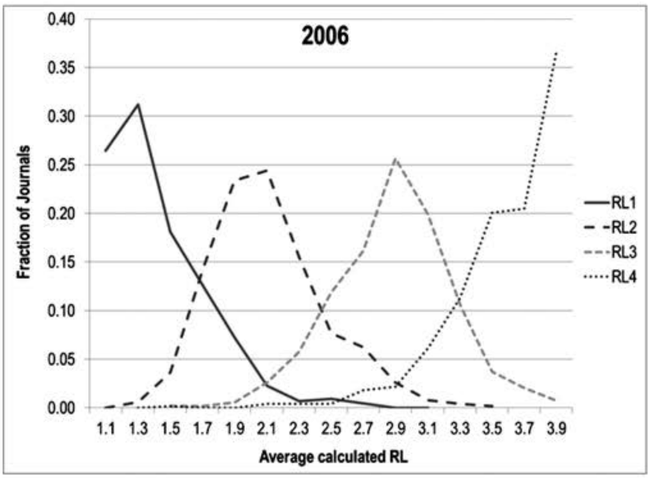
Year	#Jnl	#Jnl-CHI	#Art	#Art-CHI	WFM-%Match	CFM-%Match
1996	3,750	1,834	437,796	312,318	56.26%	57.54%
1997	3,870	1,850	455,263	319,971	57.15%	60.81%
1998	3,992	1,868	464,248	321,631	57.73%	63.12%
1999	4,162	1,898	473,366	322,805	57.67%	64.23%
2000	4,309	1,925	495,348	333,465	58.11%	64.89%
2001	4,546	1,946	511,124	335,475	58.77%	66.02%
2002	4,769	1,961	530,989	337,253	59.58%	66.80%
2003	4,906	1,982	555,579	349,337	59.97%	66.94%
2004	5,113	2,001	593,628	363,695	60.02%	66.90%
2005	5,407	2,018	638,190	379,409	58.87%	66.47%
2006	5,800	2,027	693,543	397,151	58.84%	66.72%
2007	6,027	2,033	734,091	405,138	59.57%	66.92%
2008	6,200	2,034	764,564	411,760	59.69%	66.59%
2009	6,197	2,032	780,959	412,537	59.51%	66.57%
2010	6,245	2,035	795,074	415,898	69.62%	72.37%
2011	6,041	2,018	830,901	433,164	59.56%	65.32%
Total			9,754,663	5,851,007	59.60%	65.73%

A calculation was done to compare calculated RL with the CHI research levels at the journal level. An average calculated RL was computed for each journal-year combination based on its article assignments. Each panel in Figure 1 shows four distinct peaks, one for each RL, which suggests that the model differentiates all four RL quite well across the entire time range.

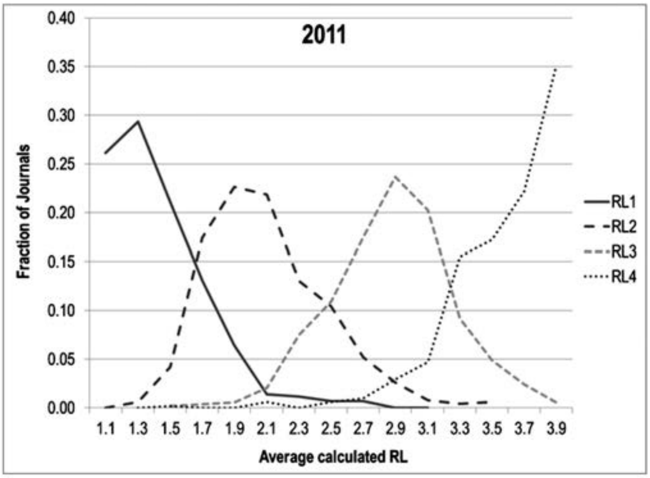
Figure 1. Fraction of journals as a function of calculated average RL for the four CHI journal research levels. Only journals with at least 20 papers in the given year are included.







… Continuation Figure 1



This model does a much better job of differentiating RL2 and RL3 than does the Lewison model, but is far more complex to calculate, and is far less transparent. The fact that our model shows distinct peaks for RL2 and RL3, while the Lewison model does not, suggests that RL2 and RL3 have their own specific language. This language should be explicitly modelled rather than interpolating between language associated with RL1 and RL4.

It is also interesting to note that while the articles in journals with CHI level assignments were roughly evenly split between categories (22.2%, 25.8%, 26.2%, 25.8% for RL1-RL4, respectively), assignment of RL to articles in non-CHI journals was strongly slanted toward applied research (42.1%, 33.0%, 14.9%, 10.0%). In many cases, journals were not assigned a research level by CHI (or were assigned a value of zero, meaning unknown) because they felt the journal could not be unambiguously assigned. The fact that the majority of the unassigned journals are assigned RL in the applied categories correlates well with this observation; unambiguously distinguishing between applied categories is certainly more difficult than distinguishing between basic research and any type of applied research.

Characterization of researchers

Using the results of the model – the calculated RL for nearly 10 million articles – we then characterized a set of researchers to learn if that characterization could be used to create an indicator of translational capacity. The assumption behind this work is that those researchers who publish nearly equal amounts of work in two RL are better situated to effect translation between those levels than are researchers who publish nearly exclusively in one level.

The disease area chosen for this first investigation was Alzheimer's Disease (AD). A set of 102,907 AD-related articles was identified in Scopus (1996–2011) using a query composed of AD terms. Of these articles, 83,808 were assigned an RL using our new model. The distribution of these

articles across RL (17.0%, 30.6%, 12.3%, 40.1% from RL1-RL4, respectively) was significantly different than the distribution across all articles (30.2%, 28.7%, 21.7%, 19.5%, respectively). We are curious about the seeming deficiency of RL3 in AD research, but will leave that to another study. Authors were obtained for all articles, and RL distributions were calculated for the top producing AD researchers.

Table 3. Research level distributions and indicators for 20 top producing researchers in Alzheimer's Disease. Bold font denotes the dominant RL.

Name/Inst	#RL1	#RL2	#RL3	#RL4	> 30%	TR	Pair
Mayeux, Richard (Columbia U.)	79	**381**	41	74	1	0.207	
Morris, John C. (U. Washington)	69	**382**	14	68	1	0.181	
Cummings, Jeffrey L. (Clev Clinic)	140	**289**	2	24	2	0.484	1–2
Trojanowski, John Q. (U. Penn)	4	175	61	**210**	2	0.833	2–4
Stern, Yaakov (Columbia U.)	85	**285**	26	31	1	0.298	
Winblad, Bengt (Karolinska Inst.)	59	**219**	21	123	1	0.562	
Petersen, Ronald C. (Mayo Inst.)	41	**302**	11	40	1	0.136	
Vellas, Bruno (CHU Toulouse)	126	**245**	10	12	2	0.514	1–2
Perry, George (UT San Antonio)	12	53	35	**291**	1	0.182	
Holtzman, David M. (Wash U. St. L.)	2	131	29	**224**	2	0.585	2–4
Butterfield, D. Allan (U. Kentucky)	2	21	13	**348**	1	0.060	
Blennow, Kaj (Goteborg U.)	17	**194**	38	135	2	0.696	2–4
Mattson, Mark P. (Nat. Inst. Aging)	3	14	15	**351**	1	0.043	
Markesbery, William R. (U. Kentucky)	10	106	27	**230**	1	0.461	
Masters, Colin L. (U. Melbourne)	6	109	18	**237**	1	0.460	
DeKosky, Steven T. (U. Virginia)	89	**146**	22	111	2	0.760	2–4
Bennett, David A. (Rush U.)	26	**265**	16	57	1	0.215	

… Continuation Table 3

Name/Inst	#RL1	#RL2	#RL3	#RL4	> 30%	TR	Pair
Soininen, Hilkka (U. East. Finland)	19	**202**	20	116	2	0.574	2–4
Masliah, Eliezer (UC San Diego)	4	51	23	**260**	1	0.196	
Frisoni, Giovanni B. (IRCCS)	53	**209**	9	65	1	0.311	

Table 3 shows that most of the top 20 AD researchers publish primarily in one RL. To identify those with high translational capacity, Table 3 also shows the number of levels containing at least 30% of a researcher's output (> 30%), and the ratio of outputs in second highest to highest RL for each researcher (TR). Seven of the 20 researchers have 30% of their output in two RL; the TR value for each of these researchers is above 0.5. The RL pair in which those seven researchers have high translational capacity is listed in the final column.

We propose that TR – the ratio of the number of papers in the second highest RL to the number of papers in the dominant RL – is a reasonable proxy for the translational capacity of a researcher within those two levels.

Conclusion

We have created a methodology to examine the translational capacity of individual researchers based on research level, and have proposed an indicator based on this approach. We note that this is a first investigation, and that it has only been done with a single disease area. Much work remains to be done to validate the proposed indicator. Yet, we also note that the research level code and calculations upon which it is based have provided robust results that strongly differentiate between all four research levels. Future investigations will look at the variability between disease areas, and will also attempt to validate this approach using expert knowledge. We also anticipate developing an extension to this approach that examines translational capacity at the team level.

References

Börner, K., Klavans, R., Patek, M., Zoss, A. M., Biberstine, J. R., Light, R. P., et al. (2012). Design and update of a classification system: The UCSD map of science. *PLoS ONE,* 7(7), e39464.

Boyack, K. W., & Klavans, R. (2011). Multiple dimensions of journal specificity: Why journals can't be assigned to disciplines. *13th International Conference of the International Society for Scientometrics and Informetrics*, 123–133.

Butler, L., Biglia, B., & Bourke, P. (1998). *Australian Biomedical Research: Funding acknowledgements and performance.*

Cambrosio, A., Keating, P., Mercier, S., Lewison, G., & Mogoutov, A. (2006). Mapping the emergence and development of translational cancer research. *European Journal of Cancer*, 42, 3140–3148.

Carpenter, M. P., Gibb, F., Harris, M., Irvine, J., Martin, B. R., & Narin, F. (1988). Bibliometric profiles for British academic institutions: An experiment to develop research output indicators. *Scientometrics*, 14(3–4), 213–233.

Grant, J., Cottrell, R., Cluzeau, F., & Fawcett, G. (2000). Evaluating 'payback' on biomedical research from papers cited in clinical guidelines: Applied bibliometric study. *British Medical Journal*, 320, 1107–1111.

Lewison, G., & Paraje, G. (2004). The classification of biomedical journals by research level. *Scientometrics*, 60(2), 145–157.

McAllister, P. R., & Narin, F. (1983). Characterization of the research papers of U.S. medical schools. *Journal of the American Society for Information Science,* 34(2), 123–131.

McMillan, G. S., Narin, F., & Deeds, D. L. (2000). An analysis of the critical role of public science in innovation: *The case of biotechnology*. *Research Policy,* 29, 1–8.

Narin, F., Pinski, G., & Gee, H. H. (1976). Structure of the biomedical literature. *Journal of the American Society for Information Science,* 27(1), 25–45.

Sung, N. S., Crowley, W. F. J., Genel, M., Salber, P., Sandy, L., Sherwood, L. M., et al. (2003). Central challenges facing the national clinical research enterprise. *Journal of the American Medical Association,* 289(10), 1278–1287.

Woolf, S. H. (2008). The meaning of translational research and why it matters. *Journal of the American Medical Association,* 299(2), 211–213.

i http://www.ncats.nih.gov/research/cts/cts.html

ii http://ec.europa.eu/research/health/medical-research/index_en.html

iii http://www.mrc.ac.uk/Ourresearch/ResearchInitiatives/Translationalresearch/index.htm

Science unfolding in technology: a cross-country analysis of scientific citations in patents[1]

Julie Callaert — Julie.Callaert@kuleuven.be / ECOOM / INCENTIM – Faculty of Business and Economics – KU Leuven, Waaistraat 6, 3000 Leuven (Belgium)

Jan-Bart Vervenne — Jan-Bart.Vervenne@kuleuven.be / ECOOM / INCENTIM – Faculty of Business and Economics – KU Leuven, Waaistraat 6, 3000 Leuven (Belgium),

Bart Van Looy — Bart.Vanlooy@kuleuven.be / ECOOM / INCENTIM – Faculty of Business and Economics – KU Leuven, Waaistraat 6, 3000 Leuven (Belgium) / IGS – University of Twente (Netherlands)

Abstract

The mapping of cross-country citation linkages between technology (applicants of citing patents) and science (authors of cited scientific articles) is highly useful for addressing questions on the sources of national technological development as drivers behind national innovative performance. Based on a methodology for matching non-patent references to the Web of Science on a large scale and in an automated manner, we perform an applied study based of such cross-country citation linkages. Our findings confirm the presence of a significant home advantage and a cross-border proximity effect for science cited in patents. At the same time, there is considerable country-variation in the extent of this home advantage. The results further show that characteristics of cross-country linkage patterns are related to technological performance on a national level. The presence of foreign linkages that complement home-based knowledge sourcing, as well as a broad geographical scope in citing foreign science, are positively related to technological performance. These leveraging effects of foreign orientation in citing science are however primarily relevant for the scientifically weaker countries.

Introduction and research objective

The concept of 'innovation systems' puts forward the interplay between different types of innovation actors (e.g. Freeman, 1994; Adams, 1990; Lundvall, 1992; Nelson, 1993; Mowery and Nelson, 1999; Baumol, 2002), including the importance of knowledge for innovative performance, growth and competitiveness. Empirical studies have attempted to evidence trends in science-technology relatedness, using various indicators to quantify the footprint of knowledge generating institutes in technology development and innovation. One indicator, which has gained increasing attention over the last decade, is the use of public science by firms as indicated by their patents referring to scientific publications (Narin et al., 1997; Verbeek et al., 2002; Van Looy et al., 2006) or citing academic patents (Henderson et al., 1998; Mowery et al., 2001; Veugelers et al., 2012). Although

1 This work was supported by the European Commission, DG Research.

some discussion remains about their actual meaning, there is convergence on the literature about the interpretation that non-patent references are indicative at least of relatedness between science and technology (Meyer, 2000; Tijssen et al., 2000; Sternitzke, 2009).

Besides indicators on the science-intensity of technologies that are based on the number of non-patent references in patented technology, great analytical potential can be unlocked by further identifying cited non-patent references in terms of authoring institutes and their geographic location, scientific disciplines and other bibliographic characteristics… The hereto required matching of cited non-patent references to bibliographic databases implies considerable methodological challenges, caused mainly by the non-systematic way in which non-patent references are registered in patent documents. Efforts in this respect have therefore remained mostly fragmented and targeted on smaller scale samples. Hence, insights into the basic mechanisms whereby science impacts innovation performance still suffers from a paucity of empirical evidence. Particularly lacking are quantitative indicators with wide coverage of institutes, countries and technologies over time.

Our contribution tries to fill this gap. Based on a methodology for matching non-patent references to the Web of Science on a large scale and in an automated manner, we perform an applied study based of cross-country citation flows between science and technology. The mapping of cross-country citation linkages between science (authors of the cited scientific article) and technology (applicants/inventors of the citing patent) is highly useful for addressing questions on the sources of national technological development as drivers behind national innovative performance. Several questions are relevant in this respect. We focus our study on the presence of a home effect in science-technology relatedness, whereby the following research questions are guiding:

(1) What are patterns of cross-country citation flows between science and patents and is there a home advantage or a geographic proximity effect in the citations from a country's corporate patents towards science?

(2) What are the relations between national-level characteristics of science-technology linkage patterns and national-level technological performance?

Data and Methodology

All patent-based data have been extracted from the PATSTAT Database (Autumn 2010 edition). Bibliographic data (scientific articles) are obtained from the ISI / Thompson Reuters Web of Science database (SCIE, SSCI, AHCI) and the ISI Proceeding databases (CPCI-S and CPCI-SSH).

The matching of non-patent references to scientific records from the Web of Science database proved a major challenge. The large scale of the data covered (matching millions of NPRs in

PATSTAT to millions of records in the Web of Science) implied that an automated process was required. As stated earlier, the development of such an automated matching algorithm implies considerable methodological challenges, caused mainly by the way in which non-patent references are registered in patent documents.

A first issue relates to the fact that not all cited non-patent references are scientific in the strict sense of the word. Callaert et al. (2006) found that about half of non-patent references refer to sources other than journal articles: industry-related documents (catalogues, manuals), newspapers, reference books or databases, gene/plant bank records and patent-related documents. Based on the method developed by Callaert et al. (2011), non-scientific NPRs were identified and excluded from the analysis. Restricting analyses and indicators the subset of real *scientific* NPRs provides a more accurate picture of science-technology linkage. At the same time, it enhances efficiency of automated matching procedures between non-patent references and scientific bibliographies, as it eliminates 'noise' among non-patent references, decreasing the volume of data that are implied in the matching algorithm with approximately 50%. Indeed, from a methodological-technical perspective, the large volume of references to be matched at both sides poses great challenges.

A second issue relates to the fact that non-patent references in patents are not formatted in a standardized way. Experimentation with several text-mining applications (a.o., a bag-of-words method and a field-based parsing method, see Magerman et al., 2010), revealed that a field-by-field matching methodology best served our purpose. The method uses seven pre-parsed fields from articles in the Web of Science that – in varying combinations – allow for the unique identification of individual scientific source documents. The considered fields are: publication year, first author's last name, journal title, volume, issue, beginning page and article title). Each field is matched separately to the NPR text strings in PATSTAT, whereby field-based match scores are calculated as 'the number of distinct terms in the WOS field that occur also in the NPR text-string, divided by the number of distinct terms in the WOS-field'. The identification of correct matches between NPRs and WoS articles is based on a set of filters that specify combinations of field-based match scores above certain thresholds. The definition of the filters is based on the fact that it requires multiple field matches to results in an overall NPR-WoS match, and that the value of some fields is more important than others for the matching process. The defined thresholds aim to balance recall (sufficiently low thresholds) with an accurate elimination of false positives (sufficiently high thresholds, enhancing precision). The developed methodology, applied to all non-patent references in PATSTAT for the period 1993–2009 succeeds in matching almost 70% of 'matchable' NPRs (i.e. references that fulfil a first set of threshold criteria for the defined match scores). The resulting matched NPR-WoS pairs served as the basis for the indicators and analyses developed for the remainder of this paper.

The indicators and analyses that will be presented next aim to shed light on the geographic dimension of technology citing science, by focusing on cross-country citation flows. We study the presence of a home advantage or geographic proximity effect in the citations from a country's

corporate patents towards scientific resources. For analysing cross-country flows of citing patents and cited scientific articles, applicant and author countries were identified respectively. Global coverage is assured by including OECD member countries, EU-27 (and EFTA) member states, EU candidate countries, BRIC countries and Taiwan (on citing and cited side). For documents allocated to more than 1 country (due to multiple applicants or authors), a full fractional counting scheme is used. At the patent side, all EPO, USPTO and PCT[i] corporate patents were considered with application years between 2000 and 2009. For scientific articles (cited), it is important to note that the available Web of Science data infrastructure was restricted to publications after 1990. Therefore, NPRs can only be matched to scientific publications from 1991 onwards.

Analyses and Results

Patterns of cross-country citation linkages between science and technology: a home advantage?

For mapping cross-country citation linkage patterns, citation matrices were developed that map countries of citing corporate patents (rows) and countries of origin of the cited scientific references (columns). Table 1 shows cross-country citation linkages between these countries, whereby the cell values represent relative intensities of citation linkages between the citing countries (rows) and the cited countries (columns). By using relative intensities rather than absolute numbers of citation links, size effects are corrected for, allowing for adequate inter-country comparison of citation patterns. The following formula was used to calculate these relative intensities: CCCij =

$$\frac{\text{\#citations of corporate citing country i to science cited country j} \,/\, \text{total \#citations of corporate citing country i}}{\text{citations to science cited country j} \,/\, \text{total corporate citations}}$$

A value higher than 1 represents a link between citing and cited country which is 'overrepresented'. The framed figures on the diagonal are of particular interest, as they indicate within-country citation flows which will be used to evaluate a home advantage in citation patterns.

Table 1. Relative intensities of cross-country citation linkages.

CITING CORPORATE PATENTS \ CITED SCIENTIFIC DOCUMENTS	AT	AU	BE	BG	BR	CA	CH	CN	CY	CZ	DE	DK	ES	FI	FR	GB	GR	HU	IE	IL	IN	IS	IT	JP	KR	LI	NL	NO	NZ	PL	PT	RU	SE	SI	TW	US	ZA
AT	6,83	0,98	0,80	0,50	0,72	0,82	2,21	0,80	1,72	0,96	1,15	0,92	1,12	1,05	0,77	0,99	0,91	1,76	0,86	0,44	0,50	1,04	1,15	0,72	0,46		0,92	0,71	1,31	0,64	1,20	0,99	0,99	1,94	0,47	0,90	0,97
AU	1,18	5,07	0,74	1,40	1,51	0,89	0,93	0,97	1,08	0,82	0,91	1,04	0,85	1,06	0,90	1,03	0,70	0,97	1,11	0,80	0,86	0,97	0,88	0,79	0,54		0,97	1,15	2,08	1,02	0,75	0,92	1,07	0,27	0,50	0,89	1,09
BE	0,97	1,00	3,62	0,91	1,33	0,94	0,99	1,04	1,80	0,96	0,93	0,91	1,24	0,82	1,27	1,20	0,95	1,38	0,79	0,74	0,73	0,27	1,12	0,75	0,60		1,34	0,88	0,83	1,10	0,92	0,98	0,94	0,23	0,51	0,84	1,48
BG	0,84	0,77	3,28			0,78	0,81	1,41		3,24	1,10	2,00	1,42		1,48	1,11		5,02		1,09			1,74	0,82	0,41		2,19	2,91		1,59			1,01			0,59	
BR	0,71	0,81	0,99		27,84	0,66	0,69	0,79			0,84	0,28	0,80		1,46	0,94	2,38	2,12		0,23	0,36		1,47	0,89	0,35		0,58	0,62	1,71	0,67	1,45		1,00		0,28	0,84	3,05
CA	0,97	1,12	1,07	0,89	1,15	2,57	0,90	0,95	0,41	0,62	0,89	1,05	1,15	0,85	0,98	1,01	0,71	0,68	0,65	0,91	0,84	0,37	0,95	0,76	0,77		1,01	1,10	0,91	0,69	0,59	1,43	0,95	0,77	0,88	0,90	1,87
CH	1,41	1,09	0,84	0,99	0,87	0,90	2,41	0,90	0,36	1,14	1,07	1,09	1,04	0,94	0,95	1,10	0,77	1,00	1,14	1,04	0,87	0,87	1,15	0,80	0,61	0,89	0,92	0,81	1,17	0,93	0,68	0,82	1,06	0,95	0,73	0,92	1,41
CN	0,99	0,85	1,07	0,91	0,93	0,81	0,58	4,24		1,29	0,77	0,76	0,61	0,61	0,86	1,06	1,02	0,37	0,79	0,76	1,21	1,90	1,02	0,96	1,01		0,51	1,23	1,20	0,39	0,17	0,54	0,52	0,79	1,41	1,13	1,07
CY		0,67	0,83			0,51	0,24	3,73			1,55	1,18	1,25	1,46	1,16	0,69	6,63			1,93	1,49		0,88	0,73	0,37		2,17			2,81	3,02		1,19		1,17	0,79	
CZ	0,70	0,64	0,39		1,32	0,49	0,46	2,77		32,79	0,83		1,59		0,56	0,56	1,58			0,46	2,14		0,42	1,16	0,35		0,46			2,68	2,88	0,83	0,85	6,71	1,12	0,97	18,21
DE	1,35	0,91	0,94	1,09	0,66	0,81	0,93	1,02	1,01	1,25	1,89	0,94	1,00	0,79	1,00	1,01	0,72	1,06	0,89	0,62	0,98	0,91	0,92	0,86	0,71	1,14	1,02	0,66	0,83	1,01	1,06	1,06	0,87	0,87	0,66	0,94	0,94
DK	1,18	0,88	1,33	1,23	0,71	0,91	0,98	0,73		1,47	0,96	7,18	1,12	1,30	0,99	1,07	0,65	0,94	1,53	0,58	1,11	1,53	0,94	0,73	0,77		1,29	1,18	0,94	1,40	1,82	0,49	1,34	0,64	0,57	0,77	1,82
ES	0,95	0,84	0,78	0,57	1,26	0,64	0,81	1,20		0,90	0,72	0,70	6,73	1,02	1,18	1,08	0,99	0,85	0,61	0,86	2,32	1,19	1,54	0,72	0,83		1,22	0,63	1,25	1,33	1,16	0,52	0,77	0,49	0,72	0,75	1,56
FI	0,83	0,78	0,95	0,47	0,88	1,00	0,70	0,99	1,62	0,58	0,86	0,68	0,71	9,58	0,66	0,82	2,10	2,49	1,06	0,60	0,75	2,44	0,81	0,59	1,06		0,83	1,11	0,87	0,48	0,35	0,85	1,35		0,98	1,09	1,10
FR	0,71	0,92	1,10	1,33	0,90	0,84	0,79	1,16	0,63	0,87	0,88	0,80	1,12	0,92	2,33	0,96	1,29	0,92	0,85	0,74	1,15	0,57	1,14	0,83	0,87	1,56	0,82	1,65	0,98	1,02	1,14	0,76	0,72	0,51	0,87	0,96	0,82
GB	0,89	0,96	0,95	0,75	0,89	0,90	0,97	0,81		1,39	0,94	0,87	1,09	0,75	1,00	1,88	0,94	0,98	0,97	0,67	1,01	0,61	1,04	0,81	0,73		0,96	0,85	1,19	0,98	1,01	0,92	0,86	1,01	0,72	0,95	1,68
GR	0,81	0,37	0,91		4,58	1,70	0,53	0,92			0,54	0,65	0,92		0,64	0,98	32,93	4,89			2,47		2,66	0,36	2,42		0,53					1,93		7,76		0,77	
HU	1,15	0,63	0,91		1,30	0,48	1,35				0,79	1,11	1,18	1,37	0,82	0,92	0,52	21,53	1,10	0,91	1,87		1,65	0,78	0,80		0,83	1,21				0,82	0,65			1,10	4,98
IE	1,57	0,48	1,89	0,64	0,87	0,97	0,98	1,11		1,58	0,90	0,88	1,05	1,32	0,77	1,15	0,78	1,57	12,18	0,76	1,65	2,68	1,20	0,62	0,87		0,72	1,72	0,42	1,44	0,72	1,31	1,08		0,79	0,91	0,50
IL	0,92	0,99	1,13	0,16	0,91	1,06	0,89	0,72	3,36	0,80	0,90	0,87	1,13	0,54	0,89	0,96	1,22	1,01	0,56	4,37	1,41	1,01	1,22	0,92	0,95		0,89	0,71	1,06	0,72	0,30	1,18	0,97	0,70	0,93	0,93	2,65
IN	0,93	0,72	1,08	1,28	1,20	0,87	0,81	1,34		2,60	0,70	1,26	1,34	0,81	0,90	1,18	0,33	0,70	0,83	0,69	8,28	4,69	0,91	0,88	1,17		0,78	0,56	1,06	1,11	0,72	0,52	0,74	2,22	0,81	0,87	5,40
IS	0,99	0,91	1,12	4,39		0,69	1,10	0,67		3,09	0,81	2,23	1,35	1,58	0,83	1,14	2,24	2,39	0,95	0,13	0,40	201,37	1,36	0,50	0,30		0,85	2,08	0,48	0,38		0,71	1,61			0,90	0,86
IT	0,70	0,76	0,89	1,05	0,66	0,85	0,90	0,91		0,95	0,83	0,62	1,10	0,78	0,95	0,97	1,04	0,90	1,16	0,85	1,32	1,25	3,84	0,88	0,83		0,96	1,25	0,82	1,06	1,22	0,85	0,75	0,65	0,62	0,89	2,04
JP	0,60	0,65	0,64	0,78	0,71	0,73	0,79	1,07	0,35	0,78	0,88	0,64	0,79	0,64	0,84	0,80	0,84	0,58	1,05	0,59	0,78	0,63	0,68	2,27	1,20	0,21	0,76	0,56	0,48	0,87	0,92	0,92	0,63	0,63	1,03	1,04	0,68
KR	0,46	0,61	0,65	0,48	0,38	0,75	0,77	1,53	0,67	0,88	0,94	0,62	0,66	1,08	0,72	0,73	0,72	0,71	0,83	0,70	0,74	0,40	0,68	1,20	5,22		0,86	1,11	0,48	0,46	0,57	0,92	0,69	0,91	1,56	1,05	0,53
LI	0,71	0,65	1,86			0,72	1,00	1,87			1,15	0,95	0,80	0,47	1,08	0,73	3,20		1,69	0,78		16,34	1,13	1,24	0,24	88,79	0,78	1,65			0,97		0,96		1,50	1,05	
NL	0,99	0,79	1,00	0,44	0,82	0,75	0,92	0,85	1,52	1,06	1,00	0,65	0,92	1,16	1,04	0,97	1,02	0,76	0,99	0,69	0,88	0,69	0,80	0,89	1,02	0,31	2,87	0,88	0,90	0,95	1,49	0,69	0,69	0,47	0,73	1,03	1,18
NO	0,52	1,02	0,67	2,29	1,17	1,17	0,54	0,99		1,41	0,88	1,29	1,12	1,18	0,82	1,30	0,58	0,47	0,74	0,61	0,95	14,32	1,00	0,67	0,46		1,10	15,29	0,25	1,38	1,28	0,61	1,17	3,47	0,78	0,85	1,34
NZ	0,96	2,07	0,93		1,20	1,19	0,70	0,47		2,23	0,75	1,94	1,12	1,39	0,91	0,94	0,29	0,19	1,21	0,67	0,77		0,97	0,75	0,44		1,06	0,89	29,57	0,85	1,05	0,30	1,11	1,22	0,35	0,75	0,27
PL	0,76	1,04	0,21			0,70	0,61	1,71		2,94	1,10	0,91	0,64	0,75	0,82	0,56	2,56	1,14			0,38		0,79	1,08	1,88		0,37			25,24	1,55		0,46		0,30	1,03	1,63
PT		0,99	0,81		0,68	0,84	0,82	1,02			0,86	0,87	1,84	3,22	1,50	0,96			0,86	0,71	2,19	8,31	1,61	0,67	0,54		0,59	1,26	2,62	0,69	34,07	0,86	1,02	3,45	0,29	0,76	1,56
RU	0,45	0,41	0,76		0,84	0,63	0,58	2,02		1,75	1,12		1,02	0,89	0,71	1,08			1,07	0,88	1,82		0,80	0,79	1,79		0,74	0,78	1,09			15,97	0,36			0,87	
SE	1,03	0,91	1,08	0,55	0,62	0,82	0,92	0,88		1,06	0,86	1,05	1,22	1,22	0,96	1,14	0,78	0,87	0,98	0,60	1,02	0,57	1,01	0,76	0,69	0,52	0,79	0,88	1,50	0,93	0,61	0,73	3,84	0,55	0,78	0,96	0,82
SI	2,53	0,19	0,47		0,79	0,39	0,55				0,50		0,48	0,42	1,17	0,73		1,27	2,00		4,69	19,38	1,13	0,97	1,26			0,73	1,02	0,80	5,18		1,70	100,51	0,33	1,08	5,45
TW	0,56	0,50	1,39	0,34	0,47	0,54	0,30	1,44		0,12	0,95	0,30	0,88	0,34	0,54	0,78	0,77	0,37	0,37	1,06	3,02		1,44	1,22	2,32		0,62	0,54	0,08	0,41	0,38	0,33	0,38		5,71	1,11	0,13
US	1,00	1,05	1,03	1,10	1,13	1,07	1,02	0,98	1,28	0,98	0,94	1,01	0,97	1,03	0,98	0,98	1,08	1,08	1,00	1,18	0,98	0,84	1,01	0,85	0,98	1,34	0,99	1,06	1,01	1,04	1,02	1,06	1,06	1,14	1,09	1,03	0,86
ZA	0,73	0,84	0,41	4,05	2,07	0,94	0,12	0,62		1,43	1,26	1,76	1,66		0,72	1,42		2,21	1,75		1,86		1,42	1,00	0,91		0,72	0,64				0,87	0,74			0,78	33,24

A visual inspection of the matrix reveals that, overall, countries cite their own science more intensively than science from foreign countries (cell values on the diagonal representing higher intensities). At the same time, it becomes apparent that the extent of such a 'home advantage' differs between countries. More specifically, countries where the within-country citation intensity is lowest are US, UK, DE (index below 2) and JP, FR, CH, CA, NL (index below 3). Most of these countries are at the same time technologically strong, suggesting that reliance on foreign science leverages technological performance. A more thorough analysis of the implied premise is conducted below, where we further analyse the relation between characteristics of ST linkage patterns and technological performance on a national level. When considering citation linkages beyond country borders, the matrix suggests that geographic proximity matters. Examples of groups of neighbouring countries that display relatively strong citation linkage intensities include: Austria, Switzerland, Germany – Denmark, Norway, Sweden – Sweden, Finland – France, Belgium.

To statistically verify the above observations, an analysis of covariance (ANCOVA) was performed. The dependent variable is the index of relative citation intensity between citing and cited country. It is transformed logarithmically to meet distributional requirements for the model. Independent variables are: home advantage; geographic distance between citing and cited country and difference in scientific productivity between citing and cited country. The home advantage variable is represented by a dummy for 'within-country' (1) versus 'foreign' (0) citation links. If a home advantage is present, a positive relation between 'home' and citation intensity is expected. Based on the observation in the matrix (table 2) that the home advantage appears more outspoken for some countries, an additional interaction effect was introduced between citing country and home advantage. Turning our attention to cross-border citation patterns, a geographic distance measure is included to assess the strength of a proximity effect in technology-science citation patterns. The measure has been adopted from the CEPII[ii] GeoDIst data files (see http://www.cepii.fr/anglaisgraph/bdd/distances.htm), which contain measures of bilateral distances (in kms) for most countries of the world. The measure introduced in our model is a weighted distance measure (transformed logarithmically to meet distributional requirements for the model). It calculates distances between two countries based on bilateral distances between their biggest cities, whereby the inter-city distances are weighted by the share of the city in the overall country's population. If a geographic proximity effect is present in science-technology citation linkages, then a negative relation between this distance measure and citation linkage intensity is expected. The model includes controls for country specificities (citing and cited country). Domain-specificities are accounted for by introducing technological domains, classified into 35 classes according to the classification developed by ISI-Fraunhofer (Schmoch, 2008). Results are presented in table 2.

Table 2. Ancova analysis: presence of home advantage in citation linkage patterns between corporate patents and scientific references.

Dependent Variable: Relative citation intensity between citing and cited country						
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Parameter Estimate (B)**
Corrected Model	3004,044a	136	22,089	116,576	,000	
Intercept	301,668	1	301,668	1592,107	,000	2,738
Home (1/0)	358,625	1	358,625	1892,709	,000	1,530
Geographic distance between citing and cited country	17,118	1	17,118	90,343	,000	-,046
Citing country * Home	217,972	32	6,812	35,950	,000	
Technology Domain (FhG35)	614,326	34	18,068	95,359	,000	
Citing Country	840,003	34	24,706	130,390	,000	
Cited Country	556,623	34	16,371	86,402	,000	
Error	3219,408	16991	,189			
Total	19716,133	17128				
Corrected Total	6223,452	17127				
a. R Squared = ,483 (Adjusted R Squared = ,479)						

Controlling for significant domain effects in citation intensities, the results confirm the presence of a home advantage in science-technology linkage patterns. Technological development in a country indeed seems to be linked primarily to the own national science base, a finding which supports the existence of the national innovation 'systems' notion. In addition, the negative effect of the geographic distance measure shows that proximity matters for foreign science-technology linkages. Moreover, the significant interaction effect between home and citing country confirms that the home effect is not equally outspoken for all countries. This is further illustrated in Figure 1, which compares the within-country citation intensity to the average citation intensity with foreign countries, for each citing country.

Figure 1. Comparison of home to foreign citation intensity.

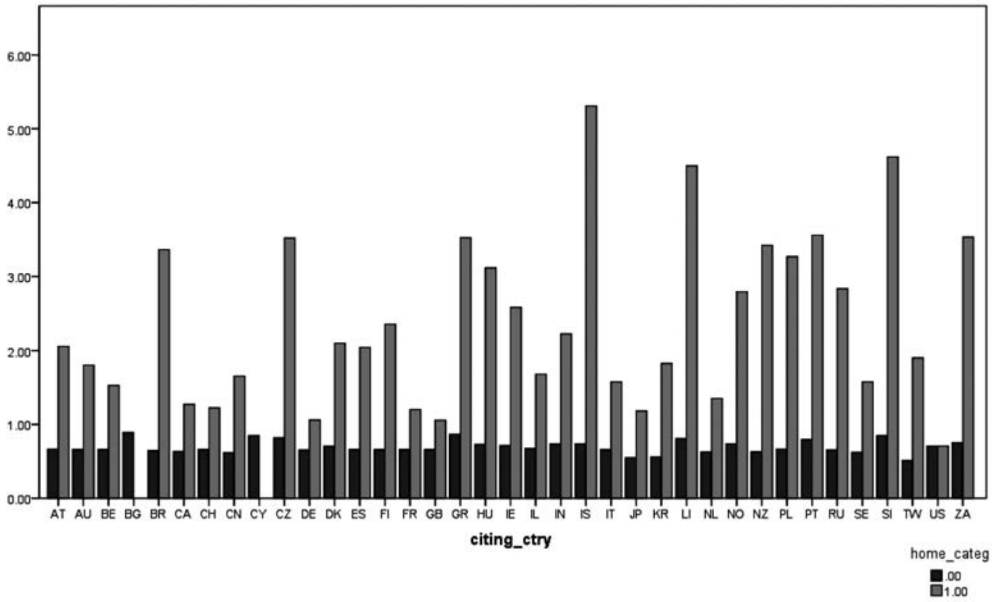


Figure 1 shows that, although the overall tendency clearly shows higher within-country citation intensities, this difference between home and foreign is less outspoken for countries like US, France, UK, Japan and Germany. Most of these countries are at the same time technologically strong, suggesting that reliance on foreign science leverages technological performance. A more thorough analysis of the implied premise is presented in the following analytical section.

National-level relation between science-technology linkage patterns and technological performance.

In this section, the relationship between characteristics of science-technology patterns and technological performance is analysed on a national level. National technological performance is measured by a country's patent volume per capita. For measuring characteristics of ST linkage patterns, a number of national-level variables have been developed, based on the available cross-country linkage characteristics. A first characteristic concerns the extent of a country's 'home advantage' in citing science, i.e. the extent to which a country's patents contain references to its own scientific articles versus references to scientific articles from foreign countries. The measure used is the ratio between 'within country citation intensity' (CCC_home, see the green bars in Figure 1) and the average citation intensity between the source country and cited foreign countries (CCC_foreign, represented by the blue bars in Figure 1). Second, as a complement to the home advantage variable, a Herfindahl-based measure was calculated that represents the concentration of foreign citations over cited countries. The Herfindahl index is the sum of the squares of

the (received) citations shares of all cited foreign countries, where the citation shares are expressed as fractions of the total number of citations to foreign countries. The result is proportional to the average citation share, weighted by citation share. As such, it can range from 0 to 1, moving from a large number of foreign countries that each receive a small number of citations (maximum geographic breadth), to a single foreign country that receives all citations (maximum concentration). In other words, a higher Herfindahl index indicates that science is cited from a smaller, more concentrated set of countries (the more geographically focused foreign citing patterns). A lower Herfindahl index indicates a geographically broader pattern of foreign citations.

Several control variables are added to the model. By introducing technological domains (Fraunhofer classification in 35 domains), we control for domain-specific effects. Moreover, several characteristics of national science and technology textures are considered that may be related to national technological performance. 'Scientific Output per 1000 Capita' represents the country's productivity in scientific research, measured by the number of scientific articles in the Web of Science for the period 1991–2009, and normalized for country size. 'Number of SNPRs' represents the science-intensity (number of scientific citations by patent) of a country's patent portfolio, broken down by technological domain (EPO, US, PCT patents, 2000–2009). The number of university patents is another indicator of national science-technology relatedness, representing the extent to which universities contribute to national technology development (EPO, US, PCT patents involving at least one university applicant[iii], 2000–2009).

The effects of science-technology linkage pattern characteristics (extent of home advantage and Herfindahl concentration index of foreign citations) may be interrelated with the scientific capacity from the home country. One could e.g. assume that scientifically weaker countries require stronger reliance on foreign science to keep their technology development up to pace. In order to test this, interaction effects between the extent of the home advantage and the Herfindahl index on the one hand and scientific productivity of the citing country on the other hand are additionally introduced into the model. The results of the ANCOVA analysis, along with estimates of the parameters (B), are presented in table 3.

Table 3. Ancova – Relation between national-level characteristics of ST linkage patterns and national technological performance[iv]

ANCOVA – Tests of Between-Subjects Effects						
Dependent Variable: Number of Patents per mio capita (Citing Country)						
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Parameter Estimate (B)**
Corrected Model	2290,597a	41	55,868	97,261	,000	
Intercept	7,185	1	7,185	12,509	,000	1,544
Scientific Output per 1000 Capita (Citing Country)	98,794	1	98,794	171,991	,000	2,362
Number of university patents per mio capita (Citing Country)	22,371	1	22,371	38,945	,000	,259
Number of SNPRs (Citing Country)	83,244	1	83,244	144,919	,000	,226
Ratio Home versus Foreign citation intensity	4,039	1	4,039	7,032	,008	-,156
Herfindahl Foreign Citations	8,354	1	8,354	14,544	,000	-1,330
Technology domain (FhG35)	299,064	34	8,796	15,313	,000	
Scientific Output per 1000 Capita (Citing Country)* Herfindahl Foreign Citations	5,744	1	5,744	10,000	,002	1,528
Scientific Output per 1000 Capita (Citing Country)* Ratio Home versus Foreign citation intensity	4,661	1	4,661	8,114	,004	,231
Error	558,333	972	,574			
Total	19157,226	1014				
Corrected Total	2848,930	1013				
a. R Squared = ,804 (Adjusted R Squared = ,796)						

Controlling for significant effects of technology domains, the results confirm that several characteristics of national science-technology systems are related to national technological performance. The relation between national scientific productivity and national technological performance is most outspoken. As for measures of national science-technology relatedness, the number of university patents is positively related to national technological output. In addition, the volume of SNPRs present in the national technology portfolio, as a second measure of national-level science-technology relatedness, shows a significant positive relation with technological performance. The positive relation between national-level measures of science-technology interaction on the one hand and national technological performance on the other hand confirms previous findings (see e.g. Van Looy et al., 2006).

Turning our attention to the indicators of cross-country linkage characteristics, the results show a significant relation between the extent of the home advantage and national technological performance. The main effect is negative, signalling that – overall – more reliance on foreign science is beneficial for national technology development activities. At the same time, the interaction effect with national-level scientific output is significant, implying that the relation between home advantage and national technological productivity depends on the level of scientific capabilities of the citing country. The results for the Herfindahl index suggest the appropriate modality of foreign citation patterns. The Herfindahl index for geographic concentration in foreign citations is negative and significant. This means that countries of which the cited science base is spread over more (foreign) countries are the technologically more performing countries. In other words, a sufficient degree of geographic diversity in cited foreign scientific prior art appears to more beneficial than a science base that is concentrated in fewer foreign countries, in terms of leveraging national technological performance. At the same time, the interaction effect between the Herfindahl index and scientific capacity of the citing country is significant. Hence, the relation between the Herfindahl index of geographic concentration in foreign citations on the one hand and national technological productivity on the other hand depends on the level of scientific capabilities of the citing country.

For a transparent interpretation of these interaction effects, the analyses were split up in two groups: countries in the lower 50% percentile of scientific productivity versus countries in the upper 50% percentile. The results of this split analysis are presented in table 4. It becomes clear that the positive effect of foreign scientific citations and a broad geographic scope in the origin of these foreign citations on national technological performance is valid for countries with a smaller home science base. For countries with large scientific capabilities, these leveraging effects of foreign and broad science citation patterns appear not to be decisive for national technological performance.

Table 4. Ancova – Relation between national-level characteristics of ST linkage patterns and national technological performance, split by level of scientific performance (Low versus High)

Dependent Variable: Number of Patents per mio capita (Citing Country)												
	Low Scientific Productivity						High Scientific Productivity					
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Parameter Estimate (B)**	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Parameter Estimate (B)**
Corrected Model	1184,370a	39	30,368	51,652	,000		283,693a	39	7,274	22,809	,000	
Intercept	,500	1	,500	,850	,357	1,132	7,146	1	7,146	22,408	,000	1,971
Scientific Output per 1000 Capita (Citing Country)	234,845	1	234,845	399,438	,000	3,702	53,635	1	53,635	168,178	,000	2,645
Number of university patents per mio capita (Citing Country)	45,655	1	45,655	77,653	,000	,773	,384	1	,384	1,204	,273	,043
Number of SNPRs (Citing Country)	20,956	1	20,956	35,644	,000	,180	14,160	1	14,160	44,399	,000	,135
Ratio Home versus Foreign citation intensity	3,277	1	3,277	5,574	,019	-,102	,001	1	,001	,003	,958	,002
Herfindahl Foreign Citations	3,327	1	3,327	5,659	,018	-,619	,074	1	,074	,232	,630	-,134
Technology domain (FhG35)	149,561	34	4,399	7,482	,000		159,305	34	4,685	14,692	,000	
Error	267,513	455	,588				147,978	464	,319			
Total	6009,784	495					12961,258	504				
Corrected Total	1451,882	494					431,671	503				
	a. R Squared = ,816 (Adjusted R Squared = ,800)						a. R Squared = ,657 (Adjusted R Squared = ,628)					

To summarize, the analysis confirms that technologically strong countries are characterized by high levels of scientific productivity and science-technology relatedness (measured by university patenting and scientific prior art in patent documents). Indeed, effects are largely driven by the scale of a country's scientific and technological activities. In addition, it is shown that characteristics of cross-country linkage patterns are related to technological performance on a national level. The presence of foreign linkage patterns to complement home-based knowledge sourcing, as well as a broad geographical scope in citing foreign science are positively related to technological performance. Moreover, significant interaction effects reveal that leveraging effects of this foreign orientation in citing science are primarily relevant for scientifically weaker countries. Countries that have a large own science base at their disposal, appear much less dependent on foreign sourcing and a broad geographic scope for enhancing or maintaining their technological performance.

Conclusions

Much of the empirical work on national innovation systems over the last decades confirms an intensification of the contribution of scientific knowledge to innovative performance. At the same time, there is a strong suggestion of heterogeneity: some institutions and/or some countries appear much more active in academic technology creation and science-technology linkage than others (Debackere & Veugelers, 2005; Van Looy et al., 2011).

We consider cross-country linkages, covering major world regions, between citing technology (patents) and cited science (articles), based on the results of a matching methodology between scientific references cited as prior art in patent documents on the one hand and individual articles in the Web of Science on the other hand. A mapping of science-technology citation linkages between countries is highly relevant for addressing questions on the sources of national technological development as drivers behind a national innovative performance.

First, our findings confirm the presence of a significant home advantage and a cross-border proximity effect for science cited in patents. The extent of the home advantage varies between countries: for several countries that are traditionally known as technologically strong (DE, FR, UK,…) and especially for the US, the home advantage is clearly less outspoken. The presence of a home advantage implies that countries can rely on their own scientific knowledge base for advancing their technological development and innovative activities, hence providing support for the existence of the national innovation 'systems' notion. At the same time, the findings show that this home advantage is complemented with science-technology citation linkages that cross national borders and often even occurring on a global scale. These findings are in line with a previous study, conducted by Veugelers et al. (2012) in which cross-country citation patterns between corporate (citing) patents and academic (cited) patents were analysed.

Second, the results confirm that characteristics of cross-country linkage patterns are related to technological performance on a national level. The presence of foreign linkages that complement home-based knowledge sourcing, as well as a broad geographical scope in citing foreign science, are positively related to technological performance. These leveraging effects of foreign orientation in citing science are however primarily relevant for the scientifically weaker countries.

So, although the relevance of science within today's innovation systems is widely recognized, the actual translation of scientific knowledge into usable technological developments – and the role enacted by universities – is not straightforward and varies greatly between countries. At the same time, for enhancing and maintaining technological performance, sufficient levels of linkage to science appear imperative. The modalities of this linkage between technology and science matter, whereby a broader geographical scope in scientific prior art is beneficial for leveraging technological performance.

References

Adams, J. (1990). Fundamental stocks of knowledge and productivity growth. *Journal of Political Economy*, 98, 673–702.

Baumol, W.J. (2002). *The Free-Market Innovation Machine: Analyzing the Growth Miracle of Capitalism*. Princeton University Press.

Callaert, J., Grouwels, J., & Van Looy, B. (2011). Delineating the scientific footprint in technology: Identifying science within non-patent references. *Scientometrics*, 91 (2), 383–398.

Callaert, J., Van Looy, B., Verbeek, A., Debackere, K., & Thijs, B. (2006). Traces of Prior Art: An analysis of non-patent references found in patent documents. *Scientometrics*, 69 (1), 3–20.

Debackere, K., & Veugelers, R. (2005). The role of academic technology transfer organizations in improving industry science links. *Research Policy*, 34(3), 321–342.

EUROSTAT (2011). Patent Statistics at Eurostat: Methods for Regionalisation, Sector Allocation and Name Harmonisation. *Methodologies & Working papers*. Publications Office of the European Union, Luxembourg, ISBN 978-92-79-20237-7.

Freeman, C. (1994). The economics of technical change: A critical survey article. *Cambridge Journal of Economics*, 18(5), 463–514.

Henderson, R., Jaffe; A.B., & Trajtenberg, M. (1998). Universities as a source of commercial technology: A detailed analysis of university patenting, 1965–1988. *Review of Economics and Statistics*, 80(1): 119–127.

Lundvall, B.-Å. (1992). *National Systems of Innovation*. London: Pinter.

Magerman, T., Van Looy, B., & Song, X. (2010). Exploring the feasibility and accuracy of Latent Semantic Analysis based text mining techniques to detect similarity between patent documents and scientific publications. *Scientometrics*, 82(2), 289–306.

Meyer, M. (2000). Patent citations in a novel field of technology: What can they tell about interactions between emerging communities of science and technology. *Scientometrics*, 48 (2), 151–178.

Mowery, D.C., & Nelson, R.R. (1999). *Sources of Industrial Leadership*. Cambridge: Cambridge University Press.

Mowery, D.C., Nelson R.R., Sampat B.N., & Ziedonis A. (2001). The growth of patenting and licensing by US universities: An assessment of the effects of the Bayh-Dole Act of 1980. *Research Policy*, 30(1), 99–119.

Narin, F., Hamilton, K.S., & Olivastro, D. (1997). The increasing linkage between US technology and science. *Research Policy*, 26, 317–330.

Nelson, R.R. (1993). *National Innovation Systems*. New York: Oxford University Press.

Schmoch, U. (2008). Concept of a technology classification for country comparisons. *Final report to the world intellectual property organization*, WIPO.

Sternitzke, C. (2009). Patents and publications as sources of novel and inventive knowledge. *Scientometrics*, 79, 551–561.

Tijssen, R.J.W., Buter, R.K., & Van Leeuwen, T.N. (2000). Technological relevance of science: validation and analysis of citation linkages between patents and research papers. *Scientometrics*, 47, 389–412.

Van Looy, B., Debackere, K., Callaert, J., Tijssen, R., & van Leeuwen, T. (2006). National scientific capabilities and technological performance: An exploration of emerging industrial relevant research domains. *Scientometrics*, 66(2): 295–310.

Van Looy, B., Du Plessis, M., Callaert, J., Meyer, M., & Debackere, K. (2011). The impact of legislative framework conditions on the technological performance of European Universities: An empirical assessment. *Paper presented at the 13th ISSI Conference 2011*, Durban, South-Africa.

Verbeek, A., Debackere, K., Luwel, M., & Andries, P. (2002). Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics*, 54(3), 399–420.

Veugelers R., Callaert J., Song X. & Van Looy B. (2012). The participation of entrepreneurial universities in technology development: Do creation and use coincide? An empirical investigation on the level of national innovation systems. Economics of Innovation and New Technology, 21 (5–6), 445–472.

i The identification of corporate patents was based on the sector allocation methodology developed at ECOOM – KU Leuven (Du Plessis et al., in: EUROSTAT 2011).

ii French research center in international economics; See: www.cepii.fr

iii Identification of university patents was based on the sector allocation methodology, developed at ECOOM – KU Leuven (see EUROSTAT, 2011).

iv All continuous variables have been transformed logarithmically for better fitting distributional characteristics. For patents with multiple applicants from different countries, a full-fractional counting scheme is used (whereby each represented country is counted as 1). For patents belonging to more than 1 Fraunhofer domain, a full counting scheme is used.

Quantitative Analysis of Technology Futures: A conceptual framework for positioning FTA techniques in policy appraisal

Tommaso Ciarli t.ciarli@sussex.ac.uk / SPRU, Science and Technology Policy Research, University of Sussex, Brighton

Alex Coad a.coad@sussex.ac.uk / SPRU, Science and Technology Policy Research, University of Sussex, Brighton

Ismael Rafols i.rafols@ingenio.upv.es / Ingenio (CSIC-UPV), Universitat Politècnica de València, València

Abstract

Quantitative techniques for exploring future developments in Science and Technology (here called Future-oriented Technology Analysis (FTA)) are increasingly important in an era of big data and growing computational power. New quantitative techniques such as webometrics, altmetrics and prediction markets are complementing more traditional S&T indicators. While these techniques hold great promise, it is unclear how robust and appropriate is their use under different contexts. In order to help users think through their distinct values and limitations, in this article we discuss quantitative FTA techniques in the light of a general analytical framework. Following Stirling & Scoones (2009), we position FTA quantitative techniques according to their representation of (the incompleteness) of knowledge – i.e. the extent to which they portray their knowledge on probabilities and outcomes as problematic. This framework illuminates the implicit assumptions about the uncertainty, ambiguity and ignorance that distinct quantitative techniques make when exploring (in some cases "predicting") the future. We distinguish between techniques that tend to 'open up' awareness of new or unexpected futures, and others that tend to 'close down' by pointing out to likely futures.

Introduction

Since the middle of the last century, organisations and policy makers began to use a large number of techniques to investigate and influence the future. Several techniques have been proposed since the 1960s, some of which rely on quantitative methods – particularly many of the most recent ones based on the use of internet data (Eerola & Miles, 2011; Porter et al., 2004). In a recent NESTA report (Ciarli, Coad, & Rafols, 2012) we reviewed and classified 26 techniques employed in Future oriented Technology Analysis (FTA) grouped in 10 different families, and we discussed the contexts in which they are most widely used. Although we use here the terminology from the FTA tradition, we notice, *for the purposes of this conference, that FTA techniques are special case of S&T indicators and/or mapping that explore future developments*.

In this article we assess the FTA techniques reviewed by asking how practitioners represent knowledge when they use quantitative techniques for FTA. In other words, we study how the properties of different techniques allow the practitioner to "construct" different states of knowledge about the future.

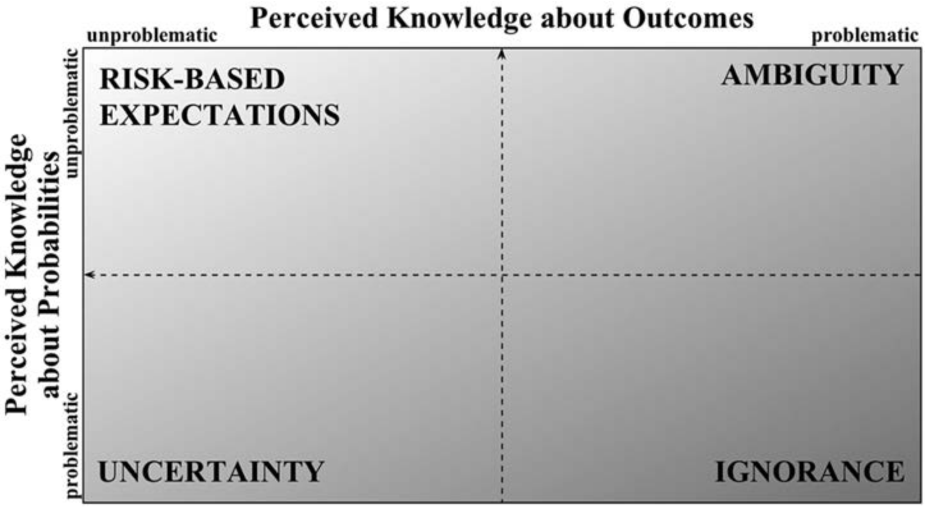
Following Stirling and Scoones (2009) we focus on two main dimensions: knowledge about outcomes and about probabilities. Under the first dimension knowledge is perceived as more or less problematic with respect to which outcome is more or less relevant when considering future states of the world. Under the second dimension knowledge is perceived as more or less problematic with respect to the probability that specific instances of an outcome will occur in the future. We take an explicit constructivist approach: the analyst using an FTA technique can only refer to a limited representation of the complexity of the world, and the outcomes of using a technique depend also on the perspective taken by the analyst and used to simplify the world.

We compare to which extent the use of different families of techniques tend to "open up" ("close down") the range of policy options (i.e. comparable outcomes) resulting from an FTA, and to which extent different techniques broaden (narrow) the range of inputs (i.e. sources of information) used in technology appraisal. We will answer questions such as: how does the use of specific FTA techniques represents knowledge and incertitude, in particular the very incomplete nature of knowledge regarding future technological outcomes? How do they change the way in which we represent knowledge on future events initially considered unknown?

Analytical framework: the incompleteness of knowledge

When an analyst decides to use a given FTA quantitative technique she makes a specific representation of knowledge, i.e. she assumes that some or all variables influencing aspects regarding the future technology can be known, others are uncertain, others are unknown and many others are irrelevant. Following Stirling and Scoones (2009), we distinguish two dimensions in the incompleteness of knowledge, as illustrated in Figure 1.

Figure 1. Representations of knowledge that users of different FTA quantitative techniques may make. Source: Stirling and Scoones (2009).



The horizontal axis describes the perceived knowledge about outcomes. On the left hand side of Figure 1 the analyst considers the outcome of the FTA analysis as not problematic, and assumes that it is fixed. For example, an FTA analysis in the 1950s or 60s might have assumed that urban means of transportation would be based on the combustion engine. Knowledge about the type of the dominant engine technology was perceived as not problematic. However, an analyst in the 2010s on urban means of transportation is likely to represent the possible type of technologies for urban transportation as relevant alternatives to evaluate. Not only there are diverse means (bicycles, cars, metro, tramways), but also she is likely to think that new technological means might appear or their research may be induced. Hence, the knowledge about the outcomes of technological innovation are represented as problematic, or unknown (right side of Figure 1), whereas in the 1960s they were often represented as not problematic, or known (left side of Figure 1).

The vertical axis describes the perceived knowledge about the likelihoods about a certain aspect of a technology, a plausible instance of one of the outcomes (the different technologies). If the analyst perceives that a certain instance can be calculated in a probabilistic manner with a known generation mechanism of the probability distribution and an expected probability of its occurrence, then she is assuming that the knowledge about likelihoods is not problematic. For example, one analyst in the 1960s may have assumed that combustion car ownership trends in a given city were sufficient to "predict" the number of automobiles in the years to come. One analyst in the 2010s might instead think that knowledge about the likelihood of combustion car ownerships is extremely problematic since it depends on a series of variables (public opinion on climate change, governmental regulations on pollution, public health measures, oil cost, and so on), which have behaved in an erratic way for the last 40 years.

This analytic framework leads to four potential "ideal manners of representing knowledge. It is in relation to these ideal representations of knowledge that we can now think on how different FTA techniques are conventionally perceived and used. We should emphasise that there is some flexibility in how an analysts use of a given FTA represents knowledge. Here we describe the conventional uses of the FTA techniques.

When neither knowledge about likelihoods nor knowledge about outcomes is represented as problematic, the analysts engage in *risk-based expectations* (top left of Figure 1). Here there is a "neat" focus on a given technology and a method to estimate one of its aspects. This would be the case of many simple quantitative FTA techniques, such as trend extrapolation. This is the type of approach which is often associated with scientific techniques, possibly because it allows quantification in a similar way that physics does. However, these methods are only valid to the extent that they ensure that exogenous conditions are controlled and fixed so that only the variables under investigation may have an effect on the outcomes. Yet, in practice, technology futures in the mid and longer terms unfold with many variables beyond control (public perceptions, energy source prices, other technologies, political and organisational preferences, events that cannot be known), changing radically and releasing unforeseen signals, having major effects on technological development.

Indeed, when dealing with future technologies, all sorts of changes in conditions – both endogenous and exogenous to the closed system analysed – assumed as stable may disturb the assumptions made by risk-based expectations FTA. If the analyst focuses on well-defined outcomes but represents the system as not being amenable to probabilistic analysis, then she moves into an area of *uncertainty* (bottom left in Figure 1). Roadmapping, which often includes quantitative description of technological trajectories, would be one such case. There is a consensus on the type of out-come desired, and hence the actors involved focus on achieving some given technological specifications. However, the roadmapping exercise is carried out without making assumptions on the likelihood of each instance of the outcome being achieved. A good example is the case of the International Technology Roadmap for Semiconductors (ITRS, http://www.itrs.net) which specifies expected outcomes, while acknowledging uncertainty without carrying out probabilistic assumptions.

Another way of representing the state of knowledge is to assume that probabilities are not problematic, but that the knowledge of outcomes is problematic, because of conflicting assessments on the desirability of these outcomes. Such state is characterised by *ambiguity*. In principle, one cannot find many examples of quantitative FTA techniques leading to ambiguity because in quantitative approaches incomplete knowledge of type of outcomes often occurs with incomplete knowledge on likelihoods. However, many approaches using risk-based expectation type of assessment over diverse potential outcomes could fall into ambiguity. This would be the case, for example, of an exercise looking into future of urban transportation where the stakeholders agreed on the likelihood that various technologies (bicycles, cars, metro, etcetera) were used in a near future on the basis of trend extrapolation, but the stakeholders did not agree on the desirability of those.

Finally, the state of knowledge can be represented as *ignorance*. This is the state of knowledge that Donald Rumsfeld, then US Defence Secretary, made famous with his quote on "unknown unknowns", namely those "things we do not know we don't know." In short, when an analyst considers that she is in a state of ignorance, she assumes that she does not know what are the potential types of outcomes (or their desirability), nor the probability that they occur. One might think that ignorance is the most sensible way to represent technological futures, given that futurology or forecasting have an extremely bad record on prediction. But ignorance is a difficult state to work with. One can try to carry out forecasting or foresight with some known potential outcomes, but how should one characterise quantitatively unknowns?

The representation of knowledge by users of FTA techniques

In Figure 2 we map the 10 groups of FTA quantitative techniques surveyed in Ciarli, Coad and Rafols (2012) according to their relative position with respect to the representation of knowledge that a user has before an FTA exercise. For all techniques the knowledge on probabilities is initially represented as problematic.

Techniques vary to a large extent with respect to how the knowledge on outcomes is represented. Some techniques are prevalently employed when the outcomes of the technology under investigation are already perceived as relatively certain. For example, the use of Trend Analyses to describe the diffusion of a technology in terms of the share of users. Other techniques are employed when the knowledge about outcomes is presented as highly problematic. In this short version of the article, we only illustrate how the use of FTA changes the representation of knowledge with reference to Trends Analyses (Figure 3) and different types of Modelling (Figure 4). Figure 5 presents all the techniques reviewed.

The choice of techniques that are part of the Trend Analyses family tends to have a closing-down effect since the beginning of an FTA exercise because of the use of very restrictive assumptions. Their use tends to increase the certainty about the representation of knowledge even more.

In the case of modelling, the effect depends on the type of modelling. Different economic methods have very different effects on technology appraisal (Figure 4). On the one hand Prediction Markets close down on one single outcome and a well defined expected value with low variance (Risk-Based Expectations). On the other hand, Input-Output models may even open up to different scenarios, although the relevant outcomes are usually defined at the outset. Simulation Models usually allow to open to a number of different, even non-predicted outcomes. Starting from a condition of ignorance – fewer assumptions on the outcomes, Quantitative Scenarios represent one relatively elaborate way to use Simulation Modelling. Their aim is to find conditions under which a large number of different outcomes can be realised. The only reduction in policy space occurs towards a perceived increase of knowledge about the likelihood of the different outcomes. This is achieved thanks to the combination of a large number of conditions defined by a number of parameters.

Figure 2. The representation of knowledge about outcomes and probabilities before an FTA.

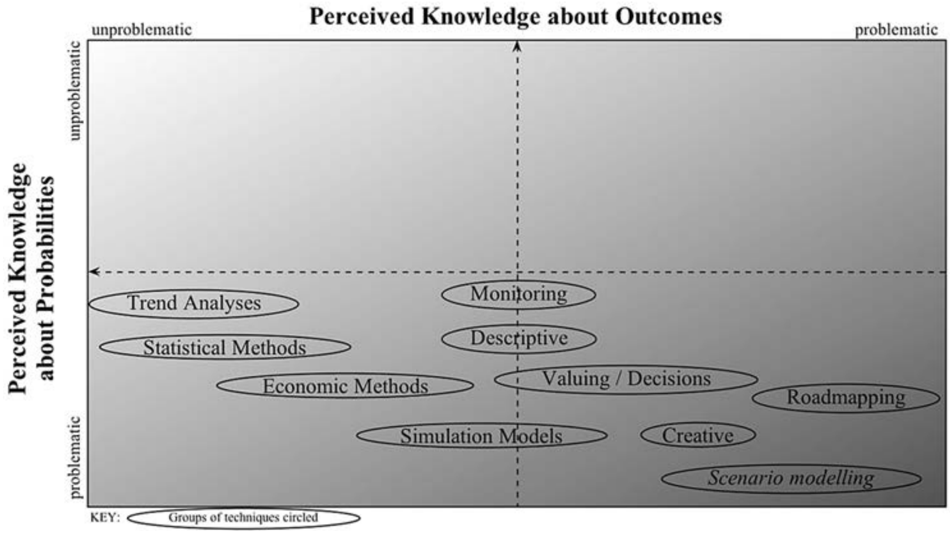


Figure 3. The change in the representation of knowledge about outcomes and probabilities using an FTA quantitative techniques. The case of Trend Analyses.

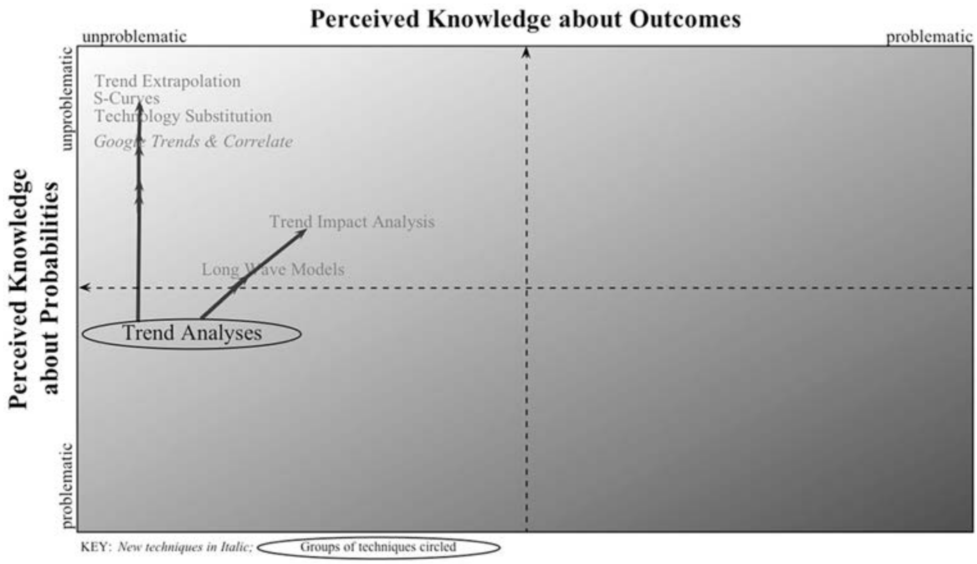


Figure 4. The change in the representation of knowledge about outcomes and probabilities using an FTA quantitative techniques. The case of different types of modelling: Economic Methods, Simulation Models and Scenario Modelling.

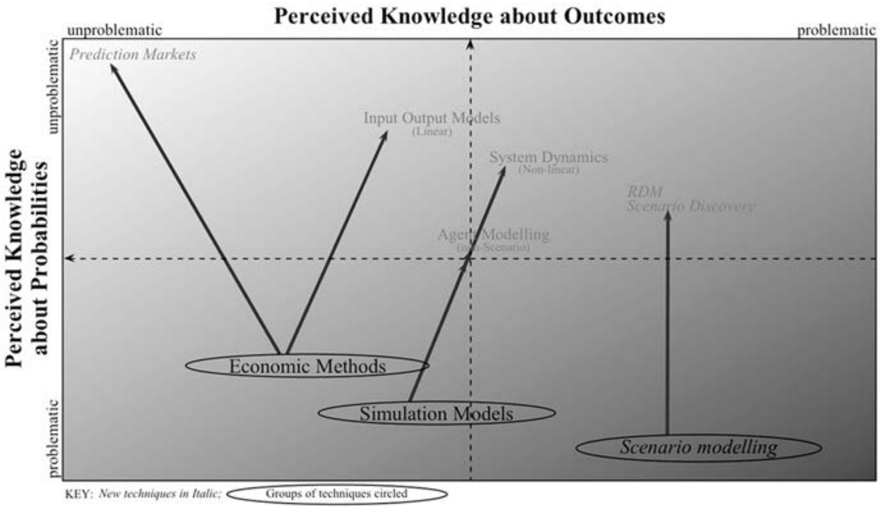
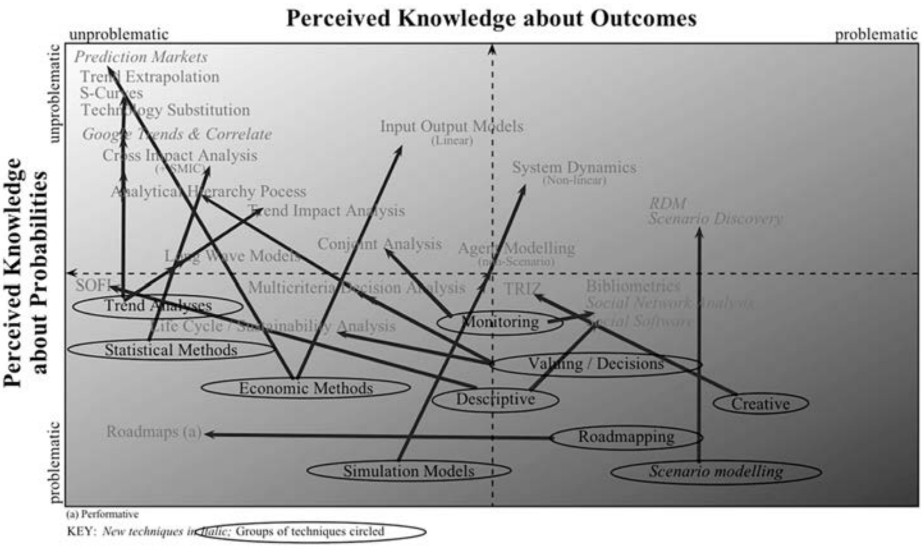


Figure 5. The change in the representation of knowledge about outcomes and probabilities: using techniques.

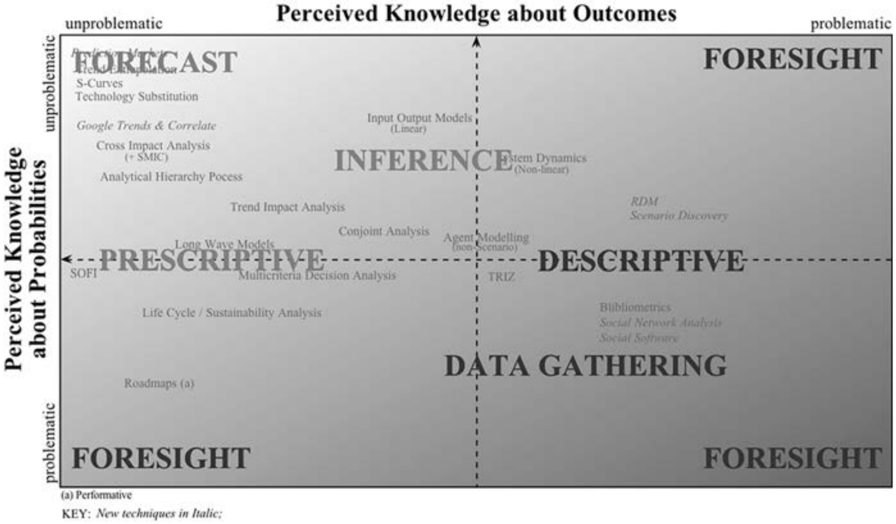


Distinguishing between foresight and forecasting

FTA techniques that focus on a given outcome and specific properties in order to provide allegedly quantitatively rigorous approaches (generally falling in the risk-based expectations zone, Figure 6, top left), do so at the expense of putting blinders to phenomena which are very difficult to quantify and which may dramatically alter the results of the "prediction" (an unexpected war, conflict, or terrorist attack). These are the type of approaches that resemble in one way or the other the traditional forecasting literature.

As a result of the repeated failure of forecasting exercises, policy-oriented analysis of future technologies puts the emphasis on foresight rather than quantitative forecasting. This practical focus acknowledges that the usefulness of foresight lies in opening up a process to discuss technological futures in the face of insurmountable incertitude, rather than trying to predict the future. The contribution of quantitative techniques to foresight seems to be best achieved via FTA techniques that represent knowledge as incomplete, either in terms of the probability of known outcomes or in the type of outcomes.

Figure 6. The representation of knowledge about outcomes and probabilities and their role in Forecast and Foresight FTA.



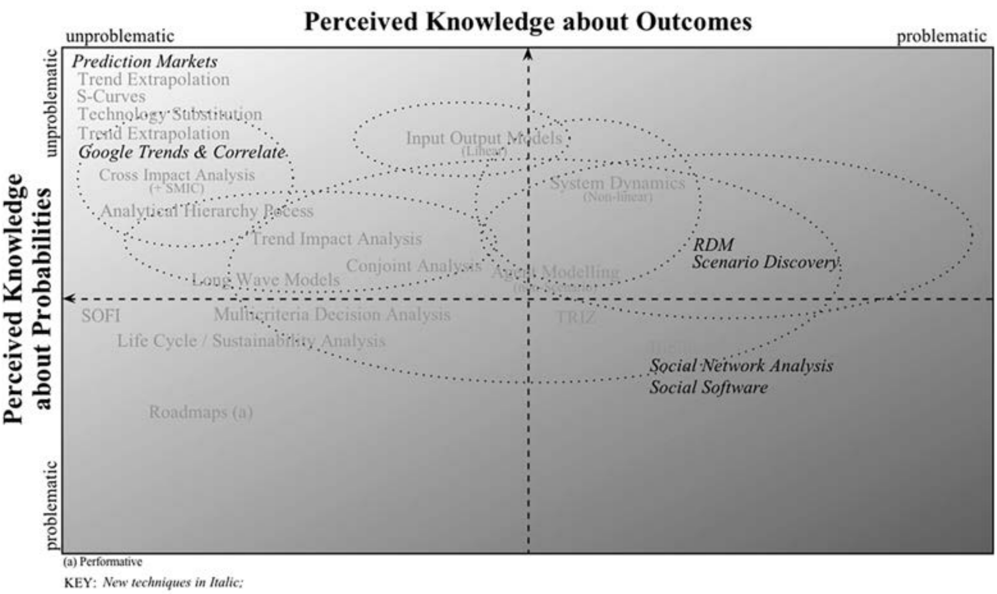
Breadth of inputs and "participation" in FTA: the role of new techniques

The advantage in using new crowd-sourcing based techniques is that they collect information from a very large number of users, rather than from a small number of experts or from structured sources that acquired an authoritative status such as patent and scientific publication datasets. In other words, they broaden the source of inputs, including outputs and opinions that were not usually included in FTA. Indeed, by monitoring the opinions and sentiments on blogs and social

network such as Twitter, it is possible to use more or less spontaneous opinions about past, present and future events (the degree of spontaneity is increasingly problematic: monitoring is already resulting in manipulations).

Although these methods collect very diverse types of information and therefore they broaden up the diversity of information sources, they do not necessarily have an opening up effect in decision making. Let us think for example in carbon-based nanotechnologies. The exponential boom in presence of fullerenes and carbon-nanotubes in databases and the web in the 1990s meant that most quantitative FTA techniques would have focused on the impact of these two types of carbon materials. This focussing (closing-down) effect might have had a blinding effect, preventing to see that there was ongoing research on other carbon-based nanomaterials, and hence it would have missed out the possibility that graphene, another carbon-based nanomaterial, would become important. Instead a simple qualitative study that discussed diverse carbon-based nanomaterials could have mentioned graphene, since it had been studied theoretically for decades. This example illustrates that quantitative power does not imply capacity for seeing more alternative futures – an opening up of perspectives.

Figure 13. New techniques for FTA.



Discussion

The improvements in computational capabilities are resulting in the creation of new FTA techniques that promise great improvements in their capacity to generate reliable predictions. The analytical framework we proposed should help discern the type of contribution to policy that they can make.

Our analysis suggests that the most promising quantitative methods for conducting FTA in a sophisticated manner are those that allow the analysis to explore states of ignorance – in which neither future technology outcomes nor their probability are known – and bring the user to a state of ambiguity, in which future outcomes are compared against different probabilities of occurring, and users can evaluate a number of trade-off.

These exercises are not capable to predict instances of outcomes, but they help explore the future in a conditional manner, acknowledging the incompleteness of knowledge. Using the techniques plotted in the ambiguity quadrant of the knowledge map, one can investigate the possibility of reaching certain outcomes under circumstances that are unknown but can be investigated in a conditional manner. We suggest that these types of agent modelling and scenario modelling are the ones which can make a more positive contribution to policy-oriented FTA – by avoiding narrow prediction and allowing plural exploration of future technologies.

Finally, monitoring methods (such raw bibliometrics or web-scraping) may be able to identify potential outcomes and be useful for activities such as horizon-scanning, but they have limited analytical potential on their own. Therefore, their usefulness depends on their implementation within a larger foresight methodology.

References

Ciarli, T., Coad, A., & Rafols, I. (2012). Quantitative Analysis of Technology Futures. Part I: *Techniques, Contexts, and Organisations* (NESTA Report). Brighton: SPRU, University of Sussex.

Eerola, A., & Miles, I. (2011). Methods and tools contributing to FTA: A knowledge-based perspective. *Futures*, 43(3), 265–278.

Porter, A. L., Ashton, W. B., Coates, G. C. nad J. F., Cuhls, K., Cunningham, S. W., Ducatel, K., … Thissen, W. (2004). Technology futures analysis: Toward integration of the field and new methods. *Technological Forecasting & Social Change,* 71, 287–303.

Stirling, A. C., & Scoones, I. (2009). From Risk Assessment to Knowledge Mapping: Science, Precaution, and Participation in Disease Ecology. *Ecology and Society,* 14(2). Retrieved from http://www.ecologyandsociety.org/vol14/iss2/art14/

The relationship between the performance of editorial board members and the impact of scientific journals: an exploratory analysis[1, 2]

Rodrigo Costas rcostas@cwts.leidenuniv.nl / Centre for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62-A, Leiden, 2333 AL (the Netherlands)

Clara Calero-Medina clara@cwts.leidenuniv.nl / Centre for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62-A, Leiden, 2333 AL (the Netherlands)

Ludo Waltman waltmanlr@cwts.leidenuniv.nl / Centre for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62-A, Leiden, 2333 AL (the Netherlands)

Paul Wouters p.f.wouters@cwts.leidenuniv.nl / Centre for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62-A, Leiden, 2333 AL (the Netherlands)

Abstract

This paper presents an extensive analysis on the relationship between the impact of scientific journals and the performance of the editorial board members that serve in the journal. A novel methodology for the data collection of the publications at the individual level has been applied. Preliminary results indicate that there is a clear correlation between the impact of the editorial board members and the impact of the journals. Further analyses are pointed out for the full version of the paper, including the more thorough analysis of this relationship across fields of science or the study of other elements that could also play a role in the determination of the impact of scientific journals.

Introduction

Editorial boards are a key element of scientific journals as they are the journal's intellectual gatekeepers and they are considered as the cornerstone of any peer-reviewed journal (Lowe & van Fleet, 2009; Cabanac, 2012). Editorial board members (EBMs) are expected to be among the most prestigious scholars in a given field (Campanario et al, 2006), and play a key role in the configuration of science. The composition of the editorial board can be seen as an indication of the reputation of a scientific journal. In this paper, we explore the hypothesis that the scientific performance of the EBMs of a journal is a proxy measure of the impact of the journal. The mechanism underlying this hypothesis may be that journals with EBMs with high impact and

1 This work was supported by a grant from the SURFSHARE program (the Netherlands) http://www.surf.nl/en/Pages/default.aspx.

2 The authors want to thank Bert van der Wurff and Marieke Klijn from CWTS for their help in the data collection for the study.

visibility would ‘attract’ authors and papers that will be also of high (or similar) impact. Thus, we wish to analyze how strong the relationship is between the scientific level of the EBMs of journals and the actual impact of the journals. Some of the previous studies that have focused on the problem have not produced conclusive evidence on this point (e.g. Lowe & van Fleet, 2009; Cabanac, 2012) and have been limited to relatively small samples. In this paper, we tackle the issue from a broader point of view, studying the relationship between EBM performance and journal impact in an extensive set of journals, based on a novel methodology for data collection and considering an important range of bibliometric indicators. This paper has emerged from a previous analysis done by the authors (Costas et al, 2012) where the possibilities of developing a predictive model for recently created Open Access journals were explored. The data used in that study are used here, but in this case focusing only on the relationship between the performance of EBMs and journal impact.

Objectives of the study

The main aim of this study is to explore the relationship between the scholarly performance of the EBMs of scientific journals and a number of commonly used journal impact indicators.

Methodological development

A total of 300 journals from 5 different Web of Science (WoS) subject categories have been selected[i]: ‘Chemistry, Analytical’ (57), ‘Economics’ (59), ‘Physics, Condensed matter’ (45), ‘Psychiatry’ (72) and ‘Radiology & nuclear medicine’ (67).

Data collection on the EBMs of the selected journals

For these 300 journals their EBMs were collected through manual searches of the websites of the journals in 2012. Different information elements for every EBM were recorded:

– *Surname and initials*: these are key elements for the identification of every EBM.
– *E-mail*: e-mails of the EBMs (only available for 4% of all the cases detected).
– *Address (and/or country)*: affiliation (or country) information of the EBMs, available for 87% of the cases, although not standardized and thus only of limited help.

As a result 14,303 unique combinations of a journal and an EBM have been recorded. For all of them, their publications covered in the Web of Science were collected. This was the most difficult step (given the problems of data collection at the individual level, cf. Costas et al 2010). In order to perform this step in an accurate and efficient way we applied a variation of the “Seed+Expand” methodology recently developed by Rijnhoudt et al (2013). The main steps of the adaptation of this methodology are explained in the following paragraphs.

The "Seed+Expand" methodology

This methodology consists basically of two main steps: first the creation of a 'seed' (a 'seed' is a (group of) publication(s) for a given EBM that with a very high probability belongs to that person) for the different EBMs of the journals; and secondly the 'expansion' of this seed through other bibliometric methodologies.

Seed. Different seeds have been created for the EBMs:

(1) *E-mail seed*: matching of the e-mail address of the EBM with the e-mail information available in the scientific publications recorded by WoS.

(2) *Author-journal seed*: publications where the name of the EBM appears in a publication in the same journal for which he/she is an EBM. This seed is based on the assumption that very likely EBMs publish in the journals for which they sever as board members.

(3) *Self-citations of Author-Journal seed*: based on the previous seed we collected the author self-citations of the EBMs to the publications detected in the previous seed. We assume that author self-citations (i.e. publications that carry the same EBM name to those publications from seed 2) very likely belong to the same EBM.

Obviously, seeds 2 and 3 are not free of limitations and very general names of EBMs can bring false positive results (i.e. publications that do not belong to the real EBM). To minimize these problems we excluded from the whole analysis all EBMs whose names are among the most common[ii] author names in the WoS database.

Expansion. Once the final seed was created the next step was to collect the full 'oeuvres' of these EBMs. For this we have used the same approach as in Reijnhoudt et al (2013) of combining the novel micro classification of scientific fields developed at CWTS (Waltman & van Eck, 2012b) and the *researcher_id* existing in Scopus (Moed et al, 2013). In other words, for a given name of an EBM with a seed, we collected all the other publications carrying the same name that belong to the same 'micro-disciplines' or that presented the same 'Scopus author-id' identified in the seed[iii].

After the expansion of the seed and before the final analysis we have excluded all publications of EBMs in the journal for which they are EBM (in other words, we have excluded publications for the EBMs coming from seed 2). The underlying idea is that one may want to estimate the impact of journals not covered by one's bibliographic database (e.g., recently established journals, as in Costas et al, 2012) based on the publications of EBMs in journals that are covered by the database. In order to approximate this as closely as possible, publications of EBMs in their own journal have been excluded from the analysis.

As a result of the execution of all the steps previously described we managed to collect publications for almost 59% of all the initial journal-EBM combinations (i.e. 8,497 combinations) thus keeping a quite large and still reliable set of EBMs and publications to perform the subsequent analysis.

Analysis and results

Indicators calculated

The main indicators calculated for the EBMs and for the different journals that were considered for this study are presented in Appendix I. In general terms we calculated 2 main types of indicators:

– *Journal indicators*: these are the bibliometric indicators calculated for the 300 journals which form the object of study. The period of analysis for these indicators is 2008–2010 for publications and citations up to 2011. The Impact Factor indicator has been calculated following the regular 2-year approach (Garfield, 2003) considering publications in 2011.

– *Indicators for the EBMs*: these are the bibliometric indicators calculated for the oeuvres of the individual EBMs of every journal based on the publications previously collected through the "Seed+Expand" methodology.

The numeric indication at the end of the label of an indicator signals the period of time of publications for which the indicator has been calculated (e.g., jfis_0810 refers to the JFIS indicator calculated based on publications from the period 2008–2010). For indicators of citation impact the citation window always includes at least one full year (cf. Waltman et al, 2011).

Statistical analysis

A Pearson correlation matrix has been calculated in order to study the main relationships among all the indicators. It is important to take into account that the statistical analysis has been limited to journals with at least 10 EBMs, thus reducing the potential effects of journals with fewer EBMs (which could be easily influenced by outliers). We have also excluded journals with less than 100 publications in the period 2008–2010. Thus a total of 286 journals finally entered in the analysis.

Correlation matrix of the EBM-based indicators with the impact of the journals

Table 1 presents the correlation matrix of the EBM-based indicators with the impact of the target journals. The Pearson correlation coefficient has been used. For simplicity reasons in the table we focus on the correlations of all the indicators with the two journal impact indicators (*Impact Factor* and *JFIS 2008–2010*), thus it is possible to see which indicators have the highest correlations with them.

Table 1. Correlation matrix of indicators with the two journal impact indicators

	Impact Factor	jfis_ 0810	AvgOfind _p_0110	AvgOfind_ tcs_0110	AvgOfind_ mcs_0110	AvgOfind_ mncs_0110	AvgOfind_ mnjs_0110	AvgOfind_ pptop10_0110	AvgOfind_ uncited_0110	h_index
Impact Factor	1	.809	.362	.704	.707	.499	.502	.543	-.488	.525
jfis_ 0810	.809	1	.057	.363	.551	.624	.672	.658	-.371	.204
AvgOfind_ p_0110	.362	.057	1	.763	.262	.041	.058	.074	-.293	.675
AvgOfind_ tcs_0110	.704	.363	.763	1	.703	.383	.342	.404	-.420	.797
AvgOfind_ mcs_0110	.707	.551	.262	.703	1	.712	.663	.760	-.608	.555
AvgOfind_ mncs_0110	.499	.624	.041	.383	.712	1	.817	.887	-.399	.207
AvgOfind_ mnjs_0110	.502	.672	.058	.342	.663	.817	1	.870	-.498	.218
AvgOfind_ pptop10_0110	.543	.658	.074	.404	.760	.887	.870	1	-.476	.273
AvgOfind_ uncited_0110	-.488	-.371	-.293	-.420	-.608	-.399	-.498	-.476	1	-.445
h_index	.525	.204	.675	.797	.555	.207	.218	.273	-.445	1

Table 1 shows, among some expected correlations (e.g. the strong correlation among the two journal-based indicators), how the two main impact indicators for journals (i.e. the Impact Factor and the JFIS) present medium to high correlations with some of the EBM-based indicators, particularly the average TCS and MCS of the EBMs for the Impact Factor and the average of the individual MNCS, MNJS and PP(top 10%) of the EBMs for the JFIS, thus suggesting a relatively good correlation of the impact of scientific journals with the average impact performance of the EBMs of the journals. Notice that the non field normalized indicators of the EBMs correlate better with the Impact Factor (which is also a non field normalized indicator) while the field normalized indicators correlate better with the JFIS indicator (which is field normalized itself). Hence, as might be expected, field normalized indicators correlate best with other field normalized indicators, while non normalized indicators correlate best with other non normalized indicators.

In Figure 1, the linear correlations of the two journal-based indicators with the most strongly correlating EBM-based indicators are presented (i.e. Impact Factor vs. MCS of EBMs and JFIS vs. MNJS of EBMs).

Figure 1. Linear correlation IF vs. EBMs MCS and JFIS vs. EBMs MNJS

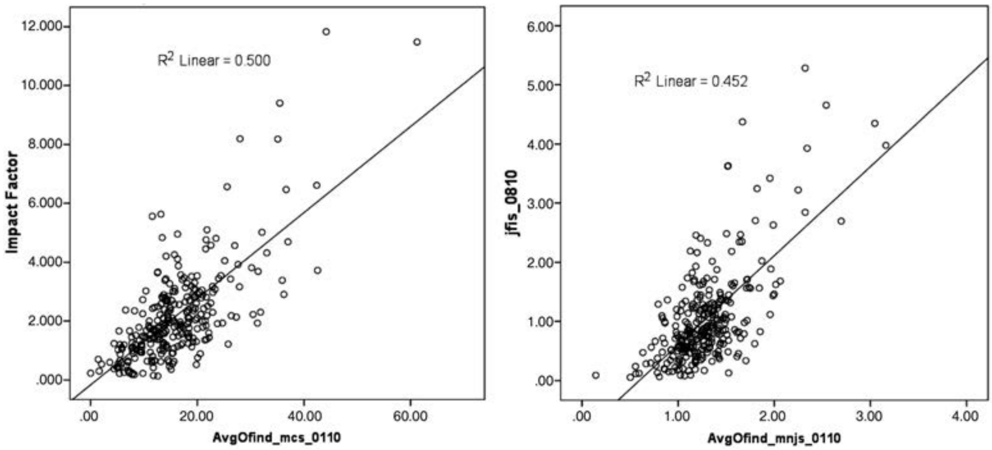


Figure 1 presents the scatter plot of the linear correlations, showing how the R^2 is 0.50 in the case of the Impact Factor with MCS of the EBMs of the journal, while the JFIS has a correlation with an R^2 of 0.45 with the MNJS of the EBMs of the journal. In general, we can see that the highest impact journals (in terms of both Impact Factor and JFIS) are fairly well separated from the lower impact journals. Among the lower impact journals, however, the relationship between the EBM-based indicators and the journal indicators is not very strong.

Discussion

In this paper we analyzed the impact of a selection of journals in relationship with the performance of the EBMs of the journals. Data collection on EBMs is a difficult and time consuming step. This difficulty comes mainly from the lack of standardization of the information on EBMs in scientific journals. However, we have collected a quite unique data set and to the best of our knowledge there are no other studies that have performed such an extensive and broad analysis on EBMs of scientific journals.

The results of this paper indicate that bibliometric indicators based on the performance of the EBMs of scientific journals are quite strongly correlated with the impact of the journals (with Pearson correlations between 0.5 and 0.7). Hence, high-impact journals tend to have EBMs whose external publications (i.e. not published in the same journal) on average are cited quite frequently, while journals with a lower impact tend to have EBMs whose external publications on average receive fewer citations.

Considering these results and regarding the final version of the paper, we plan to explore the potential development of "predictive" models of the impact of journals based on EBM-based information, also the exploration of other possible factors that could play a role in the configuration of the impact of journals[iv] and a more thorough analysis of the correlation between the performance of EBMs and the impact of journals across different scientific disciplines.

References

Baccini, A. & Barabesi, L. (2011). Seats at the table: the network of the editorial boards in information and library science. *Journal of Informetrics,* 5(3), 382–391.

Cabanac, G. (2012). Shaping the landscape of research in information systems from the perspective of Editorial Boards: a scientometrics study of 77 leading journals. *Journal of the American Society for Information Science and Technology,* 63(5), 977–996.

Campanario, J.M., González, L. & Rodríguez, C. (2006). Structure of the impact factor of academic journals in the field of Education and Educational Psychology: citations from editorial board members. *Scientometrics,* 69(1), 37–56

Costas, R. & Bordons, M. (2007). The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics,* 1(3), 193–203

Costas, R., Calero-Medina, C., Waltman, L. & Wouters, P. (2012). Surfshare Report: Development of indicators based on the performance of Editorial Board Members to predict the potential impact of scientific journals. *In press*.

Costas, R., van Leeuwen, T.N. & Bordons, M. (2010). A bibliometric classificatory approach for the study an assessment of research performance at the individual level: the effects of age on productivity and impact. *Journal of the American Society for Information Science and Technology,* 61(8), 1647–1661.

Garfield, E. (2003). The meaning of the Impact Factor. *International Journal of Clinical & Health Psychology,* 3(2), 363–369.

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America,* 102(46), 16569–16572.

Lowe, D.J. & van Fleet, D.D. (2009). Scholarly achievement and accounting journal editorial board membership. *Journal of Accounting Education,* 27(4), 197–209

Moed, H.F., Aisati, M. & Plume, A. (2013). Studying scientific migration in Scopus. *Scientometrics,* 94(3), 929–942.

Reijnhoudt, L., Costas, R., Noyons, E., Börner, K. & Scharnhorst, A. (2013). "Seed+Expand": a validated methodology for creating high quality publication oeuvres of individual researchers. *ArXiv* <http://arxiv.org/abs/1301.5177>

Waltman, L., van Eck, N.J., van Leeuwen, T.N., Visser, M.S. & van Raan, A.F.J. (2011). Towards a new crown indicator: an empirical analysis. *Scientometrics,* 87(3), 467–481.

Waltman, L. & van Eck, N.J. (2012a). The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology,* 63(2), 406–415.

Waltman, L. & van Eck, N.J. (2012b). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology,* 63(12), 2378–2392.

Appendix

Appendix I. Brief description of the indicators included in the study (and their labels)

INDICATORS	EXPLANATION
– Journal indicators	**Period 2008–2010/11**
jfis_0810	MNCS of the target journal
IF (Impact Factor)	Impact Factor of the target journal (2 years)
– EBM-based indicators	**Period 2001–2010/11**
AvgOfind_pptop10_0110	Average of the PP(top 10%) of the EBMs of the target journal
AvgOfind_mncs_0110	Average of the MNCS of the EBMs of the target journal
AvgOfind_mnjs_0110	Average of the MNJS of the EBMs of the target journal
AvgOfind_mcs_0110	Average of the MCS of the EBMs of the target journal
AvgOfind_p_0110	Average of the P of the EBMs of the target journal
AvgOfind_tcs_0110	Average of the TCS of the EBMs of the target journal
h-index[v]	Average of the h-index of the EBMs of the target journal

Some abbreviations:

- P: number of publications (article, review and letter).
- TCS: total citation score (total number of citations received by the P publications excluding self-citations).
- PP(top 10%): percentage of publications in the top 10% of their field.
- MCS: mean citation score.
- MNCS: field normalized mean citation score.

– MNJS: field normalized mean journal score (with the score of a journal being equal to the mean normalized citation score of the publications in the journal).
– JFIS: MNCS of the target journal (i.e. field normalized impact of the target journal).

---

i The selection of the subject categories was done based on the presence of a sufficient number of journals with at least 500 publications in the period 2002-2011 in WoS, while trying to keep a relative diversity of disciplines (i.e. including medical, social sciences and natural sciences journals in the sample).

ii From a practical perspective, a 'common name' has been considered a name that appears 700 or more times in the whole WoS (during the period 1980-2011).

iii In the Reijnhoudt et al (2013) study, values of ~84% precision and ~96% recall have been reported for this methodology.

iv An example could be the analysis of the impact of the authors in the first issue/year of a journal, from the point of view that if the first authors in a journal are highly visible (and cited) they could also have an attractive effect for other highly cited authors that could contribute to the journal in future issues and years, thus increasing the impact and visibility of the journal.

v The h-index as suggested by Hirsch (2005) has been also calculated for the researchers during the period 2001–2010, although in this case we have considered all the citations up to the last date in the database. We note that the h-index is an inconsistent indicator with important limitations (Costas & Bordons, 2007; Waltman & van Eck, 2012a).

Productivity differences influence Swedish universities' citation-based university rankings: An analysis of the relationship between work productivity and citation performance

Rickard Danell Rickard.danell@soc.umu.se / Department of Sociology, Umeå University

Abstract

The argument presented in this study is that when the quality of the research papers being produced from two organizations is the same, researcher from one organization can still be seen as the higher performer in terms of citations. The differences are in their work productivity. This is defined as the number of produced units per unit of work. This study shows that there are differences in the work productivity among Swedish universities and that the ranking of universities changes when these differences are accounted for. I also demonstrate that researchers belonging to organizations with greater work productivity are also more likely to produce highly cited papers.

Introduction

In recent years we have seen the discussion on evaluative citation analysis intensify and there have been a growing number of creative proposals for alternative methods of calculation. For example, there have been suggestions for standardized item oriented citation rates (Lundberg 2007) and tests of different evaluative models (Ahlgren et. al. 2012). Proposals have been made for methods to compare sets of documents in terms of citation rates (Colliander & Ahlgren 2011; Leydesdorff & Bornemann 2012; Schneider 2013). There are also theoretically appealing suggestions of new ways of measuring citation impact (Leydesdorff & Bornmann 2011; Bornmann et. al.2013).

However, all methods relate to number of citations received compared to number of publications produced. In this paper I will further complicate the discussion of how citation measurements should be used to rank university organizations. This paper will address the issue of variations in work productivity among university organizations and how this affects university rankings.

My argument is that even if the quality of the papers being produced from two organizations is the same, researchers in one organization can still be seen as the higher performer in terms of citations. The issue at hand is the differences in work productivity. I define work productivity as the number of produced units per unit of work, such as the number of articles produced per year of work. If there are differences in the work productivity between organizations, then relating highly cited papers to all papers will give a different picture from relating highly cited papers to researchers.

Let us consider two organizations (A and B). Both organizations have a work input of 100 full-time researchers, but organization A produces 1000 papers and organization B produces 500 papers. Therefore the work productivity of organization A is 10 and the work productivity of organization B is 5. Let us further consider that 10 percent of each produced document set is highly cited (let's say that these documents belong in the top 10 percent of highly cited papers). This means that the share of highly cited papers is 0.10 for each organization, making the organizations equal in terms of performance. However, if we relate this to the amount of work done, organization A clearly outperforms organization B. There is one highly cited paper per researcher in organization A, compared with 0.5 in organization B. It seems clear from this example that the quality of the paper output is the same, but the performance of the researchers in the organizations differ.

Materials and measurements

The Leiden 2012 ranking of universities was used to gather information on the number of papers produced at ten of the largest higher education institutions in Sweden. The Leiden 2012 ranking is based on Web of Science publications from 2005–2009. Using the Leiden ranking, I have gathered information on the number of papers, the number of publications in the top 10 percent, and the mean normalized citation score. In the Leiden ranking, this information is available both in a whole count version and a fractional count version. In this study I have used both count methods.

In order to estimate work productivity information is needed on the amount of work done in the universities during the period 2005–2009. I gathered this information from the Swedish Council for Higher Education. The information used is the number of employed researchers in full time equivalents, which I shall refer to as *person-years*. It was possible to distinguish between scientific area and different employee categories. In the analysis, the humanities have been excluded since the Leiden ranking is based on *Science Citation Index* (SCI) and *Social Science Citation Index* (SSCI). The researchers used in this study include researchers employed at a position that requires a PhD as well as those employed as doctoral students. Including doctoral students in the researcher population may not be reasonable in every case, but when Swedish universities are studied this is reasonable. In Sweden, doctoral student are employed at the universities and as part of the team of researchers.

Results

Citation performance among the ten largest Swedish Universities

Only 10 out of 15 Swedish universities are included in the Leiden ranking. Among these universities we find Sweden's oldest university, Uppsala University (UU), instituted in 1477. Also included are Lund University (LU), instituted in 1666, Gothenburg University (GU), which was awarded university status in 1954, Stockholm University (SU), which was awarded university status in 1960, Umeå University (UmU), instituted in 1965, and Linköping University (LiU), instituted in 1975.

The four specialized higher education institutions with university status are the medical university Karolinska Institute (KI) in Stockholm, instituted in 1810, The Royal Institute of Technology (KTH) in Stockholm, founded in 1827 and awarded university status in 1927, Chalmers University of Technology (CU) in Gothenburg, which has had university status since 1937, and the Swedish University of Agricultural Sciences (SLU), with campuses in several parts of Sweden. The Swedish University of Agricultural Sciences has existed in its present form since 1977.

First, let us take a look at the differences in citation performances among Swedish universities according to the Leiden 2012 ranking (Figure 1a and Figure 1b). Two measures from the Leiden ranking are used in the figures: the proportion of papers in the top 10 percent and the mean normalized citation score per paper. The figures also display differences in citation performances when different counting methods are used, such as the whole count method and the fractional count method. The universities are displayed in rank from left to right in accordance with their citation performance based on the whole count method. There are some changes in the relative position of the universities depending on whether the whole count method was used or the fractional counting method was used. There are also some differences in rank depending on whether the top ten percent indicator was chosen or the indicator based on mean normalized citation scores.

Figure 1a. Performance differences among Swedish higher education institutions (proportion of publications in the top 10 percent) (Leiden ranking 2012)

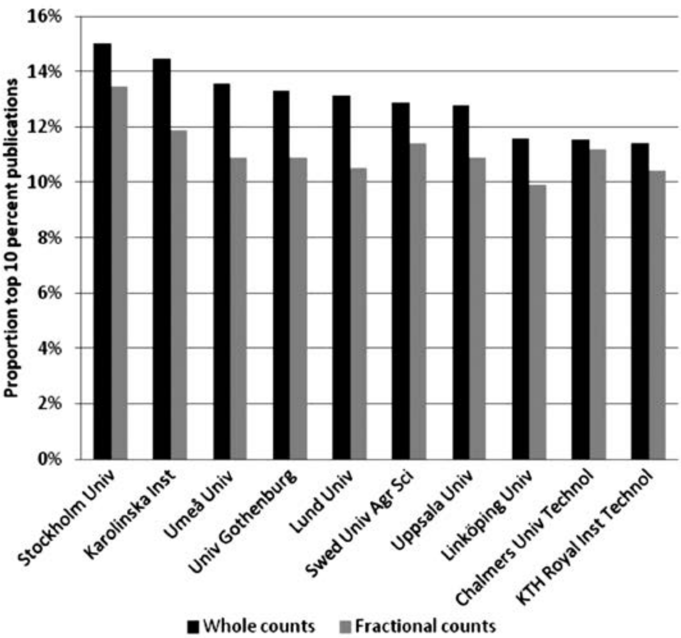
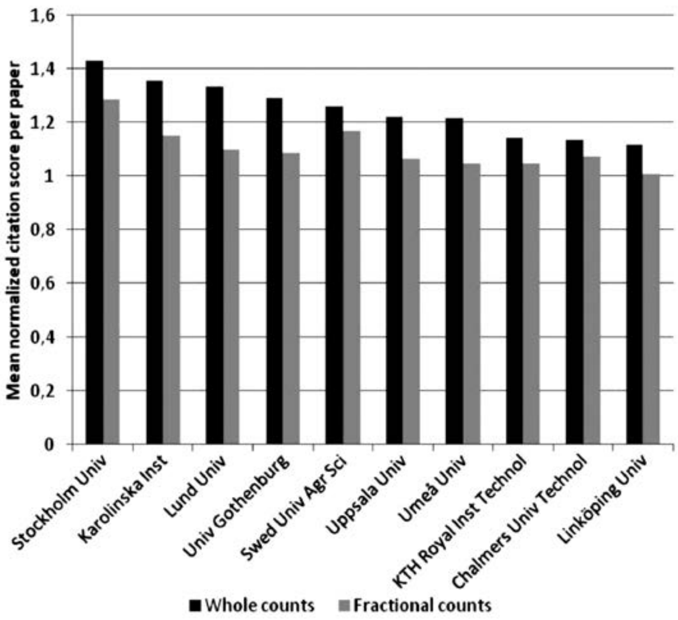


Figure 1b. Performance differences among Swedish higher education institutions (Mean normalized citation score per paper (Leiden ranking 2012))



According to the Leiden ranking, Stockholm University stands out and should be considered as the new top university in Sweden. Stockholm University remains in the lead independent of counting method and indicator. Stockholm university has the highest share of publications in the top 10 percent (both when whole counts and fractional counts are used) and Stockholm University has the highest mean normalized citation score. As expected, the use of fractional counts has a different effect on the universities, depending of the level of collaboration. Using the fractional counting method seems to reduce the differences between the universities and the coefficient of variation confirms this observation, although the changes are small. They differ from 0.09 to 0.08 as the top ten percent indicator, and from 0.08 to 0.07 for the normalized citation score indicator.

The technological universities and the agricultural university gain the most from the use of fractional counting. With the exception of Stockholm University and the Karolinska Institute, after fractionalization most of the institutions included have values that fall within the stability intervals included in the Leiden ranking.

Therefore, our conclusion should be that, with the exception of Stockholm and Karolinska, there is no performance difference between the Swedish universities included in the Leiden ranking.

Differences in work productivity among Swedish universities

Work productivity has been estimated by dividing the number of papers produced by the university with the number of person years of research done during the same period. The resulting ratio is papers per person year, or equivalently the average number of papers produced per year of work. I have used both whole counts and fractional counts when estimating the work productivity.

Figure 2. Annual number of papers per employed researcher ("person years")

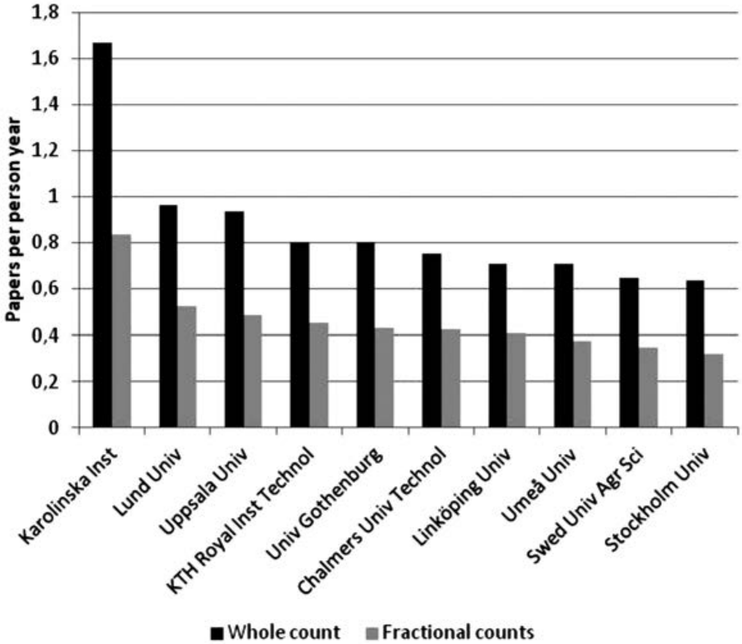


Figure 2 displays how the work productivity varies among Swedish universities. The differences in work productivity are striking. Work productivity is clearly highest at the Karolinska Institute and Lund University. In Figure 2 Uppsala University is in third place for work productivity, although the difference between Lund and Uppsala are not substantial.

The observed differences in work productivity should be considered as Web of Science productivity, which may vary from actual differences in work productivity. For example, the high productivity of Karolinska Institute may to some extent be caused by better coverage in Web of Science publications from Karolinska; the Karolinska Institute is a medical research university. The lower work productivity at the technological institutions could also be affected by the research profile and idiosyncrasies associated with technological research. Differences in work productivity among the less specialized universities (such as Lund University, Uppsala University, Gothenburg University, Linköping University, and Umeå University) are harder to explain with reference to differences in their research profile. Research profile of the organization could

be one reason for potential differences between real work productivity and Web of Science work productivity, although there could be other reasons. For example, better coverage in Web of Science, given the field of science, could be an indication of higher research quality. However, independent of whether the observed differences are real differences in work productivity, or whether differences in work productivity is the result of differences in database coverage, variation in work productivity could be a problem for citation-based bibliometric evaluations of universities.

The relationship between the researcher's overall productivity and researcher's top performance productivity

Next we will look at the relationship between work productivity and the production of highly cited publications (as defined by the Leiden 2012 ranking). Figure 3 illustrates how the overall work productivity is related to the production of highly cited articles, in this case the number (or fraction) of articles in the top 10 percent.

Figure 3. Relationship between work productivity and citation performance (fractional counts)

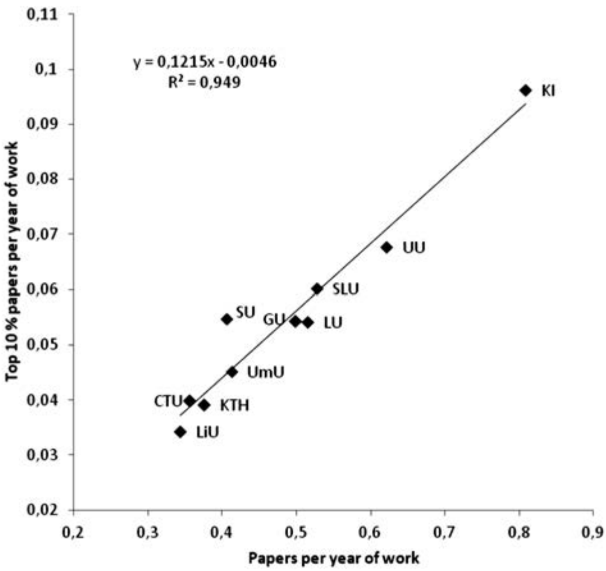


Figure 3 confirms that universities which are characterized with high work productivity also tend to have a high productivity of top 10 percent cited publications. The relationship between overall work productivity and the production of highly cited papers is well described with a linear model. In Figure 3 the model predicts a 0.12 increase in the number of highly cited articles being produced for an increase in the work productivity of 1 produced paper. When comparing the universities, the Karolinska institute (KI) looks outstanding. Independent of which counting method was

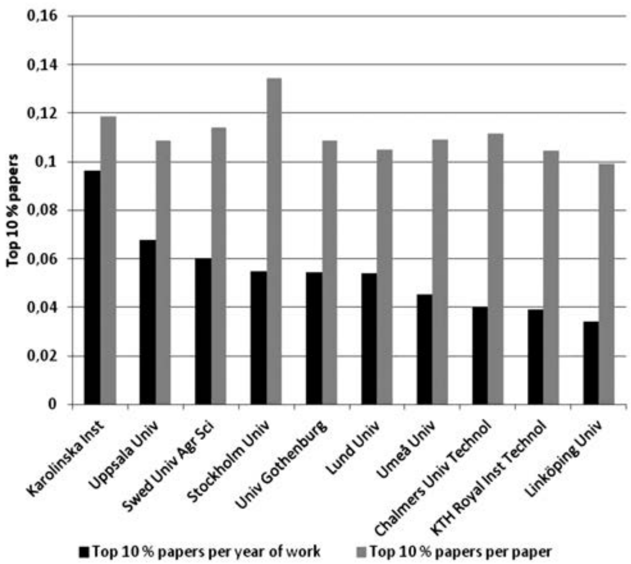
used, researchers at Karolinska institute have the highest work productivity and the highest proportion of highly cited papers.

Comparing university ranks on papers produced per year that are in the top 10% compared with how many papers are in the top 10% out of total paper production

In Figure 4 the citation performance is compared, both in terms of work productivity (i.e. number of top 10% publication per year of work) as well as the share of top 10% publications out of total papers produced. In Figure 4, the universities have been ordered in regards to work related performance. At this stage in the presentation it should not come as a surprise that we will get a different picture concerning the citation performance of universities when we relate this to the amount of work done by the researchers at the universities.

There are some notable differences between the pictures of the Swedish universities given by the different methods. First, there is a change of rank. The most notable difference is that Stockholm University loses its position as the top ranked Swedish university. Second, differences between the institutions are more marked when we relate the amount of articles produced in the top 10 percent with regards to the number of researchers. Karolinska Institute produces significantly more highly cited papers then the other institutions.

Figure 4. The number of papers produced per year who are in the top 10% compared with number of top 10% paper per year of work (fractional counts)



Conclusions

The aim of this paper was to demonstrate that even if the quality of the papers produced by two organizations is the same, a researcher in one organization can still be seen as the higher performer in terms of citations. The results presented in this article demonstrate that there are differences in work productivity among Swedish universities, and when these differences are considered, the ranking of Swedish universities changes. However, it should be noted that what has been considered in this study is not real work productivity, but what could be called Web of Science work productivity, which is something different. We should not put too much emphasis on the rankings presented in this study, but instead focus on the fact that it changed. The lesson learned is that a university organization is made up of researchers, not papers, and the relevant question concerning the performance of a university should be the performance of the researchers employed by the university.

References

Bornmann, L., Leydesdorff, L. & Mutz, R. (2013) The use of percentiles and percentile rank classes in the analysis of bibliometric data: Opportunities and limits. *Journal of Informetrics,* 7, 158–165.

Bornmann, L. & Marx, W. (2013) How good is research really? Measuring the citation impact of publications with percentiles increases correct assessment and fair comparisons. *EMBO Reports*, doi:10.1038/embor.2013.9

Colliander, C. & Ahlgren, P. (2011) The effects and their stability of field normalization baseline on relative performance with respect to citation impact: A case study of 20 natural science departments. *Journal of Informetrics,* 5, 101–113.

Leydesdorff, L., Bornmann, L., Mutz, R. & Opthof, T. (2011) Turning the Tables on Citation Analysis One More Time: Principles for Comparing Sets of Documents. *Journal of the American Society for information Science & Technology,* 62, 1370–1381.

Leydesdorff, L. & Bornmann, L. (2012) Testing differences statistically with the Leiden ranking. *Scientometrics,* 92, 781–783.

Lundberg, J. (2007) Lifting the crown: Citation z-score. *Journal of Informetrics,* 1, 145–154.

Schneider, J.W. (2013) Caveats for using statistical significance tests in research assessments. *Journal of Informetrics*, 7, 50–62.

"A number you just can't get away from": Characteristics of Adoption and the SocialConstruction of Metric Use by Researchers[1]

Gemma Elizabeth Derrick gemma.derrick@csic.es / Instituto de Politicas y Bienes Publicos, Consejo Superior de Investigaciones Cientificas, C/Albasanz 26–28, Madrid, Spain 28037

James Gillespie james.gillespie@sydney.edu.au / Menzies School for Health Policy, The University of Sydney, Sydney, NSW, Australia, 2006

Abstract

This article analyses why researchers decide to adopt Research Evaluation Metrics (REMs) as part of their role and also investigate other factors that lead them to adopt some REMs and not others. This research adapts the approach of Rogers (2003 [1962]) and combines qualitative (interview) with quantitative (survey) methodological approaches to analyse two university departments regarding individual researcher adoption, use and opinions of REMs.

Introduction

The social value of all research is the extent with which it has been adopted by the community. As such, for research evaluation metrics (REMs), the value of any metric indicator produced by the scientometric research community, lies in its utility. REMs will continually fail to offer value to the research community as long as it does not reflect what is valued in research evaluation for that community. Therefore, a measure of excellence for an REM is the extent with which it is disseminated and understood by the research community.

There has been limited previous research into the use and opinions of metrics by researchers. Overall, these studies have had limited success in investigating the knowledge and use of particular indicators. Hargens & Schuman (1990) used social comparison theory to investigate the difference in the use and knowledge of citation counts by sociologists and biochemists. The survey respondents indicated that they believed that citation counts were used for evaluation decisions (hiring, promotion and salary decisions) within the department and that a positive opinion of citations by a researcher was related to the number of citations that researcher had gained (Hargens & Schuman, 1990). Similarly, Buela-Casal & Zych (2012) used a survey of 1,704 researchers to study researcher's opinions of the Journal Impact Factor (JIF). However, despite the large survey size, neither a significantly positive nor negative opinion of the impact factor

1 I would like to acknowledge the Department of Physics and The Sydney Medical School at the University of Sydney for supporting me while I conducted this research.

could be determined. A more sophisticated survey design was employed by Asknes & Rip (2009), with researchers able to provide more in-depth descriptions of their opinions of citations. Here, Asknes & Rip (2009) showed that there was indeed ambivalence about citations among researchers, as found by Buela-Casal & Zych (2012), however researchers held the belief that citations were linked to a researcher's worth. In addition, Asknes& Rip (2009) was able to describe how researchers actively apply themselves to strategies designed to improve their chances of gaining citations. A further study using a survey, showed how researchers shape their research behaviour around REM-maximising strategies believing that they played an active role in university-based evaluations, despite university management stating that this was not the case (Van Noorden, 2010). This previous research indicates how researchers have an interesting relationship with REMs that is perhaps not best captured and described using survey-based methodologies alone.

This article acknowledges that there are different categories of REM users. Primary users have already been acknowledged in the literature and include research managers, governments and specialist REM researchers (Jonkers & Derrick, 2011). However, the way that these users learn and apply REMs is different to secondary REM users such as researchers. Indeed the attitudes, understanding and opinions of secondary REM users will affect the calculation and provision of these REMs for evaluation by these secondary users. It is therefore important to investigate the behaviours and attitudes that are influential for researchers when deciding whether to adopt REMs.

This article investigates why researchers decide to adopt REMs as part of their role and also investigates other factors that lead them to decide to adopt specific REMs but not others. This research adapts the approach of Rogers (2003 [1962]) by combining qualitative (interviews) with quantitative (survey) methodological approaches to analyse the current situation within two university departments regarding the adoption, use and attitudes towards REMs. This research represents one of the first, in-depth analyses of the current use of REMs and the opinions of secondary-users towards their use.

Theoretical Framework

Rogers (2003 [1962]) study of the Diffusion on Innovations was chosen as the theoretical framework for this research. This framework was chosen as it is a common theory used in public health and other population-based research, as well as its identification of mitigating factors as important in influencing how individuals adopt innovations, such as peer and organisational influence (Greenhalgh et al, 2004). The framework allows for an innovation to be described as physical and/or administrative. Diffusion is the process by which *"an innovation is communicated through certain channels over time among members of a social system"* (Rogers (2003 [1962])). A key premise of this model of diffusion is that some innovations diffuse quickly and widely, whereas others do so weakly, or are never adopted. Innovations are adopted by different individuals and spread at different rates in subgroups of individuals. Two main variables are responsible for these differences in diffusion; (1) The Characteristics of the Innovation; and (2) The Characteristics of Adopters.

The Characteristics of the Innovation identified by Rogers (2003 [1962]) as important for a successful diffusion are described in the Table 1 below. These characteristics of the innovation for adoption were used to formulate the survey questions in Stage 1 of this study. Further, these characteristics were tested for validity in relation to REMs and investigated further during the Stage 1 survey. For the purposes of this article, only the attributes: Relative Advantage and Compatibility are described and discussed. Rogers (2003[1962]) described the process of innovation adoption by individuals as a normal, bell-shaped distribution, with five adopter categories related to the Characteristics of Adopters. These categories were identified as; (1) Innovators (2.5%); (2) Early Adopters (13.5%); Early Majority (34%); Late Majority (34%); and Laggards (16%). Rogers (2003[1962]) proposed that identifying these categories of adopters, specific intervention strategies could be designed to aid the diffusion of a specific innovation.

Finally, for individuals making the decision to adopt an innovation or not, five steps have been identified. These stages include, *(1) Knowledge* (where the individual is first exposed to an innovation but lacks information about the innovation); *(2) Persuasion* (where the individual is interested and actively seeks information/detail about the innovation); *(3) Decision* (where the individual weighs the advantages/disadvantages of using the innovation and decides whether to adopt or reject); *(4) Implementation* (where the individual experiments with the innovation); and *(5) Confirmation* (where the individual finalizes their decision to continue using the innovation). The decision-making processes of potential adopters are also influenced by many other factors in the context, environment, or system in which the process takes place (Greenhalgh et al, 2004). This decision making process as well as the role of any peer- or organisational-based influence on individuals to adopt, were used as the basis for interview questions and further analysis. By employing a more qualitative approach to the process of REM adoption, it is hoped that these more nuanced factors of adoption and opinions towards REMs may be captured

Table 1. Characteristics of Innovations related to Diffusion

Attribute	**Definition "…the extent to which the innovation…"**
Relative Advantage	*"…is better that the idea it supersedes."*
Compatibility	*"…is consistent with the existing values, part experiences and needs."*
Complexity	*"…is easy to use and learn."*
Trialability	*"…is permitted to be experimented with prior to adoption."*
Observability	*"…is visible to others*

Methods

Definition of Research Evaluation Metrics (REMs)

Researchers were provided with the following definition of a Research Evaluation Metrics to assist them with the survey and optional follow up interview.

"An REM is a metric or measure used to quantitatively describe, evaluate, assess or rank research, journals, researchers, organisations or countries."

For the purposes of this study, the definition of a Research Evaluation Metric was purposely kept broad. This was done for two reasons. First, the original nature of this research made it necessary not to prejudge which metrics were adopted by researchers, and which ones had not. During the survey, described below, participants were given the opportunity to nominate those REMs they had adopted, and describe where they were being used. Second, the study primarily aimed to investigate the general opinions surrounding the use of REMs, rather than to focus on the researcher's opinions of the various advantages and disadvantages of a specific metric. The latter has been attempted previously, with no conclusive results (Buela-Casal & Zych, 2012). As such, it was necessary for this study to attempt a different approach and employ a broad definition of an REM.

Methodological design

The research evaluation field has traditionally used quantitative methods to examine the visibility, use, diffusion and influence of journal publications. Despite their prominence, quantitative methods are not as well suited to measure other complex aspects of the diffusion of research such as the motivations of the adopter, organisational change, research leadership and the process of implementing evaluation guidelines, which are also critical issues in research evaluation management. These more nuanced aspects of the value of REMs and their diffusion within research communities require additional non-traditional methods. The methodology presented here allows us to extend the understanding of researcher's knowledge, use and opinions of REMs beyond the capabilities of a single survey, bibliometric or other quantitative sources of information. By employing alternative methods, this research aimed to investigate the understanding of the social construction of metric use beyond the capabilities of previous studies that have reported either "ambivalence" (Asknes & Rip, 2009) or "neither positive nor negative" (Buela-Casal & Zych, 2012).

As such, this research adopts a mixed-methods approach combining Likert scale surveys with in-depth semi-structure interviews. The research framework included two, interlinking stages: Stage 1, Survey; and Stage 2, in-depth, semi structured interviews. A description of each stage, how they were analysed and then triangulated, is outlined below.

Stage 1: Survey

Survey participation was restricted to researchers with at least 3 years post-PhD from the Departments of Physics and Medicine at the University of Sydney, Australia. The departments of Physics and Medicine were selected to participate as both fields are well represented by current REMs. It is therefore valid to assume that the researchers within these departments would have had more exposure, discussion and time to experiment with various REMs. Permission from each Departmental Head was received prior to data collection took place. This permission also increased the legitimacy of the research project and was used to increase response rates.

The survey collected a combination of basic demographic information of participating researchers, information about current use of REMs, as well as a 5-point Likert scale survey of researchers' opinions of REM use. The questions were based around a number of factors influencing adoption as specified by Rogers (1983) such as Relative Advantage; Compatibility; Complexity; Trialability; Observability; Ease of Use; and Social Approval. Questions were divided into five separate parts asking questions regarding different aspects of REM adoption including: Characteristics of adoption; My use of REMs; My attitudes towards REMs; My research and REMs; and My REMs and my peers.

A total of 200 researchers completed the survey with a further 49 agreeing to take part in a follow-up, semi-structured interview (described below). The survey allowed us to identify current adopters of REMs to interview during Stage 2, which concentrated on the stages (process) of adoption of REMs. Surveys were analysed using frequency analysis using STATA and were cross referenced with the themes identified in the interview analysis also described below.

Stage 2: Interviews

Following the survey (described above), interviews were organised with those researcher identified in the survey as having adopted REMs. During the survey, researchers were specifically asked to identify which REMs they used and where they used them, thereby allowing the identification current REM adopters to be interviewed. This allowed for the interview to be designed around the specific REMs adopted by the participant as well as allow a full investigation of the stages of adoption as first described by Rogers (2003 [1962]). This selection process also ensured that the interview analysis and identification of themes (described below) was not complicated by conflicting results from participants who had not adopted REMs

A total of 26 researchers took part in a follow up semi-structured interview, either in person or by telephone, about their use of nominated REMs and opinions about the current use of REMs during evaluation. The interview questions and related prompts were modelled around the characteristics influencing adoption (Relative advantage, Compatibility, Complexity, Trialability and Observability); the process of adoption (Knowledge, Persuasion, Decision, Implementation and Confirmation) as described by Rogers (2003 [1962]) as well as the possible causes for non-adoption of other metrics. Finally, questions exploring the role of organisational and/or peer influ-

ence in the decision to adopt (or not) were also included. Interviews were continued until a satisfactory level of "saturation" was obtained.

Table 2. Coding domains and descriptions for 2nd round interview analysis

Domain	Description
Strategies/methods/use or REMs and the professional behaviours that researchers use REMs.	Ways of doing things; Communicative practices and the strategies that are central to these.
Judgements/values/meanings/beliefs/ discourses about REMs and research evaluation.	Ways of assessing and representing research excellence, ideological/paradigmatic constructions, contested viewpoints about what is valued, legitimisations, value judgements about use and researchers who use REMs. Judgements about benefits/limitations of peer review and of REMs.
Interactions/Relationships	Networks, collaborations, supports, adversaries. Includes influences, hierarchies, leverage, inequalities, vested interests. Conflicts/battles eg. Between researchers, evaluators and organisations.
Cultural Practices	This includes social processes, rules, norms and conventions, expectations.

Interview analysis

Interviews were fully transcribed and analysed using Atlas-ti. Interview analysis was based on a semi-structured, cognitive-based, grounded theory method that involved two separate rounds of coding. The first round of coding included the identification of responses in line with Rogers (2003 [1962]) five main characteristics of adoption described above and in line with the design of the interview schedule. Questions regarding additional characteristics identified in previous diffusion research (Ease of Use, Status/Social approval, Image, Usefulness, Divisibility, Voluntariness and Communicability), were also analysed. Once this first round of coding was complete, in-vivo codes were further identified within and between first round codes in order to explore themes and interlock related themes. The in-vivo codes used for this second round of analysis was based on a number of domains described in Table 2.

Throughout the interview data analysis, a process of constant comparison was employed. This meant that the codes identified were continually refined, enriched, reorganised and tested for significance. This model of analysis is characteristic of the cognitive, grounded-theory approach to qualitative data analysis (Charmaz, 2006), and ensured that the themes that emerged are a reliable and replicable reflection of the case under investigation. In addition, sufficient material was analysed to ensure that categories were saturated, that is, that all codes appeared to fit under on or more existing categories and all concepts were fully described and well-understood. Our categories and concepts were then organised under a number of general headings that form the structure of the results section below.

Results

Adoption and current use of REMs

Basic information was collected about the main REMs adopted by researchers and where these are currently being used. The results are shown in Table 3. The Journal Impact Factor, despite widespread discussion regarding its misgivings and inadequacy as a quality measure, was still a popular measure used by researchers. Not surprisingly, researcher used this metric about all other REMs when deciding where to publish (99.0%), a consideration reflective of the reward-maximising model of journal submission (Luukkonen, 1992). This model outlines how researchers make a hierarchical list of potential journals and submit their article to each journal on the list, starting from the top, until publication acceptance is achieved. The results below suggest that the JIF is a major consideration when constructing this hypothetical list. The JIF was also shown to be used in a number of different professional situations including for competitive purposes such as part of grant applications (80.0%) and promotion applications (79.7%), as well as for self-promotion activities such as on their Curriculum Vitae (71.1%) and media purposes (81.3%). Interestingly, researchers also indicated that they used the JIF to represent themselves in Research Evaluation Frameworks (90.9%) despite clear government policies stating that the JIF should not be used for these evaluation purposes (NHMRC, April, 2010). Other REMs used widely included the H-index and Citations per paper. These REMs were used widely in a number of competitive and self-promotion purposes. In addition, a smaller number of alternative REMs were also indicated, however, the three metrics: JIF; Citations per paper; and the H-index, were shown to constitute the majority of REMs currently adopted by these researchers.

Table 3. Types of REMs adopted and current use by researchers (%)

	JIF	Citations per paper	h-index	Other
Curriculum vitae	71.1	66.7	72.2	8.8
Grant applications	80.0	73.3	63.8	10.6
Promotion applications	79.7	82.3	73.4	11.4
Position reviews	78.9	64.9	66.7	7.1
Media purposes	81.3	12.5	18.8	25.1
Biography information	60.5	52.6	55.3	7.9
Professional webpage	61.5	42.3	61.5	11.5
Deciding where to publish	99.0	5.7	2.9	2.0
Research Evaluation frameworks	90.9	60.6	51.5	3.0
Evaluating other researchers	68.9	65.6	65.6	7.8
Evaluating research quality	75.4	72.3	41.5	15.3

Interview results

Peer-Generated Adoption of REMs

The interview results also showed that researchers have voluntarily adopted a number of REMs and use them in a variety of ways including on their C.Vs and grant applications, despite the absence of a formal requirement to provide them. Researchers independently introduced themselves to a variety of new metrics and adopting those metrics that were perceived as: vital for successful grant applications (58.7%); or increased the success of promotion applications (62.0%). These metrics were felt to provide a simple representation of complex information about a researcher's experience or, in other words, their "value" as a researcher.

"…when I want to sell myself, you know, when I have written promotion applications in the past, on grant applications, you try to make it clear upfront what my value is…"

Relative to the belief that "everybody uses metrics" was the widespread belief that it was essential to include REMs in grant and promotion applications. Indeed, researchers increasingly used REMs for these purposes and researchers expressed their belief that providing metrics was a key factor in guaranteeing success of an application. The survey results indicated that 62% of researchers agreed or strongly agreed with the statement "I include REMs in funding or promotion applications in order to increase my chances of success." This belief that REMs were widely used by the majority of researchers, only served to accelerate REM adoption among research peers.

"It really latched on and then people were really – continued to use it as a benchmark. And it kind of surprised me to be honest. And it's really because everybody does it, that I do it. I mean, I think that's the critical thing is this sort of widespread acceptance of something as a benchmark."

This bottom up, peer-generated pressure to include REMs on applications was further fuelled by reports from researchers who had also acted as evaluators. These researchers reflected how applications that included REMs, by the merit of REM simplicity and supposed, objectivity, were viewed more favourably during evaluations.

"…if they put – how many papers they published, how many citations for each of those papers, what their h-factors are [then] I feel more positively predisposed to them because they provided me with information in a quick, easy to understand format that I can, that I can get at…"

In addition to including REMs, to make the task of the evaluators easier and include REMs to be comparable to their competitors, was the risk researchers felt when not including them. This was felt to be the equivalent of potentially hiding information from evaluations and further emphasised the perceived necessity in including REMs on grant and promotion applications,

"If you haven't got it and you're not putting it in there, and I just think 'what have you haven't got?' and therefore you are going to get marked down."

Indeed, a large amount of peer influence that forced researchers to use REMs on grant and promotion applications was related to the belief that providing metrics was an accepted norm and that by not providing REMs, their absence sad more about the researcher's value than when a researcher provided them.

"I put it there [REMs on grant applications] because, you know, somehow if you don't put it there then it might make people wonder, 'why didn't you put it there'."

This was exacerbated by researchers stating that, when they acted as evaluators, they specifically searched for metrics on applications to guide their evaluations. The reasons stated for this behaviour ranged from time restrictions, to a need for a "yardstick" to measure and compare researcher value.

"…if they put how many papers they published, how many citations for each of those papers, what are their h-factors. I instantly feel more positively predisposed to them because they provided me with information in a quick, easy to understand format that I can – that I can get at, that I think has – personally, I think has some validity."

The same applies to specific metrics, such as the h-index;

"…it is a good metric, because it gives you a bit of a yardstick to be able to judge people on in terms of their track record and so on…"

In some circumstances, researchers remarked how that in the absence of a provided metric in applications, they would calculate the REMs they favoured for that application. This practice is alarming and further research is needed to determine how widespread this practice is among research evaluators. A more detailed exploration of the characteristics that appealed to researchers is outlined below in regards to the relative advantage of REMs.

The Relative Advantage of REMs

In terms of metrics, the relative advantage of REMs refers to the advantage the new innovation (REMs) has over the old innovation (peer review). Specifically, whether there was an agreement that REMs could replace peer review. Perhaps not surprisingly, of all of the potential identified characteristics, the perceived relative advantage of REMs was the most contentious. The results showed an incomplete resolution as to whether REMs presented a complete advantage over the more traditional, peer review.

REMs were favoured by researchers because they were perceived as simple, objective measures that provided beneficial summary information about a researcher and/or research. When discussing the benefit of simplicity of REMs, researchers described the benefit of REMs of providing a quick *"approximation"* of a researcher's professional background.

"And it's not a perfect metric by any stretch but it's a simple one. I mean, that gives you sort of a first approximation I suppose, of somebody's track record."

No distinction between REMs that were used to describe research as separate from researchers. Instead, researchers referred to how it didn't matter what the REM meant, but if in the case of rankings, what mattered was the position of the researcher and/or research and how they looked relative to their colleagues.

"Well trying to be honest, [it] doesn't really, you know, whether you're number one or whether you are number twenty, if you are up there with that, you know big US researchers and institutions, that's only where they need to go…"

REMs were also perceived to be *"objective"* measures of research value. The definition of objective in this regard does not include a measurement that is free from tampering or misuse, rather REM appearance as quantitative measures and the ability of researchers and evaluators to rank according to these numbers, appealed to researchers as *"objective"*. In addition, compared to peer-review, a naturally subjective evaluation tool, REMs were perceived as free from the biases influencing peer-review generated evaluations.

"…for me again, the quantitative background, I guess they impressed me. You know, they're trying to do something here in a rather more objective way is a bit less of, you know, personal view of that and less subjective…"

Another research independently echoed the predisposition towards metrics based on their quantitative appearance, and therefore presents a seemingly objective measure of a researcher's track record. An REM was perceived as an additional piece of information that could be provided in research grant applications in order to present their value, quickly to evaluators and positively influencing the peer-review evaluation outcome.

"It's a good, quantitative piece of evidence that you can include as to your track record."

Despite a number of REM characteristic favoured by researchers, there were a number of characteristics that made researchers cautious about their overuse and the potential loss of peer review. These included, especially for medical researchers, an understanding that REMs did not reflect the wider importance of research or their activities as a researcher.

"So these metrics are only useful to a certain extent and they're kind of a broad brush stroke."

However, researchers also gravitated away from using specific REMs and towards others, because of the way that the alternative metrics represented them. Researchers, though acknowledging that REMs represented a competitive advantage in the research "game";

"…and it's a game where everybody wants to look good, right?"

And understanding that there was a lack of guidelines about what REMs are to be used in grant and promotion applications, would automatically assume metrics that made them look better and then promote the benefits of these alternative metrics to their peers. The survey results also confirmed with this, 46.1% of researchers agreeing or strongly agreeing with the statement "I would not use an REM if it represented me badly."

"…the person that you often talk to want you to use a different index which they look better in, right?"

Researchers were cautious about this, less the new metric represent them poorly, but this was balanced with a curiosity-driven experimentation with new REMs.

"…we used it just kind of for fun, to speculate and just to play around with different metrics just to, you know, just for fun basically."

However, this was not widespread, with only 27.4% of researchers indicating that they had "Experimented with different metrics". In addition, 63.5% agreed or strongly agreed with the statement "REMs encourage researchers to 'game' or 'cheat' the system".

Out of all the characteristics of REMs that appealed to researchers, the strongest characteristics was when the researcher's existing impressions of excellent research was reinforced by corresponding excellent metrics. In these cases, researchers reflected on situations where they had come across articles or researchers that they value and had considered their REMs as equivalent to their personal appreciation of that researcher and their research. However, this was the same case with researcher's own opinions about their research contributions.

"…people are, the average person is prepared to sort of just accept the number that's given, especially if it is higher."

If researchers felt their work was adequately represented by metrics, they felt more predisposed towards the widespread use and relative advantage of REMs. However, the agreement between a researcher's personal opinion of research excellence and the corresponding metrics did not always occur. A total of 42.8% of researchers agreed with the statement, "I am sceptical about REMs".

"I've read some papers which were really useful to me and, of course, I have cited them and I'm surprised to find that they're not cited very often."

This was for a number of reasons, including a broader definition of research excellence beyond the limitations of REMs;

"…I suppose I've got my own mental metric that I use to assess whether research is good or not…"

Also included was an understanding of the restrictions of REMs as measures of research excellence,

"...because it [REM] is not actually an indicator of what you do."

In fact, the corresponding survey results show that only 7.1% of researchers agreed or strongly agreed that "The REMs I use now are better than using peer review to evaluate research."

In these cases, we expected that this mismatch would result in an increased scepticism in REMs and a corresponding decrease in their use by these researchers, but this was not the case. On the contrary, the widespread belief that REMs were essential for success seemingly reinforced researcher use of REMs. However, this mismatch between the personal definition of research value and REMs, was identified as a major factor limiting the perception of the relative advantage of REMs over peer review (old innovation), and therefore affecting their complete adoption.

Finally, REMs were perceived by researchers as *"...a number you just can't get away from"*, its increased and unregulated use in research evaluations make it a powerful tool with which researchers can engage. Specifically, the motivation behind this engagement is the perception of REMs as a marketing tool rather than as a benchmarking/evaluation tool. In relation to this view, researchers preferred metrics that they could *"control"* or *"talk around"*. These tended to be individual metrics such as the h-index rather than fixed-metrics like the journal impact factor, or institutional level metrics such as rankings.

Discussion

These results have important implications for how universities, their departments and research offices, as well as governments evaluate research in official evaluation exercises. To date, governments have either been reluctant or have changed initial plans to implement metrics into evaluation exercises. This has mainly in response to a vocal, researcher-led opposition to REM informed research evaluation as well as a predominant belief that peer review and REM-informed evaluation are mutually exclusive. However, this research shows that researchers are active, independent experimenters with metrics who provide these metrics on applications to positively influence evaluations.

Despite this active experimentation, researchers are yet to be convinced to the complete relative advantage of REMs, one factor limiting the more widespread adoption Rogers (2003 [1962]). Researchers were aware of the advantages of REMs and were drawn to their use because of their ability to communicate complex information simply and, seemingly, *"objectively"*. Despite researcher awareness of REM disadvantages, the belief that REMs are essential for grant research success, combined with the belief of their widespread use between competitors (colleagues), by peers who also act as evaluators, and the preference expressed by these evaluators about the REMs use, fuels REM adoption.

The use of REMs described in this article, further illustrates concerns over the perceived objectivity of REMs. Arguably, a researcher opinion that REMs present an objective reflection of research presents another factor limiting the use of REMs more widely for research evaluation. The results shown here indicated that not all metrics provided by researchers on grant and promotion applications are free from tampering. In addition, the perception of objectivity by researchers and research evaluators in the absence of specialist advice from the bibliometric community, will only limit further adoption if evaluations are based on REMs that are not appropriately employed. Instead of promoting metrics as a tool, it is recommended that organisations and governments focus on providing a system for independent provision of REMs including guidelines for the calculation including specified time periods, databases and types of publications.

References

Aksnes, D. W., & Rip, A. (2009). Researchers' perceptions of citations. Research Policy, 38(6), 895–905.

Buela-Casal, G. Zych, I. (2012) "What do the scientists think about the impact factor." *Scientometrics*. 92: 281–292.

Charmaz, K. (2006). Constructing grounded theory: A practical guide through qualitative analysis. Sage Publications Limited.

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Quarterly*, 82(4), 581–629.

Hargens, L.L., Schuman, H. (1990) "Citation Counts and Social Comparisons: Scientists' Use and Evaluation of Citation Index Data." *Social Science Research*. 19: 205–221.

Jonkers, K., & Derrick, G. E. (2012). The bibliometric bandwagon: Characteristics of bibliometric articles outside the field literature. *Journal of the American Society for Information Science and Technology*, 63(4), 829–836.

NHMRC (April, 2010) "NHMRC removes Journal Impact Factors from peer review of individual research grant and fellowship applications." *NHMRC Media release*. Available from http://www.nhmrc.gov.au/_files_nhmrc/file/grants/peer/impact%20factors%20in%20peer%20review.pdf.

Luukkonen, T. (1992) "Is Scientist Publishing Behaviour Reward-Seeking?" *Scientometrics*, 24(2): 297–319.

Rogers, E.M. (2003 [1962]) "Diffusion of Innovations". 5th Edition. *Free Press*: New York.

Van Noorden, R. (2010) "Metrics: A Profusion of Measures." *Nature*, 465: 864–866.

Measuring the Radicality of Innovation: A Radicality Index Based on Patent Citation Patterns

Burak Dindaroğlu burak.dindaroglu@ieu.edu.tr / Department of Economics, zmir University of Economics, Sakarya Cad. No: 156, Balçova-zmir (Turkey)

Abstract

This paper develops a simple and novel index of innovation radicality that can be computed using available large scale data on patent citations. The measure draws on the idea that radical innovation perturbs the direction and flow of innovation, and this disturbance should be visible in patent citation patterns. I construct the index for all patents granted by the USPTO between 1976 and 2006 and examine its relationship to known technological indicators. Then, I study the firm-level determinants of radical innovation. Finally, I examine the performance of the index in predicting firm performance compared to common technology indicators, such as citations, patent counts and indices of patent quality.

Viewing innovation as a cumulative process, each patent represents an innovation that builds on the previous innovation base and contributes to the invention of its antecedents. In other words, each patent occupies a unique place in the "flow" of innovation over time. The immediate precedents of the patent are observed via citations made by the patent, i.e., backward citations, and its immediate antecedents are observed via citations made to the patent, i.e., forward citations. I propose that the radicality of innovation should be highly correlated with the degree of dissimilarity between backward and forward citations of the patent associated by the said innovation. In other words, I imagine a patent as a bridge between its backward and forward citations. The dissimilarity between the precedents and antecedents of a patent, then, measures to what extent the patent "shifts" the flow of technology over time.

This idea is put into practice by building on the so-called "technological proximities" (Jaffe, 1986) which are commonly used to position firms in technological space in the study of knowledge spillovers. This is calculated as the uncentered correlation between the "technological position vectors" of economic units, which are vectors that contain the number of patents of the firm that are classified in technology class *k* in their *k*-th elements. Similarly, I construct two such vectors for a given patent, one containing the number of citations *made by* the patent to technology class *k* in its *k*-th element, and another containing the number of citations *received by* the patent from technology class k in its k-th element. The index proposed by the current paper is the distance (uncentered correlation) between these two vectors.

For a patent to be deemed radical, I also require that (1) it belongs above the 90th percentile of the distribution of patent quality (measured by the Lanjuow and Schankermann (2004) quality

index) in its narrow technology class, (2) it has received at least 10 citations, and (3) its generality (Trajtenberg et al, 1997) is larger than its originality. The first requirement ensures that the patent in question is among the highest quality patents in its technology class, while the second ensures that the patent has had an important impact on future innovations. The third requirement ensures that the "shift" in innovation flow due to the patent in question occurs in a way that is consistent with our definition of radical innovation.

The index exhibits properties that are commonly associated with radical innovation. Among all patents that satisfy conditions (1) and (2) above, more radical patents receive a larger number of citations holding patent quality (LS) and other patent characteristics constant. However, holding patent quality constant, radical innovations receive a lower share of their citations following the first 5 years after patent grant, but a higher share between the 6^{th} and 10^{th} years after grant. This is consistent with the prediction that radical innovation takes a longer time to be understood, accepted and adopted. Higher radicality is also associated with fewer claims and a lower self-citation rate, despite indicating higher quality overall.

Firm level analyses indicate that a firm's radical innovations have a large predictive power in market value equations above and beyond R&D, patents and citations. Also, studying the firm-level determinants of radical innovation, I find that radical innovations are produced more often by R&D and capital intensive firms. Larger firms have a higher propensity to produce radical innovation, but the share of radical patents among a firm's patent portfolio does not depend on its size.

Ongoing work includes seeking external verification of the index by looking at its ability to identify innovations that are deemed radical by industry experts and previous studies that used restricted samples. For this purpose, I take advantage of the study by Abernathy, Clark and Kantrow (1983), which compiled a complete list of all significant inventions in the auto industry and scored each invention according to a number of characteristics, including a ranking based on radicality. Preliminary findings based on a number of individual inventions are promising. This set of analyses requires the identification of patents that are associated with the *inventions* listed by Abernathy, Clark and Kantrow (1983), which is part of the ongoing work related to this project. A comparison of the performance of the index to the radicality index of Dahlin and Behrens (2005) is also planned for the final version of the paper.

References

Abernathy, William J., Kim B. Clark and Alan M. Kantrow (1983) Industrial Renaissance: producing a Competitive Future for America, Basic Books Inc., New York.

Dahlin, Kristina D. and Dean M. Behrens (2005), "When is an invention really radical? Defining and measuring technological radicalness," *Research Policy* 34: 717–737.

Jaffe, Adam (1986), "Technological Opportunity and Spillovers of R&D: Evidence From Firms' Patents, Profits, and Market Value", *The American Economic Review* 76: 984–1001.

Trajtenberg, Manuel and Rebecca Henderson (1997), "University versus Corporate Patents: A Window on the Basicness of Invention", *Economics of Innovation and New Technology* 5(1): 19–50.

Individual-level evaluative bibliometrics – the politics of its use and abuse

Wolfgang Glänzel wolfgang.glanzel@kuleuven.be / ECOOM, KU Leuven

Jochen Gläser Jochen.Glaser@ztg.tu-berlin.de / ZTG, TU Berlin

Ismael Rafols i.rafols@sussex.ac.uk / Ingenio (CSIC-UPV) & SPRU, Sussex U.

Paul Wouters wouters.paul57@gmail.com / CWTS, Univ. Leiden

The last five years have seen an upsurge of attempts to measure individual research performance by using bibliometric indicators. The demand for such individual-level evaluative bibliometrics has dramatically increased in the context of management reforms that make managers responsible for research performance. 'Amateur bibliometrics' – the unreflected use of ready-made indicators offered by the owners of bibliometric databases – abounds. At the same time, bibliometricians try to develop indicators that can be applied to individuals without always paying sufficient attention to the methodological and ethical problems involved. There is an obvious danger that this process will be accelerated as policy and management of science 'buy into' this new development.

These trends raise methodological, ethical and political questions. The methodological question to be addressed by bibliometricians concerns the conditions under which individual-level bibliometrics can produce information that is sufficiently valid and reliable to be used in evaluations. The political question of how expectations for evaluations can be met without distorting the science system needs to be jointly answered by the bibliometrics community, science policy and managers of research organisations. Finally, there is also the question whether the bibliometrics community should develop a guideline of standards and a code of ethics, a practice that is common for most professions.

We think that these questions need to be urgently discussed, and propose to hold a plenary session on the political and ethical issues related to individual-level bibliometrics at the STI conference in September 2013. Our aim is to initiate a discussion that leads to policy recommendations about the use and interpretations of individual-level bibliometric evaluations and to a code of conduct for the bibliometric profession.

Presentations:

(1) Brief report on the ISSI 2013 plenary on methodological aspects of individual-level bibliometrics (Wolfgang Glänzel, ECOOM, KU Leuven)
(2) Experiences with individual level bibliometrics in research management in the last decade (Paul Wouters, CWTS, Leiden University)

(3) Decision situations in science policy and management and the need for individual-level bibliometric evaluations (Marc Luwel, Hercules Foundation, Belgium)
(4) Ethical Aspects of Doing Evaluative Bibliometrics (Jochen Gläser, ZTG, TU Berlin)

The Application of Citation-Based Performance Classes in Institutional Research Assessment

Wolfgang Glänzel Wolfgang.Glanzel@kuleuven.be / Centre for R&D Monitoring (ECOOM) and Dept. MSI, KU Leuven (Belgium) / glanzw@iif.hu / Dept. Science Policy & Scientometrics, Library of the Hungarian Academy of Sciences, Budapest (Hungary)

Bart Thijs Bart.Thijs@kuleuven.be / Centre for R&D Monitoring (ECOOM) and Dept. MSI, KU Leuven (Belgium)

Koenraad Debackere Koenraad.Debackere@kuleuven.be / Centre for R&D Monitoring (ECOOM) and Dept. MSI, KU Leuven (Belgium)

Abstract

The analysis of the high end of citation distributions represented by its tail provides important supplementary information on the citation profile of the unit under study. The proposed application of characteristic scores and scales provides four performance classes that are not sensitive to ties and ensures the seamless integration of measures of outstanding and even extreme performance into the standard tools of scientometric performance assessment. While the applicability of this method was already demonstrated for cross-national comparison (Glänzel, 2013b), the present study focuses on its application at the institutional level.

Introduction

One of the objectives of previous studies (Glänzel, 2013a,b) was to analyse to what extent the tail of scientometric distributions is in line with the 'head' and 'trunk' forming the major part of the distribution and to what extent 'outliers' might be responsible for possible deviations. Two important observations are relevant in this context. One solution proposed in these studies was to use tail indices as a supplement to traditional citation-based performance indicators, such as the share of uncited papers to the mean citation rate. The analysis of the tail, which was based on ordered or ranked observations, can practically be uncoupled from the overwhelming rest of the empirical distribution. Most studies of the tail of scientometric distributions proceed from a Pareto model. The estimation of the tail parameter can directly be obtained from subsets of order statistics. This approach even allows to construct confidence intervals for its estimator. Nevertheless, the estimation of the tail index remains rather problematic since most methods are still sensitive to the cut-off point for the tail. Since already minute changes of the tail parameter might have significant consequences in an evaluative context, the recommendation in the study by Glänzel (2013a) was to favour a parameter-free solution for the assessment of outstanding performance. This might also help avoid parameter conflicts resulting from estimating parameters on the basis of head and trunk of the distributions, on one hand, and from their tail, on the other hand.

Therefore, a "reduction" of the original citation distribution to performance classes on the basis of *Characteristic Scores and Scales* (CSS) introduced by Glänzel & Schubert (1988) was proposed as an alternative parameter-free solution. Taking into account that citation standards considerably differ in the various disciplines, the method was developed for *individual subjects*. The classes obtained from this method can be applied to the comparative analysis of the citation-impact profiles of given units amongst themselves as well as with the reference standard in the given subject. It has been stressed that the calculation of a "single" indicator over these classes is not suitable as this would reduce the gained added value and thus destroy the advantages of the method. However, it has also been shown that the application to combinations of different disciplines is indeed be possible Glänzel (2013b). The study has furthermore demonstrated robustness of the method for combinations of disciplines with respect to the publication year and the citation window at the level of national research performance. In the present paper we extend the analysis to a lower level of aggregation, particularly to the assessment of research institutions. At this level the number of publications per unit is considerably lower than at the national level but more important is that we expect to observe more diverse research profiles. In particular, some institutions have a specialised profile while others are truly multidisciplinary in their research activities. The aim of the paper is therefore to also demonstrate the robustness of the method at the meso level and its independence of institutional publication profiles.

A parameter-free solution using Characteristic Scores and Scales (CSS)

An alternative to the tail analysis supplementing standard indicators is the "reduction" of the original citation distribution to a distribution over some essential performance classes, including one or more particular classes corresponding to the high end of performance, i.e., to the tail of the original distribution. A solution using six classes has already been suggested by Leydesdorff et al. (2011). According to their model, a pre-set set of six rank percentages is calculated on the basis of the reference distribution. Individual observations are then scored according to the percentage the publications in question belong to. Two particular problems arise from this approach, namely the arbitrariness of pre-set percentiles and the ties in both the reference distribution and the observations.

Characteristic Scores and Scales (CSS)

Another solution is based on the method of Characteristic Scores and Scales (CSS) proposed by Glänzel & Schubert (1988). Characteristic scores are obtained from iteratively truncating a distribution according to conditional mean values from the low end up to the high end. In particular, the scores b_k ($k \geq 0$) are obtained from iteratively truncating samples at their mean value and recalculating the mean of the truncated sample until the procedure is stopped or no new scores are obtained. The theoretical values ß$_k$ for a random variable X are simply defined using the following equation.

$$\beta_k = \begin{cases} 0 & ,\text{if } k = 0 \\ E(X \mid X \geq \beta_{k-1}) & ,\text{if } k > 0 \end{cases} \quad (1)$$

The empirical values b_k are then their estimators based on (truncated) means, where we first put $b_0 = 0$. b_1 is then defined as the mean of the original sample. The procedure is usually stopped at $k = 3$ since the number of papers remaining in the subsequent truncated sample might otherwise become too small. The *k*th citation class is defined by the pair of threshold values $[b_k{-}1, b_k)$ with $k > 0$. The last and highest class is defined by the interval $[b_k, \infty)$, with usually $k = 3$. The number of papers belonging to any class is obtained from those papers, the citation rate of which falls into the corresponding half-open interval. The problem of ties known from the percentile approach does not emerge here since all papers can uniquely be assigned to one single class. In earlier studies the resulting four classes were called *poorly* cited (if less cited than average), *fairly* (if cited above average but received less citations than b_2), *remarkably* cited (if received at least b_2 but less than b_3 citations) and *outstandingly* cited (if more frequently cited than b_3). In the present study 'Class *k*' (k =1, 2, 3, 4) is used instead for the sake of convenience.

As all location parameters, characteristic scores, too, are very sensitive to the subject field and the citation window. b_1 is, by definition, the mean value of the empirical citation distribution; all other scores are conditional means that depend on this initial value (see Eq. (1)). Therefore, characteristic scores should not be used for comparison across subject areas. In order to demonstrate this effect we have selected 20 out of the 60 subfields in the sciences according to the Leuven-Budapest classification scheme. Furthermore, all journal publications indexed as article, letter, proceedings paper or review in the 2007 volume of Thomson Reuters' Web of Science (WoS) have been selected and processed. The five-year citation window (2007–2011) has been used. Scores b_k ($k = 1, 2, 3$) and shares of papers in the four classes are given in Table 1.

Table 1. Characteristic scores and CSS-class shares of publications of publications in 2007 for 20 selected subfields according to the Leuven-Budapest scheme [Data sourced from Thomson Reuters Web of Knowledge]

Subfield*	b_1	b_2	b_3	Class 1	Class 2	Class 3	Class 4
A2	6.43	13.80	21.97	65.2%	22.6%	8.1%	4.2%
B1	16.75	39.24	79.61	69.4%	22.5%	6.0%	2.1%
B2	23.05	58.33	116.72	72.0%	20.2%	5.6%	2.2%
C1	9.37	22.04	40.48	68.2%	22.5%	6.6%	2.7%
C3	11.22	24.68	42.04	67.4%	22.2%	7.5%	3.0%
C6	8.21	23.67	51.24	73.5%	19.5%	5.3%	1.8%
E1	5.04	14.75	29.83	73.7%	18.8%	5.5%	2.0%
E2	4.71	11.90	21.97	68.2%	21.7%	7.0%	3.1%
E3	6.57	17.82	34.00	70.7%	20.2%	6.3%	2.9%
G1	15.55	38.35	74.51	70.1%	21.4%	6.3%	2.2%

... Continuation Table 1

Subfield*	b_1	b_2	b_3	Class 1	Class 2	Class 3	Class 4
H1	5.21	14.36	29.83	72.3%	20.3%	5.4%	1.9%
I1	13.52	34.87	69.24	70.2%	21.3%	6.2%	2.3%
I5	16.24	41.52	84.74	71.9%	20.4%	5.4%	2.2%
M6	11.50	28.31	51.81	68.9%	21.6%	6.5%	3.0%
N1	15.28	35.38	64.73	69.1%	21.7%	6.4%	2.8%
P4	7.25	17.71	32.75	69.6%	21.2%	6.7%	2.4%
P6	7.27	20.05	43.89	72.4%	20.7%	5.3%	1.7%
R2	10.60	23.99	42.54	72.4%	20.7%	5.3%	1.7%
R4	11.42	26.19	48.62	68.4%	22.5%	6.4%	2.7%
Z3	12.80	29.48	54.96	68.2%	22.3%	6.8%	2.6%

**Legend:* A2: plant & soil science & technology; B1: biochemistry/biophysics/molecular biology; B2: cell biology; C1: analytical, inorganic & nuclear chemistry; C3: organic & medicinal chemistry; C6: materials science; E1: computer science/information technology; E2: electrical & electronic engineering; E3: energy & fuels; G1: astronomy & astrophysics; H1: applied mathematics; I1: cardiovascular & respiratory medicine; I5: immunology; M6: psychiatry & neurology; N1: neurosciences & psychopharmacology; P4: mathematical & theoretical physics; P6: physics of solids; R2: biomaterials & bioengineering; R4: pharmacology & toxicology; Z3: microbiology

As expected, b_k values range largely among the different disciplines. The lowest value has been found in the three subfields of engineering (E1, E2 and E3) and in A2 (plant & soil science & technology), while the highest one was observed in B2 (cell biology). By contrast, the citation classes defined by the characteristic scores are by and large insensitive to the underlying subject. The share of papers cited less frequently than the average (Class 1) amounts to roughly 70%, the share of those categorised to Class 2 to about 21% and in the highest two classes one finds 6%–7% and 2%–3% of all publications, respectively. The studies by Glänzel (2007; 2013a,b) give further empirical evidence that, in contrast to the b_k scores, this distribution over classes is strikingly stable with respect to the underlying subject field, the publication year as well as the citation window. This property makes the method useful for longitudinal and multi-disciplinary studies. Application of Characteristic Scores and Scales in comparative multidisciplinary studies

Application of Characteristic Scores and Scales in comparative multidisciplinary studies

A previous study (Glänzel, 2013b) has shown that *individual* scores can be derived for each paper if performance classes are to be built across science fields. One precondition for the application of CSS to broad science fields or to all fields combined is the unique assignment of publications to performance classes. The following example describes this problem. Assume, for instance, that a paper is assigned to two subjects, here denoted by S1 and S2. According to possibly different citation standards in the two subjects, the paper is then assigned, for instance, to Class 3 in subject S1 and to Class 4 in S2 because its citation rate does not exceed b_3 in S1 but it is greater than the corresponding threshold b_3 in S2. A direct combination can, therefore, not provide any acceptable solution. A proper subject-based fractionation must be applied such that each publication is gauged against only one individual threshold value. As argued in the study by Glänzel et al. (2009) one important consequence of multiple assignments is the necessity of fractionation by

subjects and thus of calculating proper weights for the corresponding individual subject-expected citation rates. Furthermore, it was stressed that the weighting of fractional data is correct only if the sum of the individual field expectations over all publications in the system equals the citation total of the database in the combination of these fields. This will result in an 'implicit' classification without calculating any *common* thresholds b_k. Again, the procedure is based on an iteration, where the first step is identical with the procedure of calculating subfield-expected citation rates. A first fractionation is applied when the citation means of subfields is determined. This is done on the basis of the respective number of subfields to which a publication is assigned. Both publications and citations are fractionated. The second one follows when individual expectations are calculated for each paper. This expectation is then the mean value of the fractionated subfield standards. In the following step of the iteration, all papers, that have received fewer citations than their individual expectation, are removed. The above procedure is repeated on the remaining set. This is done three times in total to obtain the *individual* characteristics scores b_k* (k = 1, 2, 3) for each publication. All papers can now uniquely be assigned to one of the four classes. It should be mentioned in passing that, if the underlying paper set comprises only publications from one single subfield and fractionation is not required, the results will be identical with those described in the previous subsection. It is straightforward that, in this case, the individual thresholds are identical with the common characteristic scores.

Table 2 shows the shares of papers in the four citation classes for the most active countries in 2007. We have used $N = 2{,}000$ as threshold. Among these countries, Belgium, Denmark, the Netherlands and Switzerland have the highest shares in the upper three CSS classes with more than 40% each. Norway, Sweden, UK and USA, with slightly lower values, have a similar profile. This, of course, corresponds to the lowest share of "poorly" cited papers (Class 1) since, by definition, the content of the four classes adds up to 100%. Besides, a similar share of Class 1 papers does not imply the same distribution over the upper classes. Even very similar shares of Class 2 papers might go with different distributions over the two other upper classes as the comparison of the country pairs Belgium-Sweden, Finland-USA and Brazil-China in all fields combined convincingly illustrates.

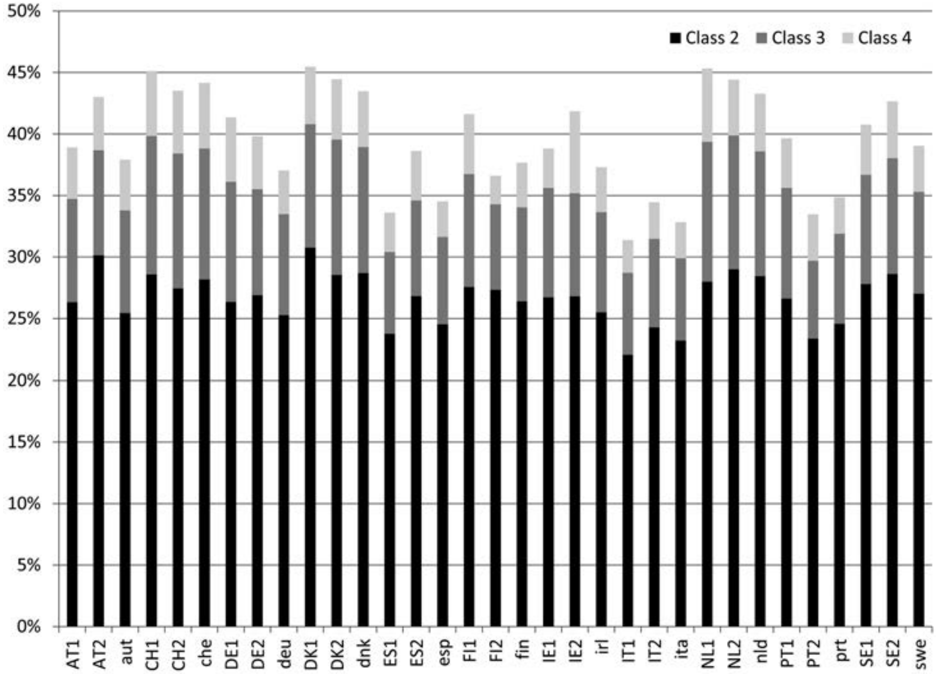
Table 2. National shares of publications in the reference CSS classes in 2007 for a 5-year citation window [Data sourced from Thomson Reuters Web of Knowledge]

Country	N	Class 1	Class 2	Class 3	Class 4	Country	N	Class 1	Class 2	Class 3	Class 4
USA	292546	60.59%	26.37%	8.96%	4.08%	AUT	9375	62.70%	25.21%	8.35%	3.74%
CHN	90041	73.05%	19.22%	5.67%	2.06%	FIN	8569	62.35%	26.40%	7.63%	3.62%
GBR	80641	62.04%	25.51%	8.56%	3.88%	IRN	7966	75.27%	18.10%	4.68%	1.95%
DEU	75924	62.98%	25.27%	8.21%	3.53%	MEX	7510	78.34%	16.19%	4.14%	1.33%
JPN	74761	73.94%	19.37%	4.99%	1.69%	NOR	7159	61.31%	26.85%	8.09%	3.76%
FRA	53927	63.64%	25.06%	7.97%	3.33%	CZE	6674	71.47%	21.19%	5.27%	2.07%
CAN	45820	62.93%	25.46%	8.06%	3.56%	SGP	6527	62.19%	26.51%	7.78%	3.52%
ITA	45069	67.18%	23.20%	6.70%	2.92%	PRT	6129	65.15%	24.57%	7.31%	2.97%
ESP	34516	65.49%	24.51%	7.11%	2.89%	ARG	5622	74.53%	19.23%	4.70%	1.55%
IND	30760	78.28%	16.58%	3.94%	1.21%	NZL	5545	64.53%	25.05%	7.50%	2.92%
AUS	29115	62.43%	26.06%	8.05%	3.46%	ZAF	5155	72.07%	19.75%	5.78%	2.41%
KOR	27338	74.02%	19.29%	5.00%	1.69%	HUN	4948	72.03%	20.35%	5.11%	2.51%
RUS	25637	87.17%	9.60%	2.29%	0.93%	IRL	4778	62.70%	25.49%	8.14%	3.66%
NLD	24547	56.74%	28.43%	10.16%	4.66%	UKR	3950	84.33%	12.51%	2.71%	0.46%
BRA	19885	75.75%	18.82%	3.95%	1.47%	THA	3566	70.70%	21.62%	6.03%	1.65%
TWN	18638	71.63%	21.44%	5.35%	1.58%	EGY	3433	79.35%	15.93%	3.70%	1.02%
CHE	18351	55.84%	28.17%	10.64%	5.35%	CHL	3229	71.63%	21.12%	5.33%	1.92%
SWE	17391	60.97%	27.02%	8.28%	3.73%	ROU	2936	78.58%	15.94%	4.36%	1.12%
TUR	16471	79.82%	15.26%	3.66%	1.26%	SVK	2309	77.48%	16.37%	4.68%	1.47%
POL	13742	78.62%	16.31%	3.51%	1.55%	SVN	2248	70.95%	21.89%	5.52%	1.65%
BEL	13656	59.37%	27.01%	9.56%	4.07%	PAK	2161	81.12%	12.82%	4.26%	1.80%
ISR	10731	67.60%	22.90%	6.76%	2.75%	MYS	2044	76.52%	17.27%	4.79%	1.42%
DNK	9485	56.54%	28.70%	10.24%	4.52%	BGR	2000	79.80%	15.10%	3.65%	1.45%
GRC	9406	70.35%	21.57%	6.04%	2.04%	World	985380	69.79%	21.50%	6.26%	2.44%

The profiles of Russia and Ukraine reflect the least favourable situation, and are even more skewed than the class distribution of Pakistan. Among the smaller countries, above all, Norway, Ireland and Singapore excel.

For the cross-institutional comparison of class profiles we have selected two universities each from eleven European countries (see Figure 1). Although the universities' profiles mostly mirror the national patterns, we find in several cases a distinctly more favourable situation than in the national standards. This is contrasted by a less favourable situation for the two South-European universities IT1 and PT2 as well as to a lesser extent for ES1, FI2 and the second Swiss university in the selection (CH2). The high standards of the selected Danish and Dutch universities are worth mentioning. Finally, 'DK1' and 'PT1' are technical universities while 'SE1' stands for a medical university. This again substantiates the subject-independence of the method (cf. Glänzel, 2013b).

Figure 1. Shares of publications of selected universities and countries in the upper three CSS classes in all fields combined in 2007 (5-year citation window)
[Data sourced from Thomson Reuters Web of Knowledge]



Discussion and conclusions

The analysis of the high end of scientific distributions is one of the most difficult and challenging issues in evaluative scientometrics. This is, of course, not merely a mathematical issue as it is always difficult to draw a sharp borderline between "very good" and "outstanding". Also the effect of outliers, i.e., of observations that might bias or even distort statistics, impressively shown by Waltman et al. (2012), is not typically a bibliometric issue. So-called censored data or data

distorting extreme values of a distribution are known in several fields, for instance, in insurance mathematics (e.g., Matthys et al., 2004). In the proposed CSS-based method, the effect of outliers is limited as the influence of individual observations on the total is marginal, and as the observations for the units under study are represented by classes instead of individual values.

Self-adjusting classes, such as those based on CSS, allow for the definition of proper performance classes without any pre-set thresholds. This is certainly one of the main advantages of the proposed method. Another one is the seamless integration of measures of outstanding performance into the assessment tools of standard performance. The method of "implicit" subject fractionation can also be used in the context of other publication and citation indicators, whenever the issue of multiple subject assignment needs to be resolved.

The studies have shown that a publication output at the meso level suffices to provide a solid basis of interpretation and further statistical analysis. A further important property has become apparent, namely the method's independence of the unit's research profile. In the small sample we have found two technical universities with more favourable citation profiles than that of medical universities or than their corresponding national reference standards.

References

Glänzel, W. & Schubert, A. (1988), Characteristic Scores and Scales in assessing citation impact. *Journal of Information Science*, 14(2), 123–127.

Glänzel, W. (2007), Characteristic scores and scales. A bibliometric analysis of subject characteristics based on long-term citation observation. *Journal of Informetrics*, 1(1), 92–102.

Glänzel, W., Schubert, A., Thijs, B. & Debackere, K., Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance. *Scientometrics*, 78(1), 2009, 165–188.

Glänzel, W. (2013a), High-end performance or outlier? Evaluating the tail of scientometric distribution. *Scientometrics*, to be published.

Glänzel, W. (2013b), *The Application of Citation-Based Performance Classes in Disciplinary and Multidisciplinary Assessment*. Unpublished manuscript.

Leydesdorff, L., Bornmann, L., Mutz, R. & Opthof, T. (2011), Turning the tables on citation analysis one more time: principles for comparing sets of documents. *JASIST*, 62(7), 1370–1381.

Matthys, G., Delafosse, E., Guillou, A. & Beirlant, J. (2004), Estimating catastrophic quantile levels for heavy-tailed distributions. *Insurance Mathematics & Economics* 34(3), 517–537.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E.C.M., Tijssen, R.J.W., van Eck, N.J., van Leeuwen, Th.N., van Raan, A.F.J., Visser, M.S. & Wouters, P., (2012), *The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation*. In: Eric Archambault, Yves Gingras, & Vincent Larivière (Eds.), Proceedings of STI 2012 Montreal (17^{th} International Conference on Science and Technology Indicators), Volume II, 791–802.

Measuring the diversity of research using paper similarities[1]

Jochen Gläser Jochen.Glaser@ztg.tu-berlin.de / Center for Technology and Society, TU Berlin, Hardenbergstr. 16–18, Berlin, 10623 (Germany)

Frank Havemann Frank.Havemann@ibi.hu-berlin.de / Berlin School of Library and Information Science, Humboldt-University of Berlin, Dorotheenstraße 26, 10099 Berlin (Germany)

Michael Heinz Michael.Heinz@ibi.hu-berlin.de / Berlin School of Library and Information Science, Humboldt-University of Berlin, Dorotheenstraße 26, 10099 Berlin (Germany)

Alexander Struck Alexander.Struck@staff.hu-berlin.de / Berlin School of Library and Information Science, Humboldt-University of Berlin, Dorotheenstraße 26, 10099 Berlin (Germany)

Abstract

The epistemic diversity of research has become a matter of concern for science policy. However, the topic is still ill-understood in science studies. Theoretical reasoning implicitly or explicitly draws analogies to biodiversity. Similarly, the calculation of diversity uses indicators that have their origin in biodiversity research. The aim of our paper is to demonstrate that research fields are quite unlike biotopes, and that the differences have consequences for the ways in which diversity can be measured. Our core argument is that the empirical objects of science studies – research fields, research processes, papers and so on – differ from the empirical objects of biodiversity research in ways that makes it difficult to apply some of the most popular diversity measures. We therefore suggest a diversity measure that operates on the paper level, i.e. does not assign papers to topics but is based on the similarity of pairs of papers, which is defined on the basis of the 'resistance distance' in a network of connected papers. We present our measure and first attempts to validate it. As a demonstration, we compare the diversity of information science papers in five countries.

Introduction

The epistemic diversity of research – the diversity of empirical objects, methods, problems, or approaches to solving them – has become a matter of concern for science policy. Attempts by science policy to increase the selectivity of research funding and the growth in strength and homogeneity of incentives for universities have led to concerns about an undue reduction of the diversity of research. Several specific warnings refer to the UK's research assessment exercise (Gläser et al. 2002, Molas-Gallart and Salter 2002, Rafols et al. 2012). A similar concern has been raised in Germany, where similar profile-building activities at all universities may make the small subjects disappear (HRK 2007).

1 This work was supported by the German Ministry for Education and Research.

All discussions about dangers to the epistemic diversity of research have in common that they cannot rely on sound theoretical backing or empirical evidence. Epistemic diversity is an ill-understood topic in science studies. It is rarely clear what the concept is intended to refer to, how epistemic diversity might affect research, and how it can be operationalized. Theoretical reasoning implicitly or explicitly draws analogies to biodiversity. Similarly, the calculation of diversity (e.g. by Rafols et al. 2012) uses indicators that have their origin in biodiversity research. This borrowing assumes functional and structural similarities between biotopes (a physical environment (habitat) and its distinctive assemblage of conspicuous species, Olenin and Ducrotoy 2006: 22) and research fields that are rarely, if ever, tested.

The aim of our paper is to demonstrate that research fields are quite unlike biotopes, and that the differences have consequences for the ways in which diversity can be measured. Our core argument is that the empirical objects of science studies – research fields, research processes, papers and so on – differ from the empirical objects of biodiversity research in ways that makes it difficult to apply some of the most popular diversity measures. We therefore suggest a diversity measure that operates on the paper level, i.e. dos not assign papers to topics but is based on the similarity of pairs of papers, which is measured as the 'resistance distance' in a network of connected papers. We present our measure and first attempts to validate it. As a demonstration, we compare the diversity of information science papers in five countries.

From biodiversity to research diversity?

Biodiversity and its measurement

Biodiversity has been an important topic of biological and environmental research for some time, with the impact of diversity on the stability and development of biotopes and species being a major concern. Two approaches to the measurement of biodiversity can be distinguished:

(a) The diversity of biotopes composed of several species is measured with a three-level hierarchical approach. Biotopes are considered as consisting of species, which in turn consist of individuals. Three factors contribute to the diversity of such a system, namely

   - variety (the number of species in the biotope),
   - disparity (the extent to which the species differ from each other), and
   - evenness (the distribution of individuals across the different species).

   Depending on the research question, these factors can be assessed separately (e.g. if only the number of species is measured) or be combined in synthetic measures such as the Shannon-Index (combining variety and evenness) or the Rao-Index (combining all three measures). This approach to diversity is applied in fields outside the biosciences as well (see Rafols et al. 2012, Stirling 2007). It requires that

- the system whose diversity is to be measured can be analytically decomposed in three levels (system, categories, and elements),
- the contribution of differences between individuals of the same species to the biotope's diversity can be neglected,
- the categories can be constructed as disjunct by assigning each element to exactly one category or by fractional assignments of elements to categories, and that
- all categories share a property that can be used to calculate disparity.

(b) The diversity of species composed of individuals is measured on the basis of a two-level hierarchical approach. In this approach, variety and evenness become meaningless because there is no intermediate level of categories to which elements can belong. The only remaining basis for measuring the diversity of the system is the disparity of individuals. While this approach is used less frequently, it can be considered to be more fundamental because it conceptualizes diversity as the degree to which the elements of a system (here: a species) differ from each other. This approach is applicable as long as a system can be delineated and elements share a property that can be used to calculate disparity.

The challenges of measuring the epistemic diversity of research

The epistemic diversity of research can be understood as the diversity of problems, empirical objects, approaches and methods. These elements of research are fully or partially shared by groups of researchers and referred to in sets of publications, where they constitutes topics common to some publications. While the diversity of elements of research processes cannot be bibliometrically measured, its representation in the diversity of topics in papers should be bibliometrically accessible. In the following, we therefore refer to epistemic diversity as the diversity of elements of research processes as it is reflected in the diversity of topics addressed in papers.

We define a topic as a focus on theoretical, methodological or empirical knowledge that is shared by a number of researchers and thereby provides these researchers with a joint frame of reference for the formulation of problems, the selection of methods or objects, the organization of empirical data, or the interpretation of data (on the social ordering of research by knowledge see Gläser 2006). This definition resonates with Whitley's (1974) description of research areas but assumes a lower degree of organization. The only demand the definition makes is that some scientific knowledge is perceived similarly by researchers and influences their decisions.

Since topics are constituted by coinciding perspectives of researchers, they have four properties that create the problems for bibliometrics:

(1) Knowledge has a fractal structure (e.g. van Raan 2000), and topics can have any size between the smallest (emerging topics that in the beginning may concern just two or three researchers) and very large thematic structures (fields or themes cutting across several fields). The 'size' of a topic can be defined in various ways – as scope (range of phenomena covered), level of

abstraction (which is again linked to the range of phenomena covered), or number of research processes or researchers influenced by it. For all these properties there is a continuum from very small to very large topics.

(2) Given the multiple objects of knowledge that can serve as common reference for researchers, it is inevitable that topics overlap. Two kinds of overlaps can be distinguished. First, any knowledge can be part of several topics at once. For example the formulae used in bibliometrics belong to mathematics insofar they are formulae but also may express bibliometric relationships. Second, any research is likely to address several topics at once, e.g. by including theories about an object, methodologies for investigating it, and empirical information about an object.

(3) All topics are emergent in the sense that they are products of autonomous interpretation and use of knowledge by researchers. Topics emerge from the interaction of these research processes, which is mediated through publications. While individual researchers may launch topics and advocate them, their content and persistence depends on the ways in which these topics are used by their colleagues.

(4) From this follows that topics are local in the sense that they are primarily topics to the researchers whose decisions are influenced and who contribute to them, and only secondarily topics to those colleagues who are outside observers. This 'locality' holds regardless of the size of a topic because it is a distinction between the common perspective of participants (however many there are) and observers. If we accept that topics are created by researchers through research (and not just 'discovered'), then we must accept that they represent coinciding perspectives on knowledge, and are local in the sense that they are created by a finite set of contributors.[i]

These properties of topics are responsible for the difficulties one encounters when attempting to delineate topics by bibliometric methods. None of the properties of a paper that can be used by bibliometrics can be assumed to be thematically homogeneous in the sense of representing only one topic. Since research processes are influenced by and address more than one topic, topics overlap in research processes, publications (and thus references), terms, journals, and authors. Furthermore, the relationships between topics are diffuse and keep changing. Any finite sub-set of papers is unlikely to completely contain a topic, which means that any hierarchy of topics is also only partially covered by the paper set. At the same time, the properties of topics described above are responsible for the severe difficulties that occur when we attempt to measure the diversity of research.

- The 'distinctness' of topics (the extent to which a common point of reference for researchers is considered as a topic) varies.
- Topics overlap one another, and may contain other topics partly or fully.
- Topics may also contain subsets that are not topics.
- Topics may simultaneously belong to higher-order topics at different hierarchical levels.

Thus, topics can be assumed to form an inconsistent poly-hierarchy that spans several levels, not all of which are populated by all trees of the hierarchy. Furthermore, topics and their relationships constantly change with newly produced knowledge. This shifting, fluid and inconsistent polyhierarchy does not lend itself easily to the measurement of research diversity. Applying the diversity measures for three-level hierarchies requires the construction of a consistent three-level monohierarchy. While this can be easily achieved – e.g. by using Web of Science categories as topics – much of the information about topics and their relationships is lost in the process. This raises questions about the validity of the diversity measures applied to such constructs.

While we would not exclude the possibility that valid three-level measures for diversity can be constructed, no easy solution suggests itself. Therefore, we use the two-level approach, which can be applied to any set of papers sharing a property that enables the calculation of pairwise dissimilarities (i.e. disparities) of papers.

Data and Methods

The two-level approach that considers diversity as a measure of the differences between elements requires only a common property of all elements that can be used to calculate the elements' similarities. If we use cited sources or shared terms as a basis for similarity, we cannot calculate the thematic similarity of two papers which do not share sources or terms. This obstacle can be overcome by including indirect connections in networks of papers and their cited sources. In such a network, similarities of all papers that belong to the same component of the network can be calculated.

We apply a parameter-free method that takes all possible paths between two nodes in a network into account. Following Klein & Randic (1993) and Tetaly (1991) we calculate 'resistance distances' between nodes. The 'resistance distance' measure models a network as a set of nodes linked by resistors and calculates the similarity of two nodes as the current resulting from all possible conduits in the network.

For the purpose of our calculation of diversity we assume that a suitable network of papers and cited sources has been identified. We conduct our experiments with the main component of the nearly bipartite citation network of papers published in the 2008 volumes of six information-science journals together with the about 14,000 sources cited by the 492 papers in the main component. In this network, the degree of a paper node equals the number of its references and the degree of a source equals the number of citations from papers to this source. The network is not strictly bipartite because there are 29 direct citation links between papers in the set. We neglect that the citation network is directed.

In the electric model we assume that each link has a conductance equal to its weight. We calculate the effective resistance between two nodes if they would operate as poles of an electric power source. The varying lengths of reference lists and distorting influences of highly cited papers

require a normalization of link weights, for which we use the inverse geometric mean of the degrees of the two nodes (Havemann et al. 2012b).

Only similarities of the 492 papers need to be calculated. Since approximate values of similarities are sufficient, we can apply a fast approximate iteration method introduced by Wu and Huberman (2004). For an extensive description of the method see Havemann et al. (2013).

In a previous experiment, we identified three informetric topics – bibliometrics (224 papers), Hirsch-index (42 papers), and webometrics (24 papers) – in this set by analyzing keywords, titles, and abstracts (Havemann et al. 2012a). These topics are used for validating the 'resistance distance' similarity measure in two ways. We first assess how well the predefined topics are reconstructed by a cluster algorithm (Ward) that uses resistance as distance measure. We then rank papers according to their resistance distance to papers in a topic.

Finally, we perform a first comparative study of diversity by comparing country sets of information science papers.

Results

The 492 papers form 120,786 pairs. The mean resistance distance R of the network is 3.711034, the geometric mean is 3.222144, and the median is 2.952775.

An important step in the introduction of a new indicator is its validation, i.e. confirming that it indeed measures what is claimed to be measured. We can currently offer three confirmations of the validity of our similarity measure.

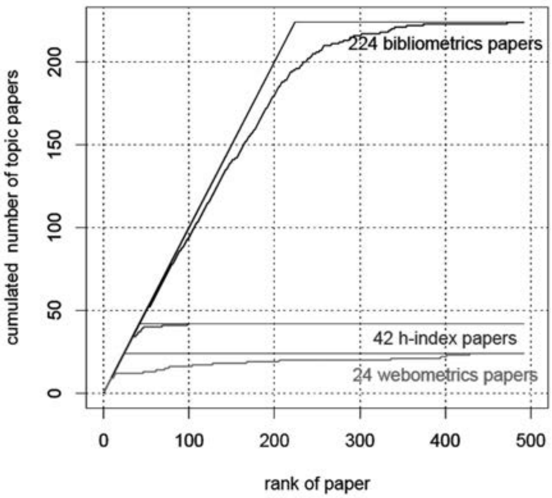
(1) If the similarity measure is valid, clustering the papers on its basis should reproduce the topics we identified by inspecting titles and abstracts of papers. This is indeed the case. We used the resistance distance to cluster the 492 papers (Ward clustering). Table 1 provides the recapture of the three topics and compares them to the performance of three approaches we developed for the identification of topics in networks of papers. It turns out that the new measure performs approximately as well as the others, which suggests that it does indeed measure the similarity of papers.

(2) If the similarity measure is valid, its application to the similarity between papers and topics should return maximal similarities between papers belonging to a topic, and lower similarities to other papers. This application requires a correction for the centrality of a paper in the whole network, which we achieve by dividing the median distance of a paper to all topic papers by its median distance to all papers (Havemann et al. 2013). We tested this measure by ranking all papers according to their corrected median distances to the papers belonging to the three topics (Figure 1). We expect the papers classified by us as belonging to a topic at top ranks.

Table 1. Recapture of the h-index, bibliometrics and webometrics topics by Ward clustering using resistance distances and three other methods (Havemann et al. 2012b). The first line for each topic gives Salton's cosine of the two sets.

	Ward clustering with resistance distances	Merging of natural communities	Hierarchical link clustering	Fuzzification of hard clusters
h_index	0.88	0.71	0.93	0.59
precision	0.86	0.56	0.91	0.35
recall	0.90	0.89	0.95	1.00
bibliometrics	0.87	0.79	0.82	0.83
precision	0.84	0.72	0.83	0.87
recall	0.91	0.86	0.81	0.80
webometrics	0.53	0.58	0.60	0.46
precision	0.61	0.53	0.85	0.45
recall	0.46	0.64	0.43	0.47

Figure 1. Cumulated number of topic papers obtained from ranking all 492 papers according to their normalized median distance to papers of the three topics. The black lines represent the ideal cases, where all papers of a topic rank above other papers.



(3) The ranking of papers according to 2) also led to a third validation of our similarity measure, a validation we did not anticipate. The ranking of papers according to their modified resistance distances to the paper by Glänzel (2008), which we assigned to the topic "h-index", resulted in a high rank of a paper by Leydesdorff (2008). When inspecting titles and abstracts

of papers, we did not assign the Leydesdorff paper to this topic because its title and abstract do not mention the h-index. However, closer inspection of the paper reveals that it cites many sources that are also cited by the papers belonging to the h-index, and that Leydesdorff does indeed discuss the h-index (ibid: 279).

We construct a measure of the diversity of the set of papers and apply it in a comparison of national subsets of papers. Since this comparison again concerns sub-sets of the whole network, a similar modification as described in 2) is applied. We correct for the number of connecting references of each paper by determining the Spearman rank correlation between pairs of resistance vectors of papers. Thus, we use Spearman's ρ as a measure of paper similarity or equivalently $1-(\rho+1)/2$ as a measure of paper distance.

We selected Belgium, France, Germany, the Netherlands, the UK and the US for our comparison. Figure 2 provides the median distances. A validation of this comparison, which consists of further statistical tests and an assessment of the similarities of papers from these countries, will be presented at the STI conference.

Figure 2. Distributions of paper distances for six countries

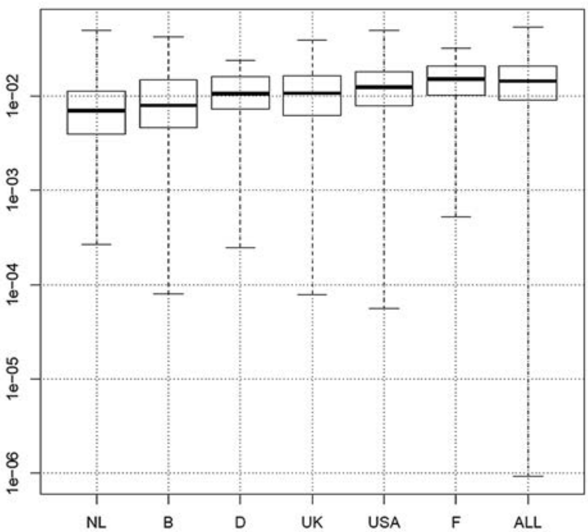
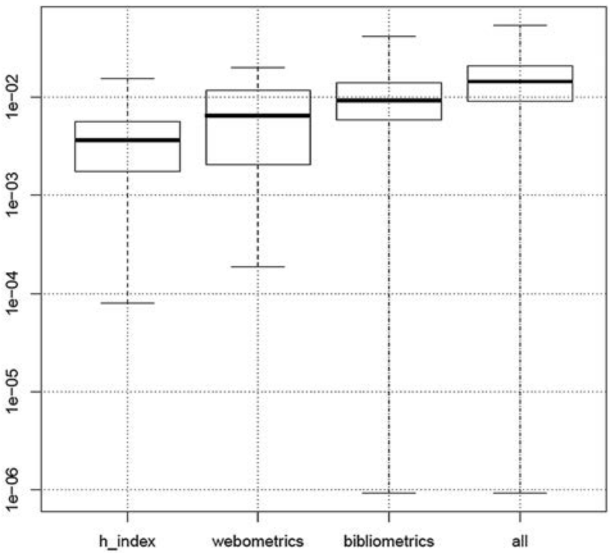


Figure 3 shows the boxplots of paper distance distributions of the three topics introduced above. Again, we do not use resistance distances directly but transform it into the distance based on Spearman's rank correlation coefficient.

Figure 3. Distributions of paper distances for three topics



Discussion

Although further validation studies of the proposed diversity measure are necessary, first results indicate that it is indeed a valid measure of similarity, and that the median resistance distance in a network of papers can serve as a measure of its diversity. Such a measure might be preferable whenever the diversity of smaller sets of papers needs to be assessed and the coarse construction of topics (e.g. on the basis of Web of Science categories, Rafols et al.2012) would distort the measurement. While the three-level approach to diversity – assigning papers to topics and measuring variety, disparity of topics, and evenness – might be preferable in many cases because it produces additional information, current methods of constructing topics are forced to ignore some major characteristics of these topics. This is why we think that using the two-level approach is an alternative in many cases.

Conclusions

Future work, which will be presented at the STI conference, includes the validation of differences in country diversity by inspecting country papers and assigning them to topics, statistical tests of country differences in diversity, and the application of the measure to a field from the sciences.

References

Glänzel, W. (2008). On some new bibliometric applications of statistics related to the h-index. *Scientometrics* 77, 187–196.

Gläser, J. (2006). *Wissenschaftliche Produktionsgemeinschaften: Die Soziale Ordnung der Forschung*. Frankfurt am Main: Campus.

Gläser, J. G. Laudel, S. Hinze & L. Butler (2002). Impact of Evaluation-based Funding on the Production of Scientific Knowledge: What to Worry About, and how to Find Out. Report to the BMBF. Retrieved February 27, 2013, from http://www.sciencepolicystudies.de/dok/expertise-glae-lau-hin-but.pdf.

Havemann, F., J. Gläser, M. Heinz & A. Struck (2012a, June). Evaluating overlapping communities with the conductance of their boundary nodes. *arXiv*:1206.3992.

Havemann, F., J. Gläser, M. Heinz & A. Struck (2012b, March). Identifying overlapping and hierarchical thematic structures in networks of scholarly papers: A comparison of three approaches. *PLoS ONE* 7 (3), e33255.

Havemann, F., J. Gläser, M. Heinz & A. Struck (2013, March). Estimating Thematic Similarity of Scholarly Papers with Their Resistance Distance in an Electric Network Model. http://141.20.126.227/~qm/Havemann2013resistance.pdf

Hochschulrektorenkonferenz (HRK) (2007). Die Zukunft der kleinen Fächer: Potenziale – Herausforderungen – Perspektiven, Bonn: Hochschulrektorenkonferenz.

Klein, D. J. & M. Randić (1993, December). Resistance distance. *Journal of Mathematical Chemistry* 12 (1), 81–95.

Leicht, E. A., P. Holme & M. E. J. Newman (2006, February). Vertex similarity in networks. *Physical Review* E 73 (2), 026120.

Leydesdorff, L. (2008). Caveats for the use of citation indicators in research and journal evaluations. *JASIST* 59, 278–287.

Molas-Gallart, J. & A. Salter (2002). Diversity and Excellence: Considerations on Research Policy. IPTS Report. Retrieved September 14, 2008 from http://ipts.jrc.ec.europa.eu/home/report/english/articles/vol66/ITP1E666.html.

Olenin, S. & J.-P. Ducrotoy (2006). The concept of biotope in marine ecology and coastal management. *Marine Pollution Bulletin* 53, 20–29.

Rafols, I., L. Leydesdorff, A. O'Hare, P. Nightingale & A. Stirling (2012). How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management. *Research Policy* 41, 1262–1282.

Stirling, A. (2007). A general framework for analyzing diversity in science, technology and society. *Journal of The Royal Society Interface* 4, 707–719.

Tetali, P. (1991, January). Random walks and the effective resistance of networks. *Journal of Theoretical Probability* 4 (1), 101–109.

Van Raan, A. F. J. (2000). On growth, ageing, and fractal differentiation of science, *Scientometrics* 47, 347–362.

Whitley, R. (1974), Cognitive and social institutionalization of scientific specialties and research areas. In: R. Whitley (ed), *Social Processes of Scientific Development* (69–95). London.

Wu, F. & B. A. Huberman (2004, March). Finding communities in linear time: a physics approach. *The European Physical Journal B – Condensed Matter* 38 (2), 331–338.

---

i This does not mean that constructing topics by global approaches is meaningless. Global approaches create distinct representations by simultaneously including insider and outsider perspectives. The selection of local or global approaches depends on the problem that is to be solved.

What do citations, downloads and readership data of an information systems journal have in common and where do they differ?[1]

Juan Gorraiz — juan.gorraiz@univie.ac.at / University of Vienna, Vienna University Library, Dept of Bibliometrics, Boltzmanngasse 5, A-1090 Vienna (Austria)

Christian Gumpenberger — christian.gumpenberger@univie.ac.at / University of Vienna, Vienna University Library, Dept of Bibliometrics, Boltzmanngasse 5, A-1090 Vienna (Austria)

Kris Jack — kris.jack@mendeley.com / Mendeley, 144a Clerkenwell Road, London, EC1R5DF (UK)

Peter Kraker — pkraker@know-center.at / Know-Center, Inffeldgasse 13, A-8010 Graz (Austria)

Christian Schlögl — christian.schloegl@uni-graz.at / University of Graz, Institute of Information Science and Information Systems, Universitätsstr. 15/F3, A-8010 Graz (Austria)

Abstract

In our article we investigate the commonalities and differences across citations, downloads and Mendeley readership data for The Journal of Strategic Information Systems. Our analysis shows a medium to high correlation between downloads and citations (Spearman $r = 0.77$) and downloads and readership data (Spearman $r = 0.73$) which is lower between citations and readership data (Spearman $r = 0.51$). The difference among the three data sets is stronger with regards to obsolescence characteristics.

Introduction

There exist already several studies which have compared download and citation data. With the advent of the social web and its growing acceptance in academia, alternative metrics seem to be a further source for the measurement of science (Bar-Ilan et al, 2012). In particular, the social reference management system Mendeley seems to be an interesting candidate.

It is the goal of this contribution to investigate the commonalities and differences across citations, downloads and readership data for an information systems journal (The Journal of Strategic Information Systems). In particular, the following issues will be addressed:

1 This report is based in part on analysis of anonymous ScienceDirect usage data and/or Scopus citation data provided by Elsevier within the framework of the Elsevier Bibliometric Research Program (EBRP). Readership data were provided by Mendeley. The authors would like to thank both Elsevier and Mendeley for their great support.

– Are most cited articles the most downloaded ones and those which can be found most frequently in user libraries of the collaborative reference management system Mendeley?
– Do citations, downloads and readership data have different obsolescence characteristics?

Both citations and downloads were provided by Elsevier in the framework of the Elsevier Bibliometric Research Program (EBRP). For all documents published between 2002 and 2011 all monthly downloads were made available from ScienceDirect (SD) and all monthly citations from Scopus until mid of 2012. Furthermore, we received the number of occurrences of full length articles in user libraries together with the date of their particular data entry in Mendeley from 2002 to 2011.

Results

Download data

Though we received the download data for all SD documents, we concentrate on full length articles in the following. As Table 1 shows (since these are critical data for ScienceDirect, only reference numbers are displayed), there was a strong increase in the number of downloads between 2002 and 2011. An analysis of the obsolescence characteristics reveals that from the downloads of a certain year, most of them allot to articles either published in the download year or one year earlier (formatted in bold). However, also older articles are downloaded relatively often.

Table 1. Yearwise relation of downloads per print publication year (2002–2011), (doc type: full length article, download year: <=2011) (n=181)*

Pub year	**n**	**Download year**											**downloads per doc – *relations*[1]**
		2002	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**All**	
2002	13	1.0	**2.3**	1.7	1.3	1.2	1.4	2.4	2.8	2.8	2.7	19.6	7.4*x
2003	21	0.0	1.3	**2.2**	1.0	1.0	0.9	1.5	1.3	1.5	1.1	11.9	2.8*x
2004	17			1.7	**2.6**	2.1	2.2	2.4	2.7	2.9	2.3	18.9	5.5*x
2005	18				1.7	2.3	1.8	2.0	2.4	2.6	2.2	15.0	4.1*x
2006	14				0.2	**2.4**	2.1	1.8	2.1	2.0	2.0	12.5	4.4*x
2007	18					0.0	**2.7**	**3.6**	3.4	3.5	2.9	16.1	4.4*x
2008	16						0.0	2.9	**3.5**	3.0	2.4	11.8	3.6*x
2009	14								3.1	**4.0**	3.1	10.2	3.6*x
2010	21									3.9	4.4	8.3	2.0*x
2011	29									0.3	**5.6**	5.9	1.0*x
all	181	1.0	3.7	5.6	6.8	8.9	11.1	16.6	21.4	26.4	29.0	130.4	

*Since the download numbers are very sensitive, we did not provide the absolute figures but only the ratios among them.

Citation data

Table 2. Year-wise citations (2002–2011) per publication year (document types: article, review, conference paper), only cited documents (n = 150).

Pub year	n	Citation year											cites per doc
		2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	all	
2002	13	2	19	**38**	**69**	**88**	**105**	**158**	**165**	**194**	**199**	1037	**79.8**
2003	14		1	6	21	27	39	35	41	40	39	249	17.8
2004	17				15	40	56	74	78	88	107	458	26.9
2005	19					16	46	78	76	93	99	408	21.5
2006	14				1	2	14	31	31	53	49	181	12.9
2007	18						1	31	74	92	85	283	15.7
2008	15							3	30	69	83	185	12.3
2009	14								3	34	57	94	6.7
2010	18									5	40	45	2.5
2011	8										14	14	1.8
all	150	2	20	44	106	173	261	410	498	668	772	2954	

Since ScienceDirect and Scopus use different document types which are not compatible to each other, the Scopus analyses were performed for the three Scopus document types “article”, “conference paper” and “review” which mainly correspond to “full length articles” in SD. Table 2 shows the year-wise citation distribution of articles, reviews and conference papers between 2002 and 2011. As can be seen, in all citation years most citations (formatted in bold) accrue to articles from the publication year 2002. In contrast, only a few documents were cited in the year of publication. This shows a clear difference to downloads which have their maximum either in the year of publication or one year later.

Readership data

Since Mendeley started in 2008, there are only very few readership data for this year. The strong increase of the users – the end of 2012 Mendeley reached two million users – makes an obsolescence analysis difficult. As can be seen in Table 3, the maximum occurrences accrued in the last one and a half years. However, also older articles, for instance published in 2002, were added to user libraries in more recent years though, due to a much smaller time window, their total is not as high compared to the download and citation data.

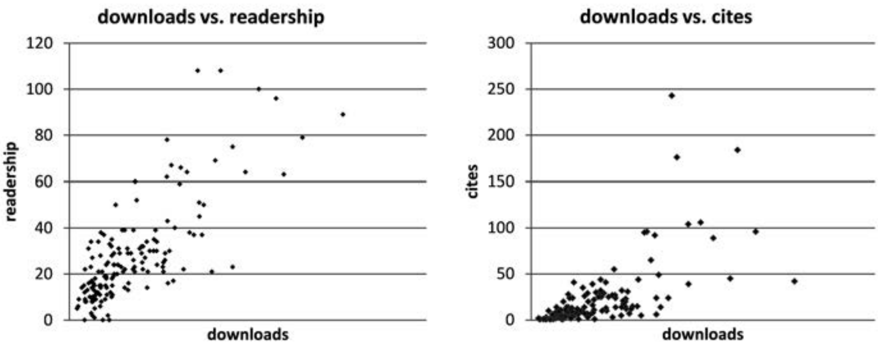
Table 3. Readership data per print publication year (2002–2011), (doc type: full length article, data extracted from Mendeley: October 2012) (n = 181)

Pub year	n	Download year						down-loads per doc
		2008	2009	2010	2011	- July 2012	all	
2002	13	7	30	126	**245**	183	**591**	**45.5**
2003	21	1	29	58	108	**145**	341	17.1
2004	17	11	36	107	158	**165**	477	28.1
2005	18	2	31	79	141	**151**	404	23.8
2006	14	6	39	88	128	**148**	409	29.2
2007	18	4	45	129	222	209	**609**	35.8
2008	16	7	36	99	182	164	488	32.5
2009	14	0	27	111	127	**150**	415	29.6
2010	21	0	0	84	**238**	191	513	24.4
2011	29	0	0	4	208	**282**	494	17.6
all	181	38	273	885	1757	1852	4741	

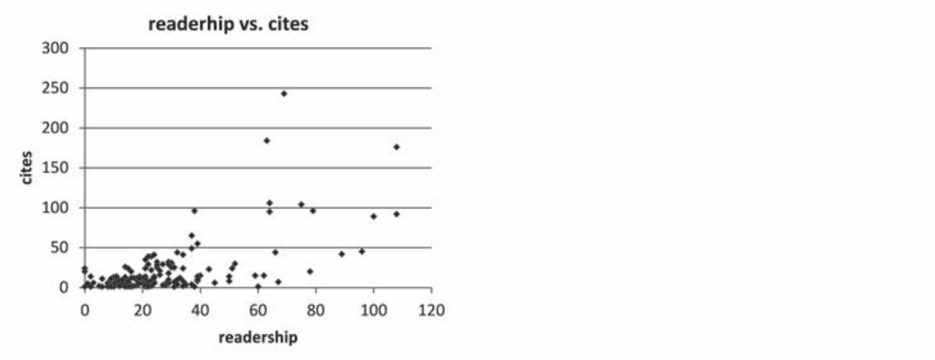
Comparison among downloads, citations and readership data

Figure 1 shows a medium to high relation among downloads, citations and readership data which is higher for downloads and citations (Spearman r = 0.77) and for downloads and readership data (Spearman r = 0.73). Among the ten most downloaded articles, seven (not the same) are in the top-10 readership and citation rankings. The correlation was lower between readership data and citations (Spearman r = 0.51) but in line with previous studies (for instance, Bar-Ilan 2012; Li, Thelwall and Giustini 2012).

Figure 1. Downloads vs. readers vs. cites, scattergram (publication year: 2002–2011, doc type: full length article, only articles cited at least once) (n = 151)

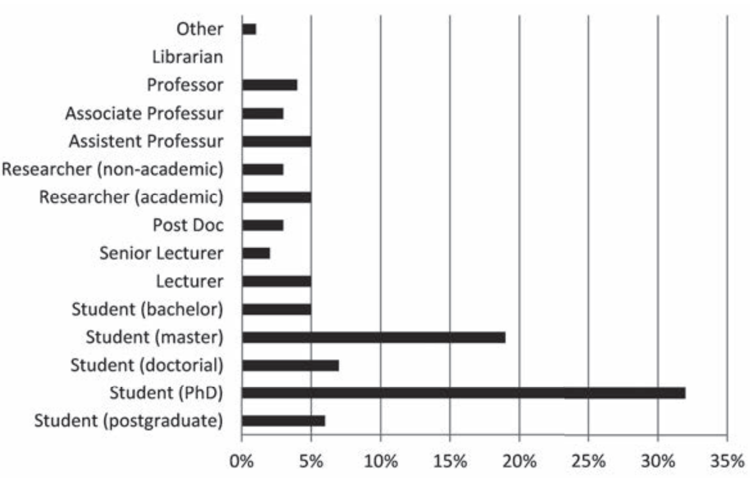


... *Continuation Figure 1*



One reason for the lower correlation between Mendeley readership and citation data could be that Mendeley users are younger. As Figure 2 exhibits, approximately two thirds of the readers of the analyzed information systems journal were students (most of them PhD and master students). This population clearly differs from that of the citing articles from the same journal.

Figure 2. Readership structure of the articles (2002–2011)



Conclusions and future research

Our study showed a medium to high correlation between downloads and citations and downloads and readership data which is lower between citations and readership data. In most cases, the most cited articles are also among the most downloaded and most read ones. An analysis of the readership data revealed that a perfect relation, at least between citations and Mendeley read-

ership data, cannot be expected because of the different populations involved. There are clear differences between citations and downloads with regards to obsolescence characteristics. The obsolescence of readership data might be more similar to those of the downloads. However, due to the dynamic development of Mendeley, a definite prediction is not yet possible.

In the near future, we plan investigations with more journals also from other disciplines (e.g. economics, oncology, linguistics, and history).

References

Bar-Ilan, J. (2012). JASIST@mendeley. *ACM Web Science Conference 2012 Workshop*. Retrieved January 23 2013 from: http://altmetrics.org/altmetrics12/bar-ilan/

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H. & Terliesner, J. (2012). Beyond citations: scholars' visibility on the social web. In *Proceedings of the 17th International Conference on Science and Technology Indicators* (98–109). Montréal: Science-Metrix and OST.

Li, X., Thelwall, M. & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement. *Scientometrics*, 91(2), 461–471.

Access and utilisation of social capital in knowledge transfer

Thomas Gurney t.gurney@rathenau.nl / Science System Assessment, Rathenau Instituut, Anna van Saksenlaan 51, The Hague, 2593 HW (The Netherlands)

Edwin Horlings e.horlings@rathenau.nl / Science System Assessment, Rathenau Instituut, Anna van Saksenlaan 51, The Hague, 2593 HW (The Netherlands)

Peter van den Besselaar p.a.a.vanden.besselaar@vu.nl / Department of Organisation Science & Network Institute, VU University Amsterdam, De Boelelaan 1105, 1081HV, Amsterdam (The Netherlands)

For start-ups choosing to locate in Science Parks, the physical and social environments are sources of social capital for firm founders. In this study we examine the social and physical proximity, contextual origins and application spheres of various sources of social capital available to start-ups prior to, and following, their choice to locate within a Science Park. Interviews are conducted with biotechnology-oriented start-up founders, specifically investigating their sources of social capital in three key phases of the firm development. We find that start-ups choose to locate in the Science Park based primarily on neo-classical location theory and social capital is primarily externally sourced with minimum involvement of the Science Park itself or with other firms located within the Science Park. The local HEI is also cited as being instrumental in sourcing and mobilising social capital.

Introduction

Science Parks have entered the literature in waves with each crest bringing new ideas and theories as to their utility to science, innovation, and society. Studies typically evaluate the utility of a Science Park (Dettwiler, et al., 2006), compare Science Parks (Fukugawa, 2006), or to compare firms on and off Science Parks (Squicciarini, 2008). The aim of this study is to highlight the social developments experienced by firm founders in the particular environment of a Science Park.

Conceptual framework

Science Parks have been wielded as a policy tool for many years, and numerous policy initiatives such as the EU Framework Programmes and the Bayh-Dole Act (which signalled a change in the intellectual property regime in favour of universities) have incentivised the formation of Science Parks across the globe (Siegel 2003).

Common threads link previous studies on science parks. These include:

(1) No common definition: Science Parks, and the utility of Science Parks, have been extensively studied, yet common definitions are hard to come by. General descriptions of a Science Park sum to a property-based, technology-oriented agglomeration of firms of varying specialisation and size, with close links and opportunities – either cognitively, geographically, structurally or commercially – between firms and to a higher education or research institution (T. K. Das & Teng, 1997; Löfsten & Lindelöf, 2005; Quintas, et al., 1992; Siegel, et al., 2003).

(2) Unique origins: Each Science Park has unique origins and context. They can be a rejuvenation response to natural disaster (Kobe Science Park), the result of agglomeration (Silicon Valley), or the result of competition from other Science Parks (Hsinchu was a response to Silicon Valley) (Koh, et al., 2005).

(3) Host of motivations for Science Park formation: They provide an environment for large firms to develop relationships with small firms; promote formal and informal links between firms, universities and other small labs (Löfsten & Lindelöf, 2005); to provide a contact space between "fast applied science" and "slow basic science" (Quintas, et al., 1992); to promote foreign investment (Koh, et al., 2005); or development on a regional or national basis (Phillimore, 1999).

(4) A Science Park must seek tenants, regardless of the ulterior motives of the firm founder. There is an expectation of economic and social return to investments. Science Parks compete with each other to attract new firms to their location (Phan, et al., 2005). Firms choosing to locate on a Science Park can be a HEI spin-off/start-up or as the subsidiary of an outside firm (without links to the HEI).

(5) Different tenants seek different benefits. There are various location theories including Neo-classical (transport, labour costs, distances and agglomeration economies), Behavioural (mediators, gatekeepers, information channels and reputational advantages) and Structuralist (innovative milieu as well as geographical agglomeration effects (Westhead & Batstone, 1998)). The network benefits of a Science Park can be financial (promoting access to investments (Koh, et al., 2005)); commercial (access to potential clients within the park); and organisational, (non-scientific or technical expertise from Science Park administration itself as well as on-park firms.)

Resources, networks and social capital

Literature related to the social capital of firm founders is of particular interest. Human and social capital studies are numerous (Audretsch, et al., 2005; Lanciano-Morandat, et al., 2009), with studies from strategic alliances (T.K. Das & Teng, 2000; Deeds & Hill, 1996) and entrepreneurial development (Ho & Wilson, 2007) also referencing the deployment of social capital.

Social capital can best be described as supplementary and enabling resources, complementing the stock knowledge, financial capital and skills of an entrepreneur (Greve & Salaff, 2001). Entrepreneurial activity is often marked by the ability of a firm founder to mobilise such social capital through their familial and social relations, as well as the networks they develop on entry to a field. For firms choosing to locate in a science park, social capital takes on physical- and market- proximity aspects (Sorenson, 2003). The presence of a university nearby also opens up the option of entraining and accruing (if the university is an alma mater) academic social capital resources. A general model related to social capital is that of Elfring & Hulsink (2003) in which operate three processes. *The discovery of opportunities* (prior knowledge about the opportunity), *securing resources* (accessing, mobilising and deploying resources), and *obtaining legitimacy* (enhancing visibility through affiliations, alliances and networks).

Aim

We previously researched (Lanciano-Morandat, et al., 2009) the development of the social networks of biotechnology-oriented firm founders. We build on this study by examining the role of science parks in facilitating the social capital of firm founders.

We do so by investigating:

(a) The social interactions in differing contextual settings between firm founder(s) and stakeholders within, and outside, a science park;
(b) The social interactions with, or mediated by, the science park administration.

Method

Science Park and firm selection

Leiden BioScience Park (LBP) is the subject of our analysis. LBP is a biomedical science cluster in the Netherlands. Nearby HEIs include Leiden University (and their Medical Centre), The Hogeschool Leiden (Applied Sciences) and other knowledge institutions such as TNO and Top Institute Pharma. There is a park administration partnered with a facilities and strategy manager.

The firms were selected on 3 primary criteria: firm formation was within the last 10 years; the firm was founded by a university or knowledge institute researcher; and lastly, the firm is from the life sciences and health sector. Following these criteria, we were able to interview and collect full patent and publication data for 9 firms.

Interviews

The interviews were semi-structured with a pre-determined list of topics to be discussed. If any of the topics were not discussed in the interview, they were asked as direct questions at the end of the

interview. The topics revolved around the nature of interactions between the firm founder and stakeholders involved during the development of the firm. The topics and interviewing coding typology (from Lanciano-Morandat et al. (2009) concerned:

(1) Actor types:
- Academia (e.g. universities, scientific advisors or students)
- Organisational and training groups (e.g. patient groups, consortia or professional networks),
- Finance (e.g. banks or venture capital firms),
- Commerce (e.g. customers, marketing firms or suppliers)
- Industrial partners (e.g. manufacturers, other biotechnology firms or pharmaceutical partners)
- Policy and regulation (e.g. lawyers, trial administrators/enablers or safety officials)
- Science Park (e.g. Science Park administrators or facilities management).

(2) The specific context of the interaction which included:
- Scientific – scientific knowledge involved in developing the firms' products or processes;
- Commercial – commercial or sales aspect of the products or processes offered by the firm;
- Financial – venture-capital or grants and the like;
- Technical – technical, including equipment, knowledge required for the functioning of research activities;
- Organisational – regulatory and/or administrative requirements for product/process development or for firm operation.

(3) The proximity of the interacting individual/institution of which the entity could be:
- Personally related – in which the entity is/was a family member, friend or close acquaintance known before the firm was founded;
- Not personally related but within the physical confines of the Science Park;
- Not personally related but from outside the physical confines of the Science Park.

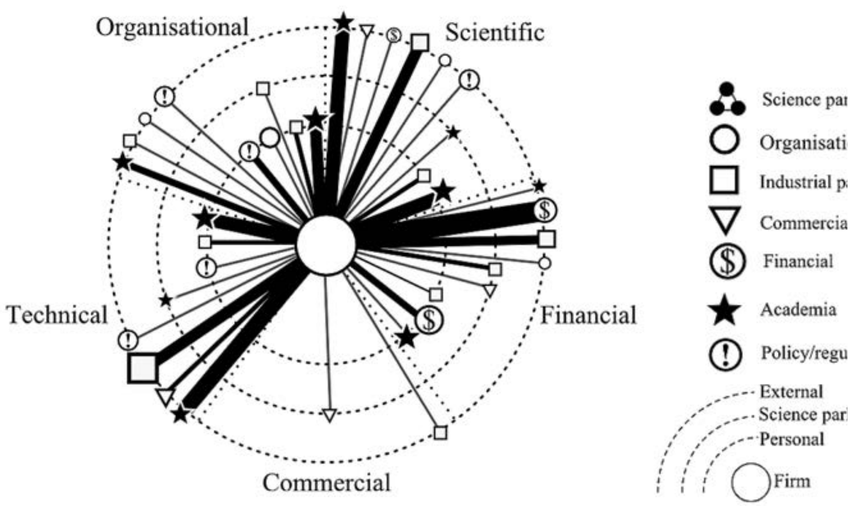
(4) The interactions are classed according to phases in the formation of the firm, specifically:
- The pre-entrepreneurial phase (this includes the time before the firm was officially incorporated, to shortly after incorporation)
- Entrepreneurial phase (wherein the technology of the firm was believed to have surpassed its viability phase)
- Managerial phase (where the duties of the founder as CEO have migrated to that of CSO)

Due to the content and depth of the issues discussed with the interviewees and in accordance with confidentiality agreements with the interviewees/firms, the results presented have been generalised with all identifying data removed.

Results

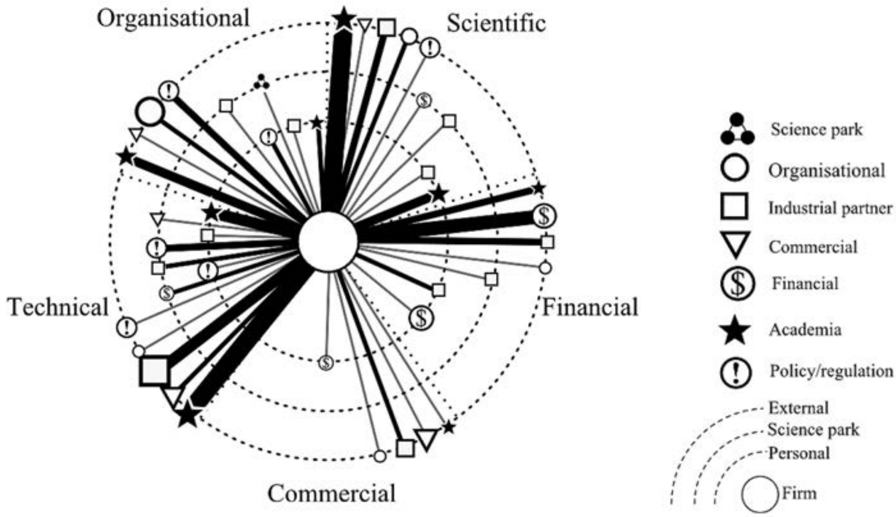
In Figure 1, in terms of funding sources, there are only incidental interactions with financial actor types in the scientific sphere, with those being small financial interactions (such as rent payments) with universities. The bulks of interactions over funding resources come from the financial sphere, and are from internal and externally-oriented interactions. In terms of proximity, personal interactions are primarily knowledge and financially oriented, along with some interactions with industrial partners. External interactions are primarily related to funding sources, industrial partners and knowledge sources.

Figure 1. Leiden collective interactions during the pre-entrepreneurial phase. *(Note: Size of node indicates average number of interactions. Edge thickness signifies count of firms reporting interactions. Grey edges signify only one firm reporting interaction.)*



In Figure 2, there was an increase in the number of commercial interactions, as firms were securing their first customers. An increase was also seen in the number of interactions with academic actor types within the technical sphere.

Figure 2. Leiden collective interactions during the entrepreneurial phase. *(Note: Size of node indicates average number of interactions. Edge thickness signifies count of firms reporting interactions. Grey edges signify only one firm reporting interaction.)*

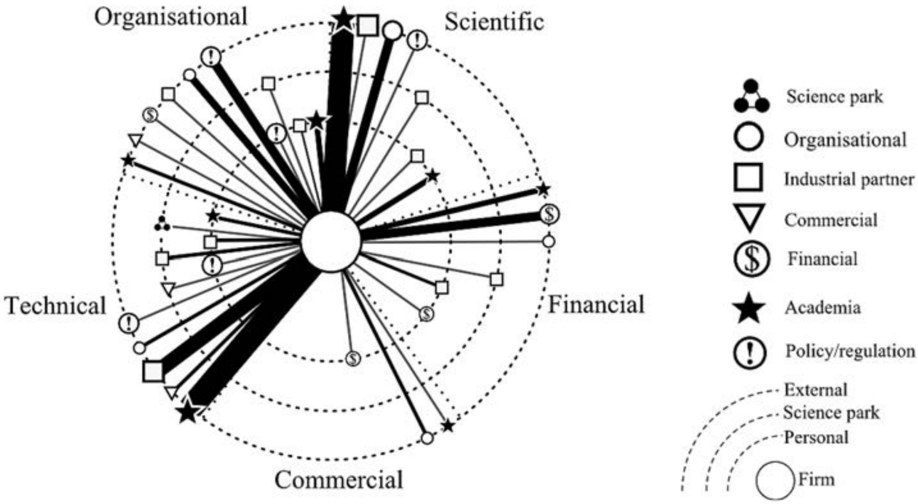


In terms of knowledge sources, most relations (and the strongest) are with academia and external to the science park in nature. In contrast to the pre-entrepreneurial phase, more firms report knowledge sources from the external relations in the financial sphere. The external interactions reported by the founders were overwhelmingly international. The founders commonly reported interactions with universities or public research institutes. However, these interactions were the result of many of the firm founders being active faculty within the universities named. In terms of proximity personal relations seem less important; the occasional firm mentions funding or industrial partners and the most pronounced are relations with academia. The role of the Science Park is more pronounced and diverse than in the pre-entrepreneurial phase, especially in the technical sphere. The Science Park itself appears as an actor in the organisational sphere and its main function seems to be to provide access to industrial partners. External relations are to knowledge sources in the scientific and technical spheres, and to industrial partners in the scientific, technical, and financial spheres. There is also a degree of emerging commercial relations.

In Figure 3, relations with policy/regulators are stronger than in preceding phases. Funding has become gradually less important in the networks of founders; mostly in the financial sphere and external, and the financial sphere is less prominent than before. In relation to network partners, the Science Park mainly serves to find industrial partners and these are found in primarily in the technical sphere and not at all in the commercial sphere. In terms of proximity, personal relations primarily draw in knowledge, industrial partners and occasionally, funding opportunities. Relations within the Science Park draw in industrial partners and little else, and as such the

Science Park now has a role in the technical sphere. External interactions result in knowledge from the scientific and technical spheres, industrial partners and organisational relations.

Figure 3. Leiden collective interactions during the managerial phase. *(Note: Size of node indicates average number of interactions. Edge thickness signifies count of firms reporting interactions. Grey edges signify only one firm reporting interaction.)*



Discussion and conclusions

The sum of the reported interactions runs contrary to many of the stated goals of a Science Park, where the firms within the park and the Science Park administration interact, exploiting the network benefits of locating on a Science Park. The principal scientific and technological pull of the Science Park as reported by the founders was the proximity to the local HEI, a motivation also reported in Löfsten & Lindelöf (2003).

The motivations cited by the firm founders to locate within the Science Park were mostly in line with neo-classical theory, where transport access and distance were important determinants to location, although one firm reported commercial and collaborative opportunities initiated by the Science Park. In terms of the Science Park being a potential innovative milieu, such as with the structuralist theories of Westhead & Batstone (1998), we found little evidence. The interactions of the firms that led to innovative outcomes were primarily with external HEIs or industry with significant contact between the firms and Leiden University, even if founders claimed another university as their alma mater. However, it is important to repeat here that for this study we considered the local university to be external to the Science Park.

The social capital of the firms was primarily determined by relations external to the science park. The networks developed by the firm founders prior to formation and with external entities were more significant sources of social capital than the Science Park or firms within the Science Park. In terms of Elfring and Hulsink's (2003) processes in developing social capital, the *discovery of opportunities*, *securing resources*, and *obtaining legitimacy*, many of the opportunities afforded to the firms came from prior personal relations. Similarly the resources acquired and mobilised by the firms were also external in nature. There was an increase of visibility and these were through collaborative efforts between the firm founders and their alma maters in terms of scientific research, and industrial partners, of which all were external to the Science Park.

This study is not without limitations. Our sample size is restricted by our selection criteria, and the diversity of the firms, in terms of their origins, sub-fields and products, is too high to conduct statistical comparisons. The level of detail and the amount of effort involved in data processing and cleaning makes our method difficult to replicate on a large scale. For our future research plans, we aim to streamline the analysis of firms.

We believe that the level of detail in our study outweighs the restrictive selection criteria and adds a new dimension to future studies on Science Parks and academic entrepreneurs who choose to locate to Science Parks. Our qualitative approach can be of help to policymakers to re-examine the purported benefits of a Science Park and if a Science Park is in fact the ideal carrier of these benefits.

The results of this paper add new weight to the need for a careful re-examination of the role of the Science Park in the regional and national spheres. As the Netherlands is a geographically compact country, many of the logistical benefits may be moot. However, our results seem to suggest that the close association of start-ups with the local university and with national industrial partners lead to a more innovative firm.

References

Audretsch, D. B., et al. (2005). University spillovers and new firm location. *Research Policy*, 34(7), 1113–1122.

Das, T. K., & Teng, B. S. (1997). Time and Entrepreneurial Risk Behavior. *ENTREPRENEURSHIP THEORY and PRACTICE*, 22(2), 69–71.

Das, T. K., & Teng, B. S. (2000). A resource-based theory of strategic alliances. *Journal of management*, 26(1), 31–61.

Deeds, D. L., & Hill, C. W. L. (1996). Strategic alliances and the rate of new product development: An empirical study of entrepreneurial biotechnology firms. *Journal of Business Venturing*, 11(1), 41–55.

Dettwiler, P., et al. (2006). Utility of location: A comparative survey between small new technology-based firms located on and off Science Parks--Implications for facilities management. *Technovation*, 26(4), 506–517.

Elfring, T., & Hulsink, W. (2003). Networks in entrepreneurship: The case of high-technology firms. *Small business economics*, 21(4), 409–422.

Fukugawa, N. (2006). Science parks in Japan and their value-added contributions to new technology-based firms. *International Journal of Industrial Organization*, 24(2), 381–400.

Greve, A., & Salaff, J. W. (2001). The development of corporate social capital in complex innovation processes. In S. M. Gabbay & R. Leenders (Eds.), *Social Capital of Organizations* (Vol. 18, 107–134): Emerald Group Publishing Limited.

Ho, M. W. Y., & Wilson, M. (2007). Biotechnology founders and employment systems of start-ups. *International Journal of Technology*, Policy and Management, 7(3), 263–279.

Koh, F. C. C., et al. (2005). An analytical framework for science parks and technology districts with an application to Singapore. *Journal of Business Venturing*, 20(2), 217–239.

Lanciano-Morandat, C., et al. (2009). Le capital social des entrepreneurs comme indice de l'émergence de clusters? *Revue d'économie industrielle*(4), 177–205.

Löfsten, H., & Lindelöf, P. (2003). Determinants for an entrepreneurial milieu: Science Parks and business policy in growing firms. *Technovation*, 23(1), 51–64.

Löfsten, H., & Lindelöf, P. (2005). R&D networks and product innovation patterns--academic and non-academic new technology-based firms on Science Parks. *Technovation*, 25(9), 1025–1037.

Phan, P. H., et al. (2005). Science parks and incubators: observations, synthesis and future research. *Journal of Business Venturing*, 20(2), 165–182.

Phillimore, J. (1999). Beyond the linear view of innovation in science park evaluation An analysis of Western Australian Technology Park. *Technovation*, 19(11), 673–680.

Quintas, P., et al. (1992). Academic-industry links and innovation: questioning the science park model. *Technovation*, 12(3), 161–175.

Siegel, D. S., et al. (2003). Assessing the impact of university science parks on research productivity: exploratory firm-level evidence from the United Kingdom. International *Journal of Industrial Organization*, 21(9), 1357–1369.

Sorenson, O. (2003). Social networks and industrial geography. *Journal of Evolutionary Economics*, 13(5), 513–527.

Squicciarini, M. (2008). Science Parks' tenants versus out-of-Park firms: who innovates more? A duration model *The Journal of Technology Transfer*, 33(1), 45–71.

Westhead, P., & Batstone, S. (1998). Independent Technology-based Firms: The Perceived Benefits of a Science Park Location. *Urban Studies*, 35(12), 2197–2219.

From inception to exploitation: research trails in biotechnology start-ups

Thomas Gurney t.gurney@rathenau.nl / Science System Assessment, Rathenau Instituut, Anna van Saksenlaan 51, The Hague, 2593 HW (The Netherlands)

Edwin Horlings e.horlings@rathenau.nl / Science System Assessment, Rathenau Instituut, Anna van Saksenlaan 51, The Hague, 2593 HW (The Netherlands)

Peter van den Besselaar p.a.a.vanden.besselaar@vu.nl / Department of Organisation Science & Network Institute, VU University Amsterdam, De Boelelaan 1105, 1081HV, Amsterdam (The Netherlands)

Abstract

The pathway of an idea generated in academic and exploited in industry is complex. For start-ups located in Science Parks, the skills and knowledge acquired during previous research needs to be successfully exploited. In this study we collect patent and publication data on single-site start-ups and their founders, located in the Leiden Bioscience Park. Using these data we examine in detail the contributions of the founders' host universities, other firms located in the Science Park, and the internationalism characteristics of the firms' collaborations pre- and post-incorporation. We link the patent applications of the firms to the firm founders' publication through the literature cited by the patent applications. We find that the sample set of firms incorporate, in addition to their own continued streams, new streams of academic research with their technological output with publishing and patenting partners varying according to their proximity to the park.

Introduction

From inception to exploitation, the knowledge pathway follows a convoluted route. For firms located on Science Parks, the transformation from exploratory research to exploited artefacts is an end-goal of the firm and, by extension, the host Science Park. However, this process begins before firm incorporation and before locating on a Science Park. Without this transformation, all the economic, technological, scientific and social benefits a Science Park purportedly offers are moot. It is for this reason that our study aims to examine, in detail, the knowledge capture and transformation mechanisms that are present from the start, based on the initial carrier of this knowledge, the firm founder, in the environment of the firm, the Science Park.

Conceptual framework

Science Parks

Science parks provide an environment for large firms to develop relationships with small firms; promote formal and informal links between firms, universities and other small labs (Siegel, et al., 2003). They provide a contact space between applied and basic science (Quintas, et al., 1992); promote foreign investment transition to a knowledge-based economy (Koh, et al., 2005). They provide technological development and renewal on a regional or national basis (Phillimore, 1999). Firms choosing to locate on Science Parks do so based on combinations of neo-classical theory (transport, labour costs, agglomeration economies) or behavioural aspects (the Science Park acts as mediator, gatekeepers, add to reputation) or structuralist theories (Science Parks create an innovation milieu) (Westhead & Batstone, 1998).

Knowledge capture

Access to resources, assets and capabilities include not only the contextual and network benefits of a Science Park but also the skills and knowledge of the personnel of the firm. These begin with, and is largely shaped by, the firm founder (Bozeman, et al., 2001). The decision to exploit the base knowledge stock is complex, and the process of becoming an "entrepreneurial scientist" (Oliver, 2004) is difficult. The process of recognising and incorporating new knowledge is that of absorptive capacity (Cohen & Levinthal, 1990). With a mind to incentives, the scientist employs a number of strategic decisions (Horlings & Gurney, 2012), guided by the perception of the trade-off between field crowdedness, problem difficulty and potential reputational gains (Zuckerman & Cole, 1994). After the firm has been incorporated, the firm founder takes into account new stakeholders and the need to balance exploiting current knowledge developing new knowledge.

Aim

We have previously developed and extended methodological tools that describe the search strategies, and exploitation activities, of researchers (Gurney, et al., 2012; Horlings & Gurney, 2012). Based on these, we aim to examine:

– The cognitive routes and developments of an idea generated in academia and exploited in the setting of a Science Park.

We do so by investigating:

(a) The links between firm founder(s) knowledge stocks and their technological output including the scientific and technological links to firms in the Science Park, and HEIs and public research institutes in close geographic proximity to the Science Park;

(b) The continuity of research conducted by the firm founder(s) – prior to firm incorporation and at certain periods after incorporation;
(c) The composition of academic and industrial collaborations of the firm and firm founder(s) including the regionalism/internationalism of collaborators;

Method

Science Park and firm selection

Leiden BioScience Park is a biomedical cluster in The Netherlands and the subject of our analysis. Nearby HEIs include Leiden University (and their Medical Centre), The Hogeschool Leiden (Applied Sciences) and other knowledge institutions such as TNO and Top Institute Pharma. There is a park administration partnered with a facilities and strategy manager. The firms were selected on 3 primary criteria: firm formation was within the last 10 years; the firm was founded by a university or knowledge institute researcher; and lastly, the firm is from the life sciences and health sector. Following these criteria, we collected full patent and publication data for 9 firms.

Patents and publications

For patent data we use the PatSTAT database prepared and developed by the EPO. All patent applications of the firm or firm founder(s) listed as an applicant, or firm founder(s) listed as inventor, were collected. We used Thomson Reuter's Web of Science (WoS), supplemented with CV data from the scientists involved to collect all publications of the firm founder(s) (up to June 2012).

These base data were parsed using SAINT (2009) and managed in a relational database. Further data were collected from the patents, specifically:

(1) In-text non-patent literature references (IT-NPLRs) – citations to publications visible in the body of the patent.
(2) Bibliographic NPLRs (B-NPLRs).

Both NPLR sets were parsed and, as far as possible, their WoS publication equivalents retrieved. The verified documents were then parsed and processed separately for a per-firm analysis and collectively for a group analysis. The addresses found within the publication and patent application data were coded for country of origin and type of entity.

Patent and publication visualisation and analysis

Publications were clustered by their shared combinations of title words and cited (van den Besselaar & Heimeriks, 2006). The degree of similarity was calculated using the Jaccard similarity coefficient. Clusters of publications were automatically assigned by their degree centrality and relative

weights of edges between nodes (Blondel, et al., 2008). The NPLR of the patent applications were included in the clustering of the publications.

Using NPLRs we linked the publications to the patent applications. Even if the patent applications do not directly cite the work of the founder(s), the NPLR that are cited cluster within the founder(s) corpus, inferring a link to the founder(s) areas of expertise. As a result, an indication of the degree of knowledge transfer and absorptive capacity from the research practices and results of the founder(s) to their technological output can be observed and elucidated.

Results

Scientific and technological output relevance

In Table 1, the publishing propensity and diversity differs by firm founder. The continuity of research involving the founder is low for all firms except one (Firm 2). The number of active streams with NPLRs co-clustering within the stream indicates a link to the founder(s) publication corpora relevant at the start of the firm.

Table 1. Founder(s) publishing research streams.

	Number of Streams					
FIRM ID	Total	Active at incorporation	New after incorporation	Active 1 year after incorporation	Active 3 years after incorporation	Active at incorporation with NPLR
1	9	5	0	5	3	5
2	20	10	2	9	7	7
3	7	1	1	1	0	0
4	7	3	1	3	3	3
5	14	3	1	2	0	0
6	4	0	2	0	0	0
7	2	1	1	1	1	1
8	3	0	0	0	0	0
9	10	1	2	1	1	0

Academic and industrial collaborations

Table 2a shows that for most of the firms with active research streams at incorporation, the authors primarily come from academia. Firm 3 and 9 have a large portion of their authors coming from industry, with Firm 9 showing a varied mix of academic and/or industrial authors.

Table 2a. Academic and industrial collaboration composition of founder(s) publishing research streams.

	Percentage author composition of papers in research streams active at incorporation				
FIRM ID	**Academic**	**Predominantly Academic**	**Academic and Industrial**	**Predominantly Industrial**	**Industrial**
1	89	1	7	1	2
2	98	1	1	0	0
3	67	0	0	33	0
4	99	0	1	0	0
5	93	0	7	0	0
6	-	-	-	-	-
7	82	9	9	0	0
8	-	-	-	-	-
9	30	10	40	20	0

Table 2b presents the assignee composition of the patenting efforts of all the firms. Firms 3, 7 and 8 have industrial assignees exclusively, and Firm 9 has a vast majority of industrial assignees (94%). Firm 1 patents almost equally with academic and industrial assignees, whilst the rest of the firms tend to academic assignees.

Table 2b. Academic and industrial collaboration composition of patent assignees pre- and post-incorporation.

		Number of patent applications and count of assignee origin pre/post inc.				
		Academic	**Predominantly Academic**	**Academic and Industrial**	**Predominantly Industrial**	**Industrial**
Firm ID	**Total**	**Pre/Post**	**Pre/Post**	**Pre/Post**	**Pre/Post**	**Pre/Post**
1	67	16/14	0/0	19/0	3/2	9/4
2	13	7/5	0/0	0/0	1/0	0/0
3	6	0/0	0/0	0/0	0/0	4/2
4	16	8/3	0/0	0/0	0/0	2/3
5	7	-/0	-/0	-/0	-/0	-/7
6	30	-/0	-/0	-/9	-/1	-/20
7	10	0/0	0/0	0/0	0/0	2/8
8	9	0/-	0/-	0/-	0/-	9/-
9	45	3/0	0/0	0/0	0/0	34/8

Local, regional and international collaborations

Tables 3a and 3b present academic collaborations of all the firms in our set. Table 3a indicates the geographic distribution of academic collaborators in the publishing streams of the firms. All the

firms publish with at least one Dutch academic address. Firm 7 publishes exclusively with Dutch academic partners and Firms 1 and 8 publish extensively with only Dutch partners. Firm 5 is the more international wherein it publishes almost exclusively with Dutch and RoW academic partners.

Table 3a. International composition of **academic collaborators** *of founder(s) publishing research streams.*

	Percentage composition of academic collaborators in research streams active at incorporation			
FIRM ID	**NL only**	**NL & EU**	**NL & RoW**	**NL & EU & RoW**
1	71.3	14.3	6.4	8.1
2	36.2	21.5	36.9	5.5
3	50	16.7	33.3	0
4	31.7	57.1	4.1	7.2
5	0	0	95.4	4.6
6	-	-	-	-
7	100	0	0	0
8	-	-	-	-
9	80	0	10	10

Links with Leiden University are considered integral to the Science Park, and are reported as such by all the firms in the set. Table Results 3b shows the composition of academic collaborations with Leiden University, other Dutch universities and Dutch knowledge institutes.

Table 3b. Composition of **Dutch academic collaborators** *of founder(s).*

	Percentage presence of Leiden University, other NL universities or knowledge institutes(KI)											
	Active at incorporation			New after incorporation			Active 1 year after incorporation			Active in 2011		
FIRM ID	**Leiden**	**Other Uni**	**KI**	**Leiden**	**Other Uni**	**KI**	**Leiden**	**Other Uni**	**KI**	**Leiden**	**Other Uni**	**KI**
1	81.3	7.3	3.8	-	-	-	81.3	7.3	3.8	75.9	12	0
2	22	23.5	11.2	100	100	0	22	23.5	3.9	23.1	22.2	3.9
3	0	20	80	0	0	0		20	80	-	-	-
4	95.5	2.7	2.7	66.6	16.6	0	95.5	27	2.7	86.5	2.7	2.7
5	0	0	0	0	0	0	0	0	0	-	-	-
6	-	-	-	75	62.5	0	-	-	-	75	25	0
7	76.9	7.7	7.7	62.5	18.7	0	76.9	7.7	7.7	69.7	15	7.7
8	-	-	-	-	-	-	-	-	-	-	-	-
9	90.9	9.1	0	55.7	27.1	20	90.9	9.1	0	67.4	21.1	20

For firms with active research streams at incorporation, the vast majority of academic collaborations are with Leiden University. This is to be expected as many of the firm founder(s) maintain active faculty positions at Leiden. There are also large numbers of collaborations with other Dutch universities. For all firms, the vast majority of academic patent collaborators of the firms are from The Netherlands except for Firm 2 who collaborates with RoW academic partners on just under a third of their pre-incorporation patent applications.

Table 4 shows the international composition of the industrial assignees of the patent applications of the firms, pre- and post-incorporation. Half the firms began to develop their patent stock prior to incorporation, and 3 of the firms developed over 80% of their knowledge stock after incorporation. All but 3 of the firms have only Dutch and/or EU industrial assignees. However, none of these collaborations are with firms located in the Science Park. The patent stocks of the 3 firms with RoW industrial partners were developed prior to incorporation. Significantly, there are no other firms from the Science Park listed as assignees in any of the 9 firms' patent applications.

Table 4. International composition of **industrial patent assignees***.*

	Industrial co-assignees as a percentage of total assignees pre- and post-incorporation			
	NL only	**NL & EU**	**NL & RoW**	**NL & EU & RoW**
Firm ID	**Pre/Post**	**Pre/Post**	**Pre/Post**	**Pre/Post**
1	62.2/13.5	18.9/2.7	2.7/0	0/0
2	0/-	0/-	0/-	100/-
3	0/28.6	28.6/0	42.9/0	0/0
4	40/40	0/20	0/0	0/0
5[a]	-/100	-/0	-/0	-/0
6[a]	-/100	-/0	-/0	-/0
7	0/80	20/0	0/0	0/0
8	5.3/52.6	42.1/0	0/0	0/0
9	81/19	0/0	0/0	0/0

a All applications are with founder firm as only industrial assignee.

Conclusions and discussion

The firm founder begins, in most cases, as an academic researcher. Much of the scientific and technical expertise developed previously was still employed by the firms at the end date of our data collection. Most firms struck balances between explorative activities and exploitative activities in punctuated equilibrium, a reported process described by Gupta et al. (2006).

The alma mater universities of the founders played a large role in the initial IP development, but after incorporation industrial partners soon supplant the universities. The share of international

collaborators in the academic output of the founders is high to begin with for most of the firms. The number of international *industrial* collaborations varies across the firms, both in patenting and publishing.

After incorporation, only firms with founders still active at the university continue a high level of international academic collaboration. For firm founders who are not active faculty, publishing activities become increasingly local, mostly with their alma mater and increasingly with Leiden University. Most of the founders interviewed found that customers for the products and services they provide are not local or even regional. Rather they are mostly from the EU or further. This would suggest that whilst there may not be a local market for their products and services, the research conducted towards developing the products and services is strongly helped by the local education and industrial sectors.

Our sample size is restricted by our selection criteria, and the diversity of the firms is too high to conduct statistical comparisons but we believe that the level of detail in our study outweighs the restrictive selection criteria and adds a new dimension to future studies on academic entrepreneurs and Science Parks. Our study can provide insight in further studies for policymakers as to the historical development and level of collaboration between firms located on Science Parks and the internationalism of academic or industrial collaborations.

Our results seem to suggest that the close association of start-ups with the local university and with national industrial partners lead to a more innovative firm. A better matching process of the products and services offered by the firms to local or national customers could boost regional development through better identification of natural complementarities between research conducted in firms and with potential customers.

References

Blondel, V. D., et al. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: *Theory and Experiment*, 2008, P10008.

Bozeman, B., et al. (2001). Scientific and technical human capital: an alternative model for research evaluation. *International Journal of Technology Management*, 22(7), 716–740.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, 35(1, Special Issue: Technology, Organizations, and Innovation), 128–152.

Gupta, A. K., et al. (2006). The interplay between exploration and exploitation. *Academy of Management Journal*, 49(4), 693–706.

Gurney, T., et al. (2012). *Knowledge Capture Mechanisms in Bioventure Corporations*. Paper presented at the 17th International Conference on Science and Technology Indicators (STI), Montreal.

Horlings, E., & Gurney, T. (2012). Search strategies along the academic lifecycle. *Scientometrics*, 1–24.

Koh, F. C. C., et al. (2005). An analytical framework for science parks and technology districts with an application to Singapore. *Journal of Business Venturing*, 20(2), 217–239.

Oliver, A. L. (2004). Biotechnology entrepreneurial scientists and their collaborations. *Research policy*, 33(4), 583–597.

Phillimore, J. (1999). Beyond the linear view of innovation in science park evaluation An analysis of Western Australian Technology Park. *Technovation*, 19(11), 673–680.

Quintas, P., et al. (1992). Academic-industry links and innovation: questioning the science park model. *Technovation*, 12(3), 161–175.

Siegel, D. S., et al. (2003). Assessing the impact of university science parks on research productivity: exploratory firm-level evidence from the United Kingdom. *International Journal of Industrial Organization*, 21(9), 1357–1369.

Somers, A., et al. (2009). *Science Assessment Integrated Network Toolkit (SAINT): A scientometric toolbox for analyzing knowledge dynamics*. The Hague: Rathenau Insitute.

van den Besselaar, P., & Heimeriks, G. (2006). Mapping research topics using word-reference co-occurrences: a method and an exploratory case study. *Scientometrics*, 68(3).

Westhead, P., & Batstone, S. (1998). Independent Technology-based Firms: The Perceived Benefits of a Science Park Location. *Urban Studies*, 35(12), 2197–2219.

Zuckerman, H., & Cole, J. R. (1994). Research strategies in science: A preliminary inquiry. *Creativity Research Journal*, 7(3–4), 391–405.

On the relation between altmetrics and citations in medicine[1]

Stefanie Haustein stefanie.haustein@umontreal.ca / École de bibliothéconomie et des sciences de l'information, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, QC. H3C 3J7 (Canada) / Science-Metrix Inc., 1335 A avenue du Mont-Royal E, Montréal, Québec H2J 1Y6 (Canada)

Mike Thelwall m.thelwall@wlv.ac.uk / School of Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY (UK)

Vincent Larivière vincent.lariviere@umontreal.ca / École de bibliothéconomie et des sciences de l'information, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, QC. H3C 3J7 (Canada) / Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, CP 8888, Succ. Centre-Ville, Montréal, QC. H3C 3P8 (Canada)

Cassidy R. Sugimoto sugimoto@indiana.edu / School of Information and Library Science, Indiana University Bloomington, 1320 E. 10th St. Bloomington, IN 47401 (USA)

Introduction

Altmetrics have been introduced as complements to citation indicators to determine the relevance and impact of scholarly documents. Systematic evidence as to whether altmetrics are valid indicators of impact and relevance is, however, still missing. This study aims to determine the relationship between eight altmetrics and citations for PubMed articles. Unlike previous studies, we do not rely on correlation coefficients as they are subject to negative biases caused by citation delays and the increasing uptake of social media. Instead we apply the sign test, which eliminates these biases by comparing only documents published at the same point in time.

Methods

This paper compares eight different altmetric sources with citation data for up to 135,331 (depending on the metric) PubMed documents published between 2010 and 2012 (Thelwall, Haustein, Larivière, & Sugimoto, 2013). Citations were counted until October 2012 and exclude first-author self-citations. Altmetric data was collected for Twitter, Facebook, Google+, Research Highlights in Nature journals, a set of 2,200 science blogs, 60 newspapers and magazines, two forums and Reddit by altmetric.com through licensed firehoses, APIs or via scraping the websites, counting an event each time a document is mentioned on these platforms. For each altmetric a specific

1 This research was funded by AHRC/ESRC/JISC (UK), SSHRC (Canada), and the National Science Foundation (US; grant #1208804) through the Digging into Data funding program.

dataset was created containing all documents with the particular altmetric and citation scores. Documents with zero-altmetric scores were excluded for the particular altmetric.

To find out whether the various altmetrics indicate impact similar to that measured by citations, both kinds of metrics must be compared. Previous studies have used correlation coefficients and found low to medium positive correlations, e.g. between citations and Tweets (Eysenbach, 2011; Shuai, Pepe, & Bollen, 2012) or readership counts from social reference managers (Bar-Ilan, 2012; Bar-Ilan et al., 2012; Li & Thelwall, 2012; Li, Thelwall, & Giustini, 2012; Priem, Piwowar, & Hemminger, 2012). The problem with altmetrics data is that there is a bias towards more recent papers as they are strongly influenced by increases in social media use. For recent papers the citation window is also too small. As correlation tests are influenced by these biases towards negative correlations, they are not ideal for comparing citations and altmetrics.

A simple sign test was used as a fair way to determine whether higher altmetrics values associate with higher citation counts. It compares each document's citation and altmetric scores only to those of the two articles published immediately before and after which are equally exposed to citation delays and social media uptake. The sign test differentiates between three possible outcomes: a) success (both altmetrics and citations for the middle article are higher than the average of the other two articles *or* both lower than average), b) failure (altmetrics higher and citation lower *or* altmetrics lower and citation higher than average) and c) null (all other cases). Success thus indicates a positive and failure a negative association between citations and altmetrics. Chronological order for the test was determined through the PubMed ID. A simple proportion test was used for each altmetric to see whether the proportion of successes was significantly different from the default of 0.5.

Results

The sign test distinguished significant results for six out of the eight altmetrics. No significant results could be found for Google+ and Reddit. Overall, there are no cases where the number of failures is higher than the number of successes and so this suggests that, given sufficient data, all the altmetrics would also show a significantly higher success than failure rate.

Table 1. The number of successes and failures as determined by the sign test comparing citations and altmetric scores.

Metric	Successes	Failures
Twitter**	24315 (57%)	18576 (43%)
Facebook walls**	3229 (58%)	2383 (42%)
Research Highlights**	3852 (56%)	3046 (44%)
Blogs**	1934 (60%)	1266 (40%)

… Continuation Table 1

Metric	Successes	Failures
Google+	426 (53%)	378 (47%)
Newspapers and magazines**	338 (59%)	232 (41%)
Reddit	103 (56%)	81 (44%)
Forums**	19 (86%)	3 (14%)

*Ratio significantly different from 0.5 at p = 0.05
**Ratio significantly different from 0.5 at p = 0.01

Conclusion

The results provide strong evidence that six of the eight altmetrics associate with citation counts, at least in medical and biological sciences but only to a small extent. Qualitative research is needed to determine the actual validity of different social media mentions as metrics to be used in information retrieval and research evaluation.

References

Bar-Ilan, J. (2012b). JASIST 2001–2010. *Bulletin of the American Society for Information Science and Technology*, 38(6), 24–28.

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). Beyond citations: Scholars' visibility on the social Web. In *Proceedings of the 17th International Conference on Science and Technology Indicators*, Montréal, Canada (Vol. 1, 98–109).

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research*, 13(4), e123.

Li, X., & Thelwall, M. (2012). F1000, Mendeley and traditional bibliometric indicators. In *Proceedings of the 17th International Conference on Science and Technology Indicators*, Montréal, Canada (Vol. 2, 451–551).

Li, X., Thelwall, M., & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement. *Scientometrics*, 91(2), 461–471.

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. http://arxiv.org/abs/1203.4745v1.

Shuai, X., Pepe, A., & .Bollen, J. (2012). How the scientific community reacts to newly submitted preprints: Article downloads, twitter mentions, and citations. *PLoS ONE*, 7(11), e47523.

Thelwall, M., Haustein, S., Larivière, V. & Sugimoto, C. (2013). Do altmetrics work? Twitter and ten other candidates. *PLOS ONE*, 8(5), e64841.

How Smart is Specialisation? An Analysis of Specialisation Patterns in Knowledge Production

Gaston Heimeriks gheimeriks@gmail.com / Innovation Studies (Copernicus Institute of Sustainable Development) Utrecht University, Heidelberglaan 2, 3584 CS Utrecht (The Netherlands)

Abstract

This paper examines the evolutionary patterns of knowledge production in Astrophysics, Biotechnology, Nanotechnology and Organic Chemistry between 1995 and 2009. In all fields, the rise and fall of research organisations over time can be attributed to their specialisation pattern of scientific knowledge production. However, the fields are characterized by distinct co-evolutionary dynamics. It is shown that the patterns of specialisation differ systematically across scientific fields, but are remarkably similar across organisations in each field. Two patterns of specialisation are identified. The first represents a turbulent pattern: concentration of research activities is low, knowledge producing organisations are of small size in terms of output, stability in the ranking is low and comparative advantages are short lasting. The second represents a stable pattern: concentration of research activities is higher than in the first group, research organisations are of larger size, stability in the ranking is greater, and comparative advantages last longer. The former group comprises biotechnology and organic chemistry, while the latter includes astrophysics and (in later years) nanotechnology. These fields differ in the number and specific nature of the capabilities they require and require different policy measures that match the existing place dependencies on the organisational level as well as the global dynamics of specific fields in order to achieve smart specialisation.

Introduction

In this study, we explore the specialisation patterns of knowledge production in in Astrophysics, Biotechnology, Nanotechnology and Organic Chemistry between 1995 and 2009. From an evolutionary perspective, we argue that the cumulative and path-dependent nature of scientific knowledge production makes it also place-dependent. This implies that research organisations are likely to specialize over time (Heimeriks & Boschma, 2012). At the same time, knowledge production is also subject to dynamics: new scientific topics and fields emerge, and new research organisations come into existence across the globe. The aim of this paper is to quantify these evolutionary patterns of knowledge production in different fields and to show how path and place dependent specialisation patterns contribute to the rise and fall of research organisations. Shift-share analysis enables to specify the changes in a set of locations over time and entropy statistics provides insight in the co-evolutionary nature of topics and research organisations.

Addressing the issue of specialisation in research and innovation is a crucial policy issue, especially for regions and organisations that are not leaders in any of the major science or technology domains (Foray, David, & Hall, 2009). This is an especially pressing issue in emerging sciences such as Biotechnology and Nanotechnology, that are characterized by rapid growth, divergent dynamics, and new complementarities creating the need for wide-ranging cross-disciplinary competences (Bonaccorsi, 2008). Grasping the fruits of these emerging techno-sciences is an objective of many government priority programs in a knowledge-based and globalising economy (Leydesdorff & Heimeriks, 2001). The question is whether there is a 'smart specialisation' alternative to policies that spreads investments thinly across many topics of research, and, as a consequence, not making much of an impact in any one area (Todtling & Trippl, 2005). A more promising strategy appears to be to encourage investment in programs that will complement existing skills and infrastructures to create future capability and comparative advantage (Hausmann & Hidalgo, 2009).

However, few of the existing studies address the question whether the specific type of research activity undertaken matter? This question is important because there are clear policy implications of this issue in terms of policies directed towards innovation and knowledge development.

The purpose of this paper is to study organisation specific specialisation patterns by identifying the extent to which research activity is either concentrated, or alternatively consists of diverse but complementary research activities, and how this composition influences output. We use the body of codified knowledge accumulated in scientific publications during the period 1995–2009 as data for our analysis. Key topics are used as an indication of cognitive developments within the scientific fields for over a period of time.

Data and Methods

In a study of aggregated journal-journal citations it was argued that one can track fields by defining 'central tendency journals'(Leydesdorff & Cozzens, 1993). In this paper, we will use four 'central tendency' journals to map the development of the fields of Biotechnology, Nanotechnology and Organic Chemistry between 1995 and 2009.

These cases for our empirical operationalization of evolving knowledge dynamics were selected as representative of patterns in global knowledge production in the natural sciences. The selection includes the emerging sciences Biotechnology and Nanotechnology as well as the more traditional fields Astrophysics and Organic Chemistry that are included in the analysis for comparison (Table 1).

Table 1. Journals and number of publications between 1995 and 2009

Field	Journal	Number of Publications
Astrophysics	Astrophysical Journal	36572
Biotechnology	Biotechnology And Bioengineering	5423
Nanotechnology	Nanotechnology	7142
Organic Chemistry	Journal Of Organic Chemistry	23282

Astrophysics is a clear example of government supported "big science". Knowledge production requires massive and unique infrastructures like observatories and satellites, which makes government funding inevitable. Astrophysics is characterized by a high importance on collaborative research, a cumulative tradition, substantial governmental funding, and an extensive use of data and physical infrastructures (Heimeriks et al., 2008).

Biotechnology is characterized by an interdisciplinary knowledge development with emphasis on applications and a variety of producers and users of knowledge (Heimeriks & Leydesdorff, 2012). The combination of problem variety, instability, and multiple orderings of their importance with technical standardization occurs especially in this field (Whitley, 2000). The knowledge base has been characterized by turbulence, with some older topics becoming extinct or losing importance (related to food preservation and organic chemistry) and with some new ones emerging and becoming important components (Krafft, Quatraro, & Saviotti, 2011).

Like Biotechnology, Nanotechnology is an emerging technoscience characterized by high growth, high diversity, and large human capital and institutional complementarities that requires a very diverse instrument set (Bonaccorsi & Thoma, 2007). Nanotechnology is highly interdisciplinary and expected to have major economic and social impacts in the years ahead. Inventive activities in nanotechnology have risen substantially since the end of the 1990s and funding has increased dramatically.

Organic Chemistry knowledge development is expected to be highly cumulative as an example of a field that has high levels of 'mutual dependence', but lower levels of 'task uncertainty' (Whitley, 2000). Organic Chemistry is a long lasting field characterized by a low to medium growth, low diversity, and low complementarity search regime. Furthermore, it is generally acknowledged that chemistry has been evolving around bilateral cooperation at national level between the universities, research institutes and firms (Bonaccorsi, 2008).

All data from these fields as retrieved from the ISI Web of Science could be organized in a relational database as the basis for the organisational analysis. In addition to organisation names, the papers provide keywords. As such, the data allows us to study the rise and fall of these organisations in co-evolution with the changing topics of research. In this study we will report the developments of the 25 most important organisations and the 25 most important topics in each field.

In order to analyse this global-local dynamic of knowledge production we use Shift-Share analysis and entropy statistics. Shift-share analysis is a widely used analytical technique for retrospectively decomposing changes in a set of locations (Knudsen, 2000). Shift-Share analysis recognizes that some local research activities are likely to be growing at a faster rate compared to the total set of research activities under study and other locations will be growing more slowly. Localized research activities are expected to specialize in certain topics and produce output in that area of expertise, thereby increasing its output in that topic.

As such, Shift-Share analysis reveals how well the local activities are performing by systematically examining the total, organisational (local), and different components of change in knowledge production. A shift-share analysis will provide a dynamic account of total organisation growth that is attributable to growth of the total production, a mix of faster or slower than average growing activities, and the competitive nature of the local activities.

However, a Shift-Share analysis is purely descriptive and because it is based on the observations between two points in time, it cannot show any truly dynamic mechanism of specialisation or diversification of search. For this purpose, we use entropy statistics for further analysis of the co-evolutionary patterns of development between organisation and topics.

Results

The increasing number of publications and the rising number of contributing countries indicate an on-going globalization. However, the analyses presented here highlight the distinct knowledge dynamics in different fields. In dynamic (emerging) fields, with high growth rates (such as Biotechnology and Nanotechnology); entrance barriers are lower for new organisations to contribute.

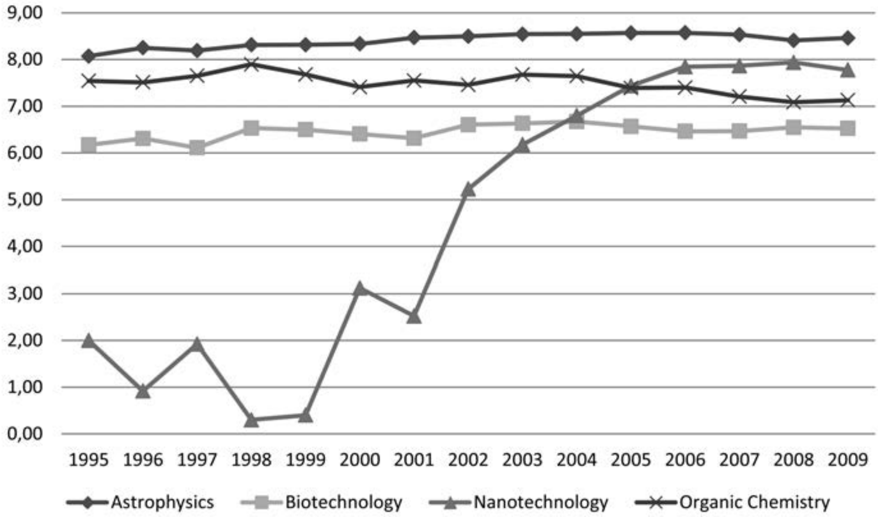
The Shift-Share analysis shows that the total growth shares are very different among fields. In Biotechnology and Organic Chemistry, total growth shares tend to be a lot smaller than in other fields. Furthermore, research organisations in these contribute only to a limited number of topics indicating a high level of required capabilities. Topics that require more capabilities will be accessible to fewer organisations.

In Astrophysics total growth rates tend to be higher than in other fields. Moreover, research organisations are diversified and contribute to many important topics in this field. In Astrophysics a substantial part of the growth of the field can be attributed to the most important research organisations. This is not the case in Biotechnology and Organic Chemistry.

These results can be further elaborated with entropy analysis. Using entropy statistics is a way to measure to degree of diversity and the mutual information between topics and research organisations and their co-evolutionary developments. Concerning the frequency distributions of topics

and organisations, the Shannon (H) values inform us that an overall increase in variety occurred in the period under study (Figure 1). This increase in diversity is the expected consequence of on-going globalization and the emergence of more topics of research.

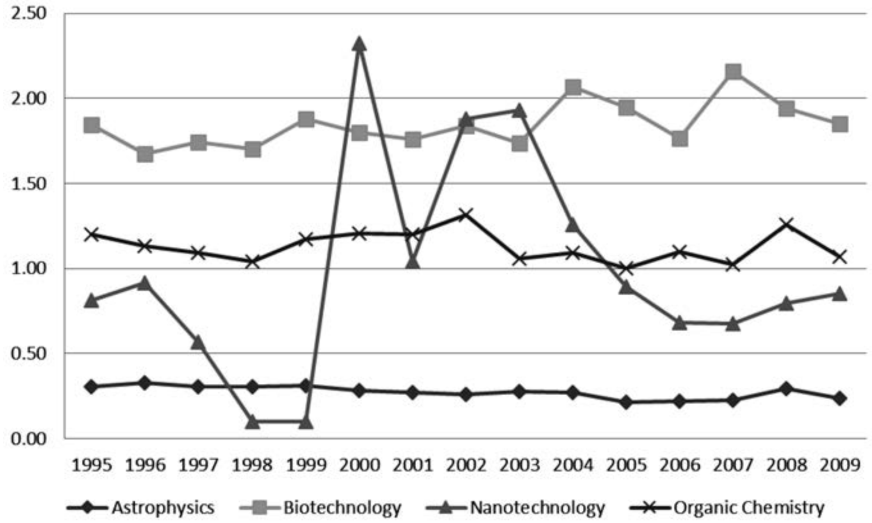
Figure 1. The two-dimensional entropy values (H) of research organizations and most important topics in the period under study.



However, the observed increasing variety does not translate into identical evolutionary patterns among the fields. The transmission values (T) inform us about the mutual information content between the organisations and topics. The transmission values show very different patterns among the fields under study (Figure 2). In Biotechnology the transmission values are relatively high and indicate a strong relationship between the research organisations and the topics under study. Astrophysics represents the other extreme; the mutual information between the two dimensions is very low compared to the other fields. The topics of research are relatively decoupled from specific research locations. In Nanotechnology, this process of de-coupling is still on-going, while in Organic Chemistry the pattern is in between Astrophysics and Biotechnology.

The transmission values between topics and organisations show a pronounced development in Nanotechnology. This development coincides with the surge in funding of nanotechnology when it became a priority funding area in most advanced nations in the period 2000–2003 (Leydesdorff and Schank 2008).

Figure 2. The transmission values between research organizations and most important topics in the period under study.



Consequently, the scope of opportunities for research organisations around the world to contribute within the constraints of the existing body of knowledge is different in each field. Biotechnology showed the highest level of local specialisation while Astrophysics provides a wide range of research topics for the most important organisations in the field. This is also the case for Nanotechnology in later years, although to a lesser extent.

Additionally, the Shift-Share results show that it is difficult, if not impossible, to maintain a local comparative advantage in Biotechnology and Organic Chemistry. In these fields, the average organisation shares are lower than the topic mix values averaged over the period under study. In more established fields with large infrastructural needs, such as Astrophysics, all important research organisations manage to achieve local comparative advantages over a longer period of time (table 2).

Table 2. Comparative advantages in different fields between 1995 and 2009

	number of organizations with a comparative advantage	number of topics contributing to comparative advantages	total number comparative advantages
Astrophysics	18	7	31
Biotechnology	2	2	2
Nanotechnology	9	11	19
Organic Chemistry	2	5	5

Discussion and Conclusion

The analyses showed that in all fields, path and place dependent processes of knowledge production can be identified. Shift-Share analysis reveals that organisations show a pattern of specialisation over time. We account for these specialisation patterns by assuming that each topic of research requires local capabilities, and that a research organisation can only contribute to topics for which it has all the requisite capabilities.

Topics (and fields in general) differ in the number and specific nature of the capabilities they require, as research organisations differ in the number and nature of capabilities they have. Topics that require more capabilities will be accessible to fewer organisations (as is the case in most topics in Biotechnology), while research organisations that have more capabilities (as is the case in Astrophysics) will have what is required to contribute to more topics (i.e., will be more diversified).

The patterns of research activities differ systematically across the scientific fields under study. However, these patterns are remarkably similar across organisations within each scientific field. Two patterns of specialisation are identified. The first represents a turbulent pattern: concentration of research activities is low, knowledge producing organisations are of small size in terms of output, stability in the ranking is low and comparative advantages are short lasting. The second represents a stable pattern: concentration of research activities is higher than in the first group, research organisations are of larger size, stability in the ranking is greater, and comparative advantages last longer. The former group comprises biotechnology and organic chemistry, while the latter includes astrophysics. Nanotechnology develops towards a stable pattern of knowledge production.

The results show that it is difficult, if not impossible, to maintain a local comparative advantage in Biotechnology and Organic Chemistry. In more established fields with large infrastructural needs, such as Astrophysics, all important research organisations manage to achieve local comparative advantages over a longer period of time.

For research and innovation policy this implies that while generic measures can sometimes be helpful, there is clear need for disaggregated measures that match the existing place dependencies on the organisational level as well as the global path dependent dynamics of specific fields. Smart specialisation strategies need to take into account the two dependencies that this study brought to the fore. First, in fields of knowledge production were research organisations have the capabilities to contribute to many topics, the number of new entrants tends to be low. The key policy is to identify the commonalties and infrastructures in these fields that allow for diversity in knowledge production. Reversely, in fields where organisations contribute to a small number of topics, there are more opportunities for new entrants to establish a niche for themselves. Policy should focus on developing a narrow set of research activities in order to yield greater innovative output

References

Bonaccorsi, A. (2008). Search Regimes and the Industrial Dynamics of Science. *Minerva*, 46, 285–315.

Bonaccorsi, A., & Thoma, G. (2007). Institutional complementarity and inventive performance in nano science and technology. *Research Policy*, 36(6), 813–831.

Cozzens, S. F., & Leydesdorff, L. (1993). Journal systems as macroindicators of structural changes in the sciences. In Van Raan H.F. Moed, A.J. Nederhof and R.W.J. Tijssen (Ed.), *Science and technology in a policy context* (219–233). Leiden: DSWO/Leiden University Press.

Foray, D. (2004). *The Economics of Knowledge*. Cambridge, MA/London: MIT Press.

Foray, D., David, P. A., & Hall, B. (2009). *Smart Specialisation. The concept. Knowledge Economists Policy Brief: Expert group on Knowledge for growth*.

Hausmann, R., & Hidalgo, C. A. (2009). The Building Blocks of Economic Complexity. *Proc. Natl. Acad. Sci.*, 106(26), 10570–10575.

Heimeriks, G, & Leydesdorff, L. (2012). Emerging Search Regimes: Measuring Co-evolutions among Research, Science, and Society. *Technology Analysis and Strategic Management*, 24(2).

Heimeriks, G., & Boschma, R. A. (2012). The path-and place-dependent nature of scientific knowledge production in biotech 1986–2008.

Heimeriks, G., Van den Besselaar, P., & Frenken, K. (2008). Digital Disciplinary Differences: An analysis of computer mediated science and "Mode 2" knowledge production. *Research Policy*, 37, 1602–1615.

Heimeriks, G., & Vasileiadou, E. (2008). Changes or transition? Analysing the use of ICTs in the sciences. *Social Science Information*, 47(1), 5–29.

Heimeriks, Gaston, & Leydesdorff, L. (2012). Emerging search regimes : measuring co-evolutions among research, science, and society. *Technology Analysis & Strategic Management*, 24(1, January 2012), 37–41.

Knudsen, D. C. (2000). Shift-share analysis: further examination of models for the descritption of economic change. *Socio-Economic Planning Sciences*, 34, 177–198.

Krafft, J., Quatraro, F., & Saviotti, P. (2011). The knowledge-base evolution in biotechnology: a social network analysis. *Economics of Innovation and New Technology*, 20(5), 445–475.

Leydesdorff, L. (1989). Words and Co-Words as Indicators of Intellectual Organization. *Research Policy*, (18), 209–223.

Leydesdorff, L. (2001). Indicators of Innovation in a Knowledge-based Economy. *Cybermetrics*, 5(1).

Leydesdorff, L., & Cozzens, S. (1993). The delineation of specialties in terms of journals using the dynamic journal set of the Science Citation Index. *Scientometrics*, (26), 133–154.

Leydesdorff, L., & Heimeriks, G. (2001). The Self-Organization of the European Information Society: The case of "Biotechnology. *Journal of the the American Society for Information Science and Technology*, 52(14), 1262–1274.

Leydesdorff, L., & Schank, T. (2008). Dynamic Animations of Journal Maps: Indicators of Structural Change and Interdisciplinary Developments. *Journal of the American Society for Information Science and Technology*, 59(11), 1810–1818.

Lucio-Arias, D., & Leydesdorff, L. (2009). An Indicator of Research Front Activity: Measuring Intellectual Organization as Uncertainty Reduction in Document Sets. *Journal of the American Society for Information Science & Technology*, 60(12), 2488–2498.

Nelson, R. (1994). The Co-evolution of Technology, Industrial Structure, and Supporting Institutions. *Industrial and Corporate Change*, 3(1), 47–63.

OECD. (1996). *The Knowledge Based Economy*. Paris: OECD.

Saviotti, P. p., & Mani, G. s. (1995). Competition, variety and technological evolution: a replicator dynamics model. *journal of evolutionary economics*, 5, 369–392.

Todtling, F., & Trippl, M. (2005). One size fits all?: Towards a differentiated regional innovation policy approach. *Research policy*, 34(8), 1203–1219.

UNESCO. (2010). UNESCO Science Report 2010. The Current Status of Science around the World. Paris: UNESCO.

Whitley, R. (2000). *The intellectual and Social Organization of the Sciences* (2nd ed.). Oxford: Oxford University Press.

Proximity mechanisms in European research networks: a case study of the water sector

Pieter Heringa p.heringa@rathenau.nl / Science System Assessment, Rathenau Instituut, PO Box 95366, 2509 CJ Den Haag (The Netherlands)

Laurens Hessels l.hessels@rathenau.nl / Science System Assessment, Rathenau Instituut, PO Box 95366, 2509 CJ Den Haag (The Netherlands)

Marielle van Zouwen Marielle.van.der.Zouwen@kwrwater.nl / KWR Watercycle Research Institute, Groningenhaven 7, 3433 PE Nieuwegein (The Netherlands)

Introduction

Transnational research policies are an increasingly powerful factor in European science systems. In the EU, since about 1980 the Framework Programmes (FPs) are an important instrument for the support of pre-competitive R&D with specific attention for interactions between academic researchers and their stakeholders. FPs aim to foster research networks across different organizational sectors, and across national borders. Promoting trans-sectoral and transborder research collaboration will remain important policy aims in the Horizon 2020 initiative for the current decade. Also the common vision for the European Research Area, adopted by the Council of the European Union, emphasizes the need for reinforced cooperation and coordination. Previous studies have shed light on the influence of European research policies have on international research collaborations. Although researchers increasingly collaborate with colleagues abroad, they still prefer to work with colleagues located nearby rather than those far away (Frenken et al., 2007).

Most studies on international research collaborations focus on scientific research networks, with a relatively homogeneous composition. Heterogeneous research networks, including firms, governmental bodies and research organizations have been studied relatively little. Another limitation of current literature is the focus on joint patents and co-publications as indicators of research collaborations. Given the nature of the European FP's , not all funded projects will generate such output. The current paper aims to contribute to the understanding of the dynamics of international research networks by a case study of European research collaborations in the water sector. Given the transnational societal challenges in combination with a (knowledge) infrastructures typically organized in national systems, water seems an interesting case for studying the influence of transnational research policies.

Theoretical framework

Various studies suggest that proximity is a key concept for understanding the configurations of collaboration in knowledge production (Boschma, 2005). The basic premise is that proximate

people have a tendency to collaborate, as it is easier to communicate with people who are close. On the other hand, the advantage of collaboration may disappear when people become "too close". There is a substantial body of work on the relation between geography and innovation (Autant-Bernard et al., 2007). Gravity models show that geographic proximity can explain co-authorship in scientific publications (Hoekman et al., 2010). Ethnographic studies, for example on business development and technology acquisition around CERN, show the importance of cognitive and social proximity for successful collaboration (Autio et al., 2004).

The current paper builds on the existing proximity literature in order to understand the dynamics of research networks. Two aspects of our study are relatively novel in the proximity literature: we use joint project participation as an indicator of research collaborations and we analyze heterogeneous networks including knowledge institutions, firms and governmental organizations.

We include four dimensions of proximity in our analysis, each describing a different aspect of the relationship between the collaborating actors:

- Geographical proximity: physical distance
- Social proximity: social embedment or trust
- Organizational proximity: similarity in incentives and routines
- Cognitive proximity: similarity in the professional knowledge base

Main question of this paper is what dimensions of proximity most strongly influence international research collaborations in the water sector.

Indicators and dataset

We use data on projects in Framework Programme 1 to 7 (last updated March 2010). Our dataset has been extracted from the EUPRO database (Barber et al., 2008), which is a cleaned and standardized version of the public CORDIS database of the European Commission. CORDIS contains detailed information about all funded projects, such as project title, project participants, location and organization type of participants, budget, start date, end date, and keywords. EUPRO includes standardized geographical information on each participant, including the NUTS2 region, and a consistent classification of 7 types of organizations: Industry, University and other educational institutions, Publicly funded research organizations, Governmental institutions, Consultants, Non-commercial and non-profit organizations, and Others.

Based on literature reviews and expert advice we have produced a set of keywords reflecting the water sector, defined as all human activities around the water cycle (production/purification and transport of drinking water, collection, transport and treatment of wastewater, water storage, water use and water management, etc.). Using this set of keywords, we have selected 5112 projects from the database for our analysis.

We operationalize the proximity dimensions as follows:

– geographical proximity: using NUTS2 classification of European regions we calculate the distance between two project participants in kilometres
– social proximity: how many FP-projects have the partners jointly participated in before?
– organizational proximity: are the partner organizations of a same type of organization?
– cognitive proximity: what is the similarity in keywords of previous FP-projects both organizations have participated in?

We analyze the extent to which each dimension of proximity can predict the probability of research collaborations. We will carry out these analysis on the complete dataset, and on subsets of projects funded by particular funding instruments, such as People, Capacities and Cooperation, in order to investigate the possible influence of funding conditions on the choice of collaboration partners.

Expected results

We will present our results in the form of curves depicting the relationships between each proximity dimension and the probability of research collaborations. This will yield an overview of the relative influence of each proximity dimension on research collaborations. By conducting these analyses also on subsets funded by different funding instruments within FP we will be able to characterize the collaboration dynamics of the different funding instruments.

Our paper will close with a reflection on the usability of our indicators of proximity and of the possibilities to analyse research collaboration based on FP participation data. In addition, we will explore the policy implications of our empirical findings.

References

Autant-Bernard, C., P. Billand, D. Franchisse, and N. Massard. 2007. Social distance versus spatial distance in R&D cooperation: empirical evidence from European collaboration choices in micro and nanotechnologies. *Papers in Regional Science* 86:495–520.

Autio, E., A.-P. Hameri, and O. Vuola. 2004. A framework of industrial knowledge spillovers in big-science centers. *Research Policy* 33:107–126.

Barber, M., B. Heller-Schuh, T. Roediger-Schluga, and T. Scherngell. 2008. The sysres EUPRO database manual. NEMO Deliverable D4.1 Vienna.

Boschma, R. 2005. Proximity and innovation: a critical assessment. *Regional Studies* 39:61–74.

Frenken, K., J. Hoekman, and F. van Oort. 2007. Towards a European Research Area NAi publishers, Rotterdam.

Hoekman, J., K. Frenken, and R.J.W. Tijssen. 2010. Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy* 39:662–673.

Enlightenment literature and the complex ecology of public knowledge

Diana Hicks* dhicks@gatech.edu / School of Public Policy, Georgia Institute of Technology, Atlanta, GA, 30332–0345 (United States)

Jian Wang jianwang@gatech.edu / Institute for Research Information and Quality Assurance (iFQ), Berlin, Germany / School of Public Policy, Georgia Institute of Technology, Atlanta, GA, 30332–0345 (United States)

* Corresponding author

Abstract

A great deal of attention has been paid to referencing within the academic literature in the bibliometric, indicators and science studies communities. As attention increasingly turns to the societal benefits of research, the relationship between knowledge created inside and outside the academic community needs to be explored to enable understanding of how scholarship can influence society and vice versa.

We begin this work by examining the non-academic source most cited in the Web of Science, the *New York Times*. The *New York Times* is referenced surprisingly often in the scholarly literature. 1.25% of US papers published in 2010 referenced the NYT, accounting for 15,000 references. The *New York Times* receives more citations from academic journals than the *American Sociological Review*, *Research Policy*, and *Harvard Law Review*. NYT referencing has grown faster than number of papers since 1980, with growth accelerating after 2006. Papers in all fields reference NYT articles, though law exhibits the highest rate of referencing, followed by social sciences, medical sciences then science and engineering fields.

Although NYT referencing is widespread, academics have not become journalists and journalists have not become academics. Examining the contexts of NYT references, we found authors most often leveraging the differences between the scholarly and newspaper genres to advance their arguments. Sometimes the NYT is referenced because the NYT is the subject of the research. Sometimes New York City is the subject, which leads to extensive use of the NYT as a source. More often, authors seek to establish the importance of their topic by using press coverage as evidence of public concern. The *New York Times* is also used as a primary source, for example in discussing a specific event or inserting words spoken by an influential person. Perhaps half of NYT referencing uses the NYT as a primary source.

The traditional image of a homogenous entity called science with a fragile sense of power and integrity defending rigid boundaries serves us less well in framing scholarly engagement with the NYT than do conceptions of a heterogeneous, changing enterprise. The idea of scholarship as a

heterogeneous evolving entity is reminiscent of Mode 2. Homogeneous science with a rigid boundary evokes Mode 1 associations while increasing engagement with the NYT aligns with elements of Mode 2.

We argue that referencing of the *New York Times* makes sense within a framework that recognizes the heterogeneity of scholarship. Especially in the social sciences, scholars work within multiple genres, including the international journal article, but also books, national journals and enlightenment literature. Taking seriously the enlightenment literature highlights the existence of closed, disciplinary journals that contrast with open, porous journals participating in the complex ecology of public knowledge. Such openness indicates the flexibility needed to interact successfully with societal debates in the hopes of influencing them in a positive direction. However, we do not know which approach is more influential in society: the closed, autonomous and high academic impact work or the open, porous, enlightenment approach.

Bibliometric classification of emerging topics

Marcus John Marcus.John@int.fraunhofer.de / Fraunhofer Institute for Technological Trend Analysis INT, Appelsgarten 2, 53879 Euskirchen (Germany)

Frank Fritsche Frank.Fritsche@int.fraunhofer.de / Fraunhofer Institute for Technological Trend Analysis INT, Appelsgarten 2, 53879 Euskirchen (Germany)

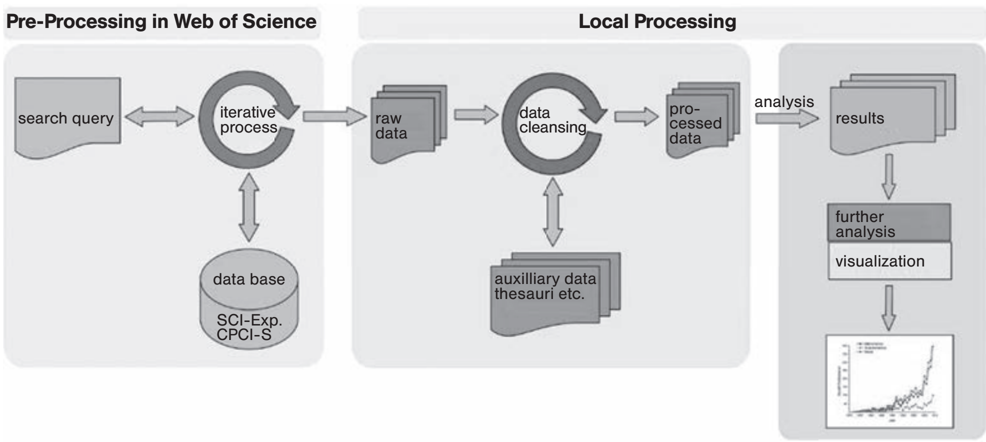
Introduction

The early and correct detection of emerging topics is a challenge in technology forecasting. In recent years there has been a growing number of publications which examine the applicability of different bibliometric methods to this problem (see e.g. Chen, 2006, Shibata, Kajikawa, Takeda, & Matsushima, 2008, Schiebel, Hörlesberger, Roche, François, & Besagni, 2010). Since bibliometrics is an inherently retrospective method, it remains an open question whether it is possible to detect trends for the future by lurking into scientific communication. This contribution tries to clarify this question by examining historic scientific trends and looking for specific patterns, in order to distinguish between different types of emerging topics. It will be demonstrated that it is indeed possible to distinguish between different kinds of emerging topics by means of bibliometric quantities. Furthermore, it will be discussed whether, and if so how it is possible to establish a scheme for the classification of emerging topics using these quantities. The bibliometric results will be accompanied by some comments regarding the scientific contents of the considered themes.

Method

To this end we established a bibliometric workflow, which in principle consists of three different phases as depicted in Figure 1. During the first phase a suitable search query is iteratively constructed, which tries to delineate the scientific research field of interest as accurately as possible.

Figure 1. Schematic of the bibliometric workflow.



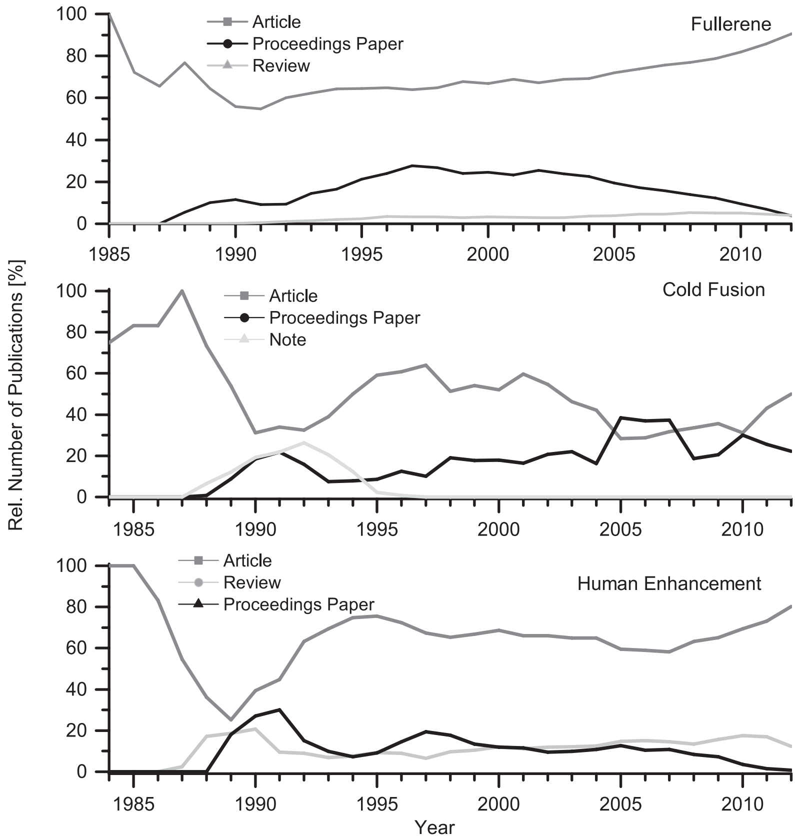
Subsequently the bibliometric data set obtained with this search query is then processed, revised, and prepared for further analysis.

For each considered research field the process of scientific communication is examined with regard to unveiling the existence of specific patterns, which allow to distinguish between different types of emerging topics. To this end, a number of bibliometric observables and their time dependence is analysed, e.g.:

(1) The publication dynamics, i.e. the number of papers per year.
(2) The time dependence of the document types (see Figure 2).
(3) The time-dependent size of the giant component of the co-author network (Bettencourt, Kaiser, & Kaur, 2009)

The method described above has been applied to different scientific themes. Exemplarily, the research on fullerenes, which can be traced back to the paper of Kroto, Heath, O'Brien, Curl, & Smalley (1985), describing the discovery and initiating a whole new research field, will be discussed. Another example is the research on cold fusion (Fleischmann & Pons, 1989), which is considered as a prototypical example of pathological science. A third example is the rather new field of human enhancement (see e.g. Eckhardt, Bachmann, Marti, Rütsche, & Telser, 2011).

Figure 2. Time dependent development of the different document types. The data are normalized to the number of publications in a specific year. For better readability only the smoothed data are presented, using the moving average (window length is 3 years).



Results

The exemplarily depicted time dependence of the scientific communication channels, viz. the document types used for the publication (see Figure 2), clearly demonstrates that each of the three different themes has established its own pattern. In the case of fullerene research the pattern comes up to one's expectations: The communication starts with a high amount of articles. After some time conferences become an important communication channel for the researchers and only after some years, the first review papers appear. The latter reach their maximum one year after the Nobel prize in Chemistry was awarded to Curl, Kroto and Smalley in 1996. Contrarily, the communication in the field of human enhancement relies on review papers already in an early stage. Furthermore, the communication about the science of cold fusion is at least partly published as notes.

The results for the analysis of the giant component of the co-author network can be distinguished in a similar way. In case of the research on fullerenes this network displays a high connectivity after a short time already. Contrarily, such a high connectivity is never reached in case of the research on cold fusion. Furthermore, this process of densification is still taking place in case of the last theme, namely human enhancement.

Conclusion

These results suggest that the following types of emerging topics can be distinguished:

(1) The generic case of an emerging topic, which can be traced backed to a single publication, initiating the research in this field. An example for this case is the research on fullerenes.
(2) The aggregating case of an emerging topic, exemplarily observed in the case of human enhancement. In this case information and techniques from various possibly disparate fields of science aggregate to a completely new field.
(3) The case of pathological or failed science, as seen in the case of cold fusion.

The applicability of techniques for the detection of emerging topics might differ for each of these three types.

References

Bettencourt, L. M. A., Kaiser, D. I., & Kaur, J. (2009). Scientific discovery and topological transitions in collaboration networks. *Journal of Informetrics*, 3(3), 210–221.

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359–377.

Eckhardt, A., Bachmann, A., Marti, M., Rütsche, B., & Telser, . (2011). *Human enhancement. TA-Swiss*: 56/2011. Zürich: vdf Hochschulverlag.

Fleischmann, M., & Pons, S. (1989). Electrochemically induced nuclear fusion of deuterium. *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, 261(2, Part 1), 301–308.

Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F., & Smalley, R. E. (1985). C60: Buckminsterfullerene. *Nature*, 318, 162–163.

Schiebel, E., Hörlesberger, M., Roche, I., François, C., & Besagni, D. (2010). An advanced diffusion model to identify emergent research issues: the case of optoelectronic devices. *Scientometrics*, 83(3), 765–781.

Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2008). Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation*, 28(11), 758–775.

Towards the Development of an Article-level Indicator of Conformity, Innovation and Deviation

Richard Klavans rklavans@mapofscience.com / SciTech Strategies, Inc., Berwyn, PA 19312 (USA)

Kevin W. Boyack kboyack@mapofscience.com / SciTech Strategies, Inc., Albuquerque, NM 87122 (USA)

Abstract

An article-level indicator of conformity, innovation, and deviation is generated by comparing an article's belief system to socio-cognitive norms. Articles that reinforce a highly related set of norms are considered conforming. Articles that reinforce a diverse set of norms are considered innovative. Articles that do not follow normative behavior exhibit a high level of deviation. Data on 241,450 articles are used to create a set of four categories related to conformity and innovation – *uniformity*, *conformity*, *innovation* and *deviation*. Assignment of articles to these categories, and comparison to citation rates, shows that uniformity and conformity result in lower citation rates, while innovation and deviation result in higher citation rates. While a higher than average fraction of the most highly cited articles end up in the innovation and deviation categories, a significant fraction of these highly cited articles (17%) are nevertheless found to be conforming. These results challenge the prevailing belief that all highly cited articles are innovative.

Introduction

The development of an article level indicator of innovativeness is critical if institutions and nations are intending to pursue an innovation strategy. Such indicators are needed both for planning (assisting in the selection of innovative proposals) and evaluation (determining if the institution has taken appropriate risks or is primarily funding safe, i.e., conforming, research). This is a central question for science policy as evidence by a recent commentary in *Nature* where the authors claim that the NIH has a "conform and be funded" profile (Nicholson & Ioannidis, 2012). Every funding institution is sensitive to this issue, and tries to create policies that will attract innovative research proposals.

The current state of the art in the development of article-level indicators of innovation is woefully inadequate. The *Nature* article mentioned above focused on the highest impact US-authored article in biomedical research and noted that the NIH did not fund a significant percentage of these papers. However, the authors did not measure innovation. Rather, they assumed that articles that are highly cited (i.e., high impact) are also highly innovative. This is a common, and perhaps mistaken, assumption. Many highly innovative papers may not be high impact papers, and many high impact papers may not be innovative. Confusing high impact with a high level of innovation can easily lead to policies that reduce innovative output. For example, it may be easier

to continue funding a stream of research that has been producing highly cited articles than to fund new innovative work with a higher potential for breakthrough. This concern, that agencies may be over-funding research that is conforming at the expense of the innovative, is widespread among researchers. A valid article-level indicator that can distinguish between conforming and innovative articles is therefore needed.

This study reports on work done to develop an article-level indicator of conformity and innovation. We start by reviewing a theory that describes how one might construct such an indicator. This is followed by a description of the data and methods used to develop the indicator. A highly detailed structural model and map of science comprised of nearly 20 million articles from 1996–2011 is used to establish socio-cognitive norms for all of science. References and their positions in full text articles are then used to describe the belief system upon which each article is based using distances on the map between the article and co-cited pairs of references. Modes in the distance data suggest that there are several socio-cognitive norm types: uniformity (extreme agreement with a single socio-cognitive norm), conformity (agreement with local socio-cognitive norms), innovation (partial agreement with a wider group of socio-cognitive norms) and deviation (little agreement with socio-cognitive norms). Articles are then characterized by these norms, and citation outcomes are examined for the resulting four groups of articles.

Theoretical Basis

The proposed indicator builds upon the computational theory of innovation by (Chen et al., 2009). This theory suggests that an article is embedded in a network of concepts, and that the most innovative articles are those associated with structural holes – i.e., those articles that bridge multiple existing clusters of documents rather than reinforce a single cluster. The development of an indicator based on this theory requires three tasks. First, socio-cognitive norms – the entire network of all related concepts – must be established. Second, the sets of concepts (or the belief systems) upon which a corpus of articles builds must be identified. Third, these sets of concepts must be analysed to see if any patterns exist that are conducive to be used as an indicator.

Chen's article provides an example. Co-citation analysis is used to generate a model of the socio-cognitive structure of a field, creating clusters of co-cited references. These clusters of references represent the socio-cognitive norms upon which a new article can build. For each new article, one can look at the actual pairs of references and see where they are located. If the pair of references is in the same cluster, it signals that the article is conforming. If the pair of references is in different clusters, it signals that the paper is innovative. Each new article contains many pairs of references, and thus many signals. Combining all of these signals from an article tells us whether it tends to be conforming or innovative.

Based on Chen's theory, we hypothesize that if the belief system of an article is conforming, it follows socio-cognitive norms, and that these norms will tend to be densely located in a local area

of the network. By contrast, if the belief system is innovative, the norms will tend to be more widely dispersed in the network. The innovative article will link disparate elements of the network.

Approach

The proposed indicator is based on a map of science and technology created using Scopus and US patent data from 1996–2011. In this study, we only use the article clusters, and not the patent clusters. A total of 19,012,183 articles were assigned to 149,613 clusters using the direct citation method recently described by Waltman & van Eck (2012) and the direct citation similarity from Boyack & Klavans (2010). The article clusters, represented as colored nodes in Figure 1, were located on a 2-D map using a force-directed algorithm (Martin, Brown, Klavans, & Boyack, 2011) where cluster:cluster relatedness was based on words from titles and abstracts using the BM25 text similarity (Boyack et al., 2011).

Figure 1. Map of the socio-cognitive structure of science and technology.

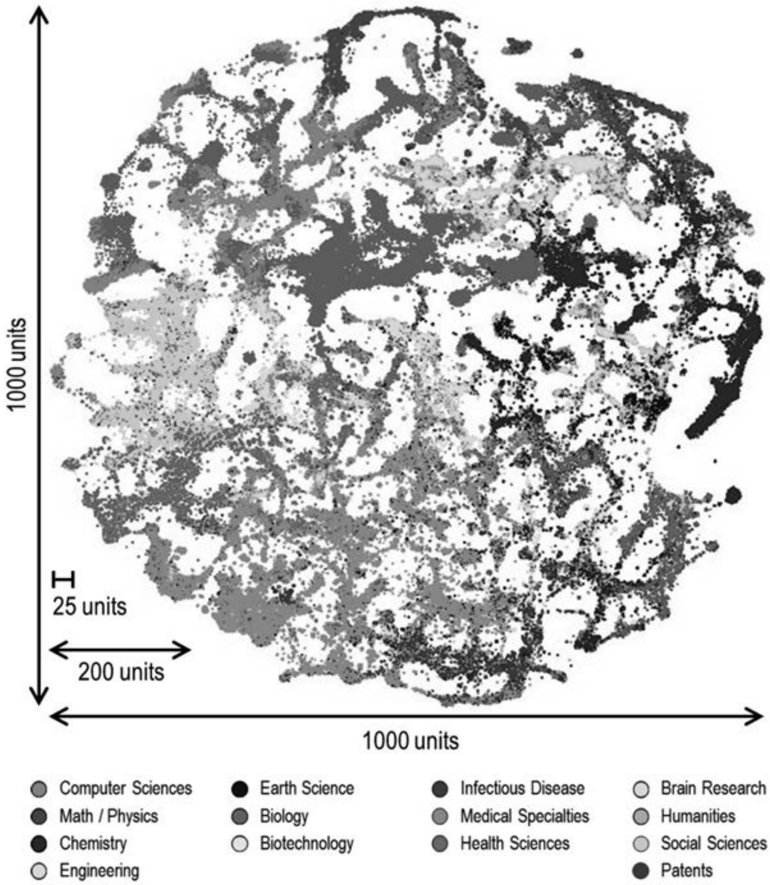


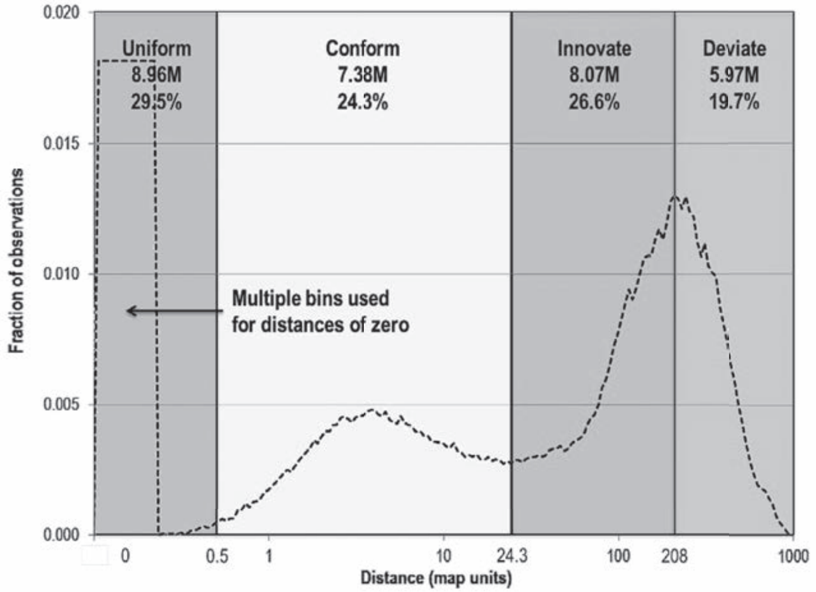
Figure 1 is a visual description of the socio-cognitive landscape of science. The map has been scaled such that the ranges in both axis directions are 1000 units. Each cluster has an x,y position in this space. One can imagine that referencing multiple clusters (or belief systems) that are very close to each other on the map is actually a form of conformity. Referencing of clusters that are more distant from each other is less conforming, and thus more innovative.

The second step mentioned above is to identify the sets of concepts (or the belief systems) upon which a corpus of articles builds, and to measure how similar they are using map distances. As mentioned in the theory section, one tradition is to focus on co-citation pairs. We refine this traditional technique in two ways. First, we chose to use triplets – the citing article coupled with pairs of co-cited references – rather than simply using co-cited pairs of references. Second, rather than using all possible pairs, we use only those co-cited pairs of references that are proximate to each other in the full text. It has been shown that proximate pairs of references are more similar than pairs that are further apart in the text (Liu & Chen, 2012). We use only those pairs of references that appear within 375 characters (an average sentence length) of each other in the text. For each triplet, all three articles are located on the map shown in Figure 1. For each triplet, distances between each pair of articles are then calculated.

We started with a sample of 241,450 articles (all full text articles published by Elsevier in 2007 containing at least two references that could be located on the map). These articles contained 21,370,337 triplets based on proximate references. Of these, nearly half, or 10,126,221 triplets, could be located on the map shown in Figure 1. The remaining triplets contained references to books, journals not included by Scopus as a citing journal, and articles published before 1996 and thus could not be located. Distances were calculated for each of the three pairs of articles in each triplet, resulting in a total of 30,378,663 individual distances.

The distribution of distances, using nearly 200 bins based on $log_n(distance)$ is shown in Figure 2. 8.87 million (29.2%) of the 30.38 million pairs were located in the same cluster (a distance of zero). These are signals of the strongest kind of conformity, which we call *uniformity*. The remainder of the distribution is clearly bimodal. Many of the distances between articles are relatively short (within 24.3 units on our map with the 1000 unit scale). These 7.38 million pairs are considered signals of *conformity* because they are relatively close to each other on the map, but not uniformly so. The right-most peak in the distribution represents pairs that are more distant from each other, thus linking more disparate topics in an innovative way. We have chosen to split this into two groups which can both be classified as innovative. While the first group is *innovative*, the second group (which starts at a distance of 208) *deviates* much more strongly from the norm.

Figure 2. Distribution of distances (in map units) between pairs of articles in triplets containing one citing and two co-cited articles.



Given the groupings shown in Figure 2 (*uniform*, *conform*, *innovate*, *deviate*), the degree of innovativeness of each article can be calculated by taking the average distance for all article pairs associated with its triplets. The natural log of each distance is used in the calculation to avoid overweighting by large distances. The thresholds described above (D = *ln*[0.5; 24.3; 208]) are then used to characterize articles into 4 categories. The first category is used to represent *uniformity* – almost all triplets and pairs are in the same article cluster. The second category is *conformity* – most links are between very proximate clusters. *Innovation* only occurs when the average distance across all triplets is greater than 25. *Deviation*, the stronger form of innovation, contains articles whose triplets have an average distance of 208 or more.

Innovation and Citation Impact

Using the distances and the approach from above, the 241,450 articles in this study were placed in the categories shown in Figure 2. Numbers of articles in each category are listed in Table 1. To test our hypothesis that some high impact papers are not innovative, we obtained citation counts for each article (as of end-2011), and also identified each citing article and its location in the map. We then calculated distances between the each of the 241k articles in this study and each of its citing articles. From this, we assigned each individual citation to a category (*uniformity*, *conformity*, *innovation*, *deviation*) based on its distance. This allows us to know the degree to which an article is recognized by the different conformity or innovation categories.

Table 1. Citation counts by conformity / innovation category for articles published in 2007.

Type	#Art	#C-Uni	#C-Conf	#C-Inn	#C-Dev	#C-Tot
Uniformity	33,477	4.11	1.17	0.71	0.49	6.49
Conformity	53,079	5.10	4.02	1.70	0.78	11.59
Innovation	131,241	3.90	3.24	4.87	2.82	14.83
Deviation	23,653	2.61	1.03	3.32	8.94	15.90
ALL	241,450	4.07	2.91	3.44	2.65	13.06

Several key observations can be made from Table 1.

- Articles in the uniformity category receive fewer citations overall than those in any other category.
- Articles in the uniformity and conformity categories receive the majority of their citations from those same two categories. Conforming papers will tend to be cited by other articles in the same local area of the map.
- Conforming articles receive more conforming citations than do innovative articles.
- Articles in the innovation and deviation groups receive more citations overall than those in the conforming categories. In the aggregate, innovative articles have higher impact than conforming articles.
- Articles in the innovation and deviation categories receive the majority of their citations from those same two categories. Innovative papers will tend to be cited by articles outside their local area of the map.

We also examined a smaller set of articles – the 1,318 articles that received at least 100 citations by the end of 2011 – to answer our question about highly cited articles and innovation. Table 2 shows that one sixth of these highly cited papers are, in fact, conforming rather than innovative. Although this is only half of the expected value (from all articles), it does show that many high impact papers are falsely assumed to be innovative. While high impact papers are likely to be highly innovative, there are many false positives and false negatives to consider.

Table 2. Fractions of articles and highly cited articles by type.

Type	#Art	%	#Art-HC	%
Uniformity	33,477	13.9%	23	1.8%
Conformity	53,079	22.0%	205	15.6%
Innovation	131,241	54.4%	872	66.2%
Deviation	23,653	9.8%	218	16.5%
ALL	241,450		1,318	

Discussion

We have presented a theory-based article-level indicator of innovativeness that relies on a highly detailed model and map of science, and upon identification of the socio-cognitive belief systems embedded in individual article. One advance in this methodology is the use of triplets (a citing article and a pair of co-cited references that are proximate in the full text of the article). This departs from current practice of only using pairs (i.e. co-citation analysis or co-word analysis). We suggest that the citing paper needs to be considered when evaluating innovativeness.

The thresholds used to determine whether an article is conforming or innovative are based on a close examination of the distributional characteristics of a very large number of triplets from an extensive corpus that covers all of science. Figure 2 shows that the distribution of topic similarity within triplets is tri-modal, with peaks at zero, 4 map units and 208 map units. The peaks and troughs of this distribution were used to create four categories of conformity and innovation. The thresholds used for this indicator are based on the science map in Figure 1. However, we hypothesize that distances based on any global science map, whether created using text, citation, or hybrid methods, will show a similar tri-modal distribution of distances. We also note that this indicator requires use of a global map, and cannot be created from a local (or single field) map.

The proposed indicator of innovativeness predicts future citation rates. Conforming articles (including the uniform category) receive citations from other conforming articles, but receive little attention in the form of citations from the broader environment. Articles in the innovation and deviation categories see the opposite – they get fewer citations from conforming articles, but much more attention from the broader environment. Overall, as the level of innovation increases, as measured by map distances, impact also increases.

We note that the results of this study both agree and conflict with recent results from Roth et al. (2012). Our result showing that conforming papers receive more citations from articles that are close in the map is similar to Roth's result showing that citation counts are highest for the most normal and disciplinary (i.e., conforming) science. However, our overall results show that non-conforming (innovative) papers are more highly cited than conforming papers. The fact that Roth did not consider citations from outside the field may account for this difference. More study is needed on the subject.

There are a number of shortcomings in the proposed indicator that still need to be addressed – specifically field effects, replication and validation. This study was based on a broad data set, and ignores field effects, which we plan to study in the future. Regarding replication, we suggest that similar experiments be run on global maps created using alternative mapping approaches. This study used a hybrid map, combining direct citation clusters with a text-based layout. One could use another hybrid (co-citation and text) or rely solely on a text-based approach (such as topic modeling). Replication studies could also increase the number of references in the map. Approximately half of the triplets we identified had at least one reference that could not be located. Unas-

signed high-frequency references (e.g., those from before 1996) could be included in future replication studies.

We are especially concerned about methods for validating the indicators. While a few case studies are useful, large scale methods for validating these types of indicators are needed. It is not sufficient to note that highly innovative papers have more impact. It is more important to determine if the innovative effect associated with high impact papers actually occurs. Innovative papers are expected to change the flow of science by combining and redirecting current flow (i.e., Chen's description of 'turning points') or by seeding the emergence of a new area of study. Innovative papers are expected to disrupt the status quo by redefining the problem that researchers are working on.

This area of research is rich in methodological possibilities for validation, many of which are relatively easy to pursue. We emphasize that the more difficult paths, the ones that involve validation, are of greater importance to scientific progress.

References

Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61(12), 2389–2404.

Boyack, K. W., Newman, D., Duhon, R. J., Klavans, R., Patek, M., Biberstine, J. R., et al. (2011). Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches. *PLoS One*, 6(3), e18029.

Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pelligrino, D. (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics*, 3, 191–209.

Liu, S., & Chen, C. (2012). The proximity of co-citation. *Scientometrics*, 91(2), 495–511.

Martin, S., Brown, W. M., Klavans, R., & Boyack, K. W. (2011). OpenOrd: An open-source toolbox for large graph layout. *Proceedings of SPIE – The International Society for Optical Engineering*, 7868, 786806.

Nicholson, J. M., & Ioannidis, J. P. A. (2012). Conform and be funded. *Nature*, 492, 34–36.

Roth, C., Wu, J., & Lozano, S. (2012). Assessing impact and quality from local dynamics of citation networks. *Journal of Informetrics*, 6, 111–120.

Waltman, L., & Van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, 63(12), 2378–2392.

Measuring the quality of PhD theses

Svein Kyvik svein.kyvik@nifu.no / NIFU (Nordic Institute for Studies in Innovation, Research and Education), Wergelandsveien 7, 0167 Oslo (Norway)

Taran Thune taran.thune@nifu.no / NIFU (Nordic Institute for Studies in Innovation, Research and Education), Wergelandsveien 7, 0167 Oslo (Norway)

Abstract

This paper presents a survey methodology for the assessment of the quality of doctoral theses. Foreign members of evaluation committees in Norway in 2010 were asked to give their opinion on: (a) Originality, (b) depth and coverage, (c) theoretical level, (d) methodological level, (e) skills in written presentation, (f) contribution to the advancement of the field, and (g) external relevance. The assessment of the various quality elements is very coherent; less than 20% regarded the various elements as 'excellent', slightly more than 40% as 'very good', and 25–30% as 'good'. Preliminary results of a regression analysis indicate that North-Americans give Norwegian PhD-theses better rating than do their European peers. Examiners' acquaintance with supervisors has no effect on the assessment of five of the quality elements. Experienced examiners give consistently better rating than their less experienced peers. Academic field explains very little of the variation. Finally, theses from traditional research universities get significantly higher rating than theses at institutions that recently have been upgraded to university status with regard to 'originality', depth and coverage', 'theoretical level', and 'methodological level'. The paper discusses the possibilities for applying this methodology to assess the quality of PhD theses in other national or institutional contexts.

Purpose of the study

Very little has been written on what constitutes the quality of doctoral theses, neither in policy documents, nor in the scholarly literature on doctoral training. Policy documents are primarily occupied with learning outcomes of doctoral education, and to the extent they discuss the dissertation, this is done in general terms. Two documents can serve as an example. The European University Association has formulated some principles for doctoral education which state that the outcome of research training must 'testify to the originality of the research and be suitable for dissemination within the scientific community' (EUA 2005). The Qualifications Framework of the European Higher Education Area state that doctoral candidates should 'have made a contribution through original research that extends the frontier of knowledge by developing a substantial body of work, some of which merits national or international refereed publication.' In the scholarly literature, the issue of quality of doctoral theses has primarily been studied by reviewing written reports by evaluation committees (see e.g. Johnston 1997, Mullins & Kiley 2002).

The purpose of this paper is to present a survey methodology for the assessment of the quality of doctoral theses. In a recent evaluation of Norwegian PhD education (Thune et al. 2012), foreign members of evaluation committees were asked about their opinion, and this methodology should be of interest also to other researchers interested in quality assessment of research training.

Theoretical and conceptual frameworks

There is no common understanding of the notion of scientific quality. Nevertheless, there is some scattered empirical evidence based upon interviews and surveys among scientists that several different aspects of what constitutes scientific quality in research work might be agreed upon. Hemlin (1993) has distinguished between different attributes of quality; like correctness, stringency, novelty, depth, breadth, intra-scientific relevance and extra-scientific relevance, and various aspects of the research, like research problem, method, theory, results, reasoning, and writing style. The elements of this conceptual framework were largely supported in a questionnaire study among academics. The results of these studies across all fields including the humanities indicated that originality and correct methods were regarded as the highest ranked elements of scientific quality. Based on a literature review and interviews with scientists, Gulbrandsen (2000) argues that the concept of research quality should be decomposed into four quality elements that describe different criteria of good research; *originality*, *solidity*, *scholarly / scientific relevance*, and *practical / societal utility*. This conceptual scheme has been used as a basis for our study on the quality of PhD theses.

Data and methods

In order to ensure the quality of doctoral theses, the standard regulations pertaining to Norwegian doctoral training determine that, where possible, at least one member of the evaluation committee should come from a place of learning abroad. This is also the normal practice in most faculties. One method of assessing the quality of the theses thus is to ask the foreign members of the committees for their opinion. Of the members who took part in the evaluation of PhD theses in 2010, in total 1,159 responded to the survey, giving a response rate of 79%.

The foreign members came from 45 different countries, of which the 15 most frequent contributors to the PhD committees were Sweden (303), UK (164), Denmark (133), USA (125), Germany (74), Netherlands (46), France (44), Finland (36), Canada (34), Italy (31), Spain (26), Belgium (20), Switzerland (17), Austria (14), and Iceland (12). Of the total number of foreign committee members who responded to this survey, 93% came from these 15 countries and 69% from the five countries with most evaluators.

A distinction was made between other Nordic countries (Sweden, Denmark, Finland, and Iceland), other European countries, and North-America (USA and Canada) in order to examine

possible regional variation in response pattern. Of the foreign members, 42% came from other Nordic countries, 41% from European countries other than the Nordic countries, and 14% from North-America. Only 3% came from other regions of the world.

The majority of the foreign members of the PhD committees appear to be experienced evaluators. Over the last decade, about 25% had examined more than 15 PhD theses, and equally as many between 10 and 15 theses. Most of them also had evaluated PhD theses for candidates based abroad. In this same time period, 60% had examined more than two theses in other countries. Only one third of the respondents had previously been a member of a Norwegian PhD committee.

The foreign members were asked to answer the following question: How would you describe the quality of the latest Norwegian PhD thesis that you evaluated, when it comes to: (a) Originality, (b) depth and coverage, (c) theoretical level, (d) methodological level, (e) skills in written presentation, (f) contribution to the advancement of the field, and (g) external (applied/societal/cultural/industrial) relevance?

These quality properties all relate to the four elements suggested by Gulbrandsen (2000). While *originality* is a common element in both studies, we have further deconstructed the notion of *solidity* into depth and coverage, theoretical level, methodological level, and skills in written presentation. *Contribution to the advancement of the field* resembles scholarly / scientific relevance, and *external relevance* is more or less identical to practical / societal utility.

Response alternatives were excellent, very good, good, acceptable, poor, and for external relevance 'not relevant/uncertain'. The latter category was included because some research is still regarded by scientists themselves as exclusively basic, without any intention of application (Gulbrandsen & Kyvik 2010). Hence, we have in this study taken this fact into consideration.

We have used the various notions of quality of PhD theses as dependent variables, in order to study whether examiners put more weight on some of the quality elements than on others. These are continuous variables where 'excellent' is given the value 5 and 'poor' the value 1.

Based on previous studies on the assessment of doctoral theses, some independent variables have been assumed to, or have proven to, create methodological challenges: Examiners' relationship to the supervisor; the experience of the supervisors; and regional affiliation of the examiners.

Examiners' relationship to the supervisor

Tinkler & Jackson (2000) and Mullins & Kiley (2002) argue that the notion of the 'independent' evaluator is problematic in an academic environment characterized by research collaboration, previous supervisor/student relationships, and disciplinary networks. A considerable proportion of the members of evaluation committees probably have an academic or personal acquaintance with the supervisor(s), and for that reason might give a more positive assessment of the theses

than a random sample of examiners would express. Accordingly, it might be important to control for the relationship with the supervisor.

The experience of the examiners

An Australian interview study reports that inexperienced examiners were regarded to be less positive than their experienced colleagues (Mullins & Kiley 2002). Two reasons for this difference were suggested. One was that young examiners had not experienced their own PhD students being scrutinized. The other reason was that their frame of reference was quite limited; often restricted to their own PhD work more than the standards of a broader range of PhD theses. Thus, it might be important to control for the experience of the members of the evaluation committee.

Regional affiliation of the examiners

Previous evaluations of research training in Denmark (Forskerakademiet 1999) and Norway (Research Council of Norway 2002) indicate that there might be regional variations in the attitudes of external members of doctoral theses committees. Hence, it will be important to control for the regional affiliation of external members of the committee. In this study, we have distinguished between examiners from the other Nordic countries, the rest of Europe, North America (USA and Canada), and the rest of the world.

Academic field of PhD candidate

The evaluators were themselves asked to decide within which of the following academic fields the PhD thesis should be classified: Humanities; social sciences, including law and educational science; natural sciences, including mathematics; medical and health sciences; engineering and technology; agricultural and veterinary sciences; or 'other, please specify'. Respondents who chose the latter category were recoded into one of the academic fields above.

University of PhD candidate

We have also investigated whether the quality of PhD theses are rated differently according to institutional affiliation.

A regression analysis was undertaken to control for the effects of these variables.

Results

The assessment of the various quality elements is very coherent; less than 20% regarded the various elements as 'excellent', slightly more than 40% as 'very good', and 25–30% as 'good'. Only 1–3% characterized the latest thesis they had evaluated as 'poor', probably meaning that the candidate

failed. Assessments of the 'theoretical level' of the theses are slightly less favorable than those of the other elements, while assessments of 'depth and coverage' and 'skills in written presentation' are slightly better. Differences between academic fields are generally small. Overall, PhD theses in the natural sciences get slightly higher grading than average, and social science theses slightly below. While 65% of the theses in the natural sciences are graded 'excellent' or 'very good' when it comes to 'contribution to the advancement of the field', this applies to 50% of the theses in the social sciences. North-Americans give Norwegian PhD-theses better rating than do their European peers, who in turn are more positive than members from the other Nordic countries. This finding applies to all quality elements. When it comes to 'the contribution to the advancement of the field', 26% of the North-Americans characterized the theses as 'excellent', while 15% from other European countries and 9% from the other Nordic countries did so. Overall, there were very small differences in responses between those who knew the supervisor beforehand and the rest of the evaluators. Finally, the quality of PhD theses from 'old' universities was found to be higher than those from 'new' universities.

Preliminary results of the regression analysis indicate that examiners' acquaintance with supervisors has no effect on the assessment of five of the quality elements; and only a weak significant effect on the evaluation of 'methodological level' and 'depth and coverage'.

Experienced examiners give consistently better rating than their less experienced peers.

The finding presented above that examiners from North America give better rating than examiners from Europe outside the Nordic countries, and that Nordic evaluators are the least positive, is confirmed by the regression analysis.

Academic field explains very little of the variation. PhD theses in the natural sciences get significantly better rating than theses in the other fields with regard to 'originality', but significantly lower rating in terms of 'external relevance'. This finding is consistent with the results of other studies showing that natural science research is generally defined to be predominantly basic and to a little extent applied (Gulbrandsen & Kyvik 2010).

PhD theses in the fields of engineering and technology, the social sciences, and the humanities get a significantly lower rating with regard to 'theory' and 'methods'.

Theses in the humanities also get a significantly lower rating with respect to 'skills in written presentation.'

With regard to institutional affiliation of the PhD candidates, there are no significant quality differences between theses from the four traditional research universities and the specialized universities. Theses at the five institutions that recently have been upgraded to some kind of university status get, however, significantly lower rating than theses awarded by the traditional university institutions with regard to 'originality', depth and coverage', 'theoretical level', and

'methodological level'. With reference to Gulbrandsen (2000), this means that PhD theses at the traditional universities and specialized university institutions are rated higher in terms of *originality* and *solidity* of the research, while there are generally small differences with regard to scholarly / scientific relevance, and practical / societal utility.

Discussion

This paper will discuss the implications of these findings for the survey methodology used in this evaluation, and the possibilities for other researchers to apply this methodology to assess the quality of PhD theses in other national or institutional contexts.

References

EUA (2005). *Salzburg II* Recommendations. Brussels: European University Association.

Forskerakademiet (1999) *Udenlandske bedømmeres vurdering af danske ph.d.-avhandlinger*. Århus: Forskerakademiet.

Gulbrandsen, M. (2000). *Research Quality and Organisational Factors: An Investigation of the Relationship*. Trondheim: Norwegian University of Science and Technology.

Gulbrandsen, M. & Kyvik, S. (2010). Are the concepts basic research, applied research and experimental development still useful? An empirical investigation among Norwegian academics, *Science and Public Policy*, 37, 343–353.

Hemlin, S. (1993). Scientific quality in the eyes of the scientist. A questionnaire study, *Scientometrics*, 27, 3–18.

Johnston, S. (1997). Examining the examiners: an analysis of examiners' reports on doctoral theses, *Studies in Higher Education*, 22, 333–347.

Mullins, G. & Kiley, M. (2002). 'It's a PhD, not a Nobel Prize': How experienced examiners assess research theses, *Studies in Higher Education*, 27, 369–386.

Research Council of Norway (2002) *An Evaluation of Norwegian Training of Researchers*. Oslo: Research Council of Norway.

Thune, T., Kyvik, S., Sörlin, S., Olsen, T.B., Vabø, A. & Tømte, C. (2012). *PhD Education in a Knowledge Society. An Evaluation of PhD Education in Norway*. Oslo: NIFU.

Tinkler, P. & Jackson, C. (2000). Examining the Doctorate: institutional policy and the PhD examination process in Britain. *Studies in Higher Education*, 25, 167–180.

Discoveries and citations over time in solar cell technology[1]

Katarina Larsen katarina.larsen@abe.kth.se / Division of History of Science, Technology and Environment, KTH Royal Institute of Technology, Teknikringen 72, Stockholm, 100 44 (Sweden)

Clara Calero-Medina clara@cwts.leidenuniv.nl / Centre for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62-A, Leiden, 2333 (the Netherlands)

Abstract

This study is analysing knowledge networks and science domains in solar cell technology to uncover how emerging fields evolve over time in the quest for more efficient, cheaper and robust solutions in solar cell technology. The results from the study and methodology development contributes with an interesting new empirical setting for citation network analysis applied to solar cell technology with a new understanding of 'how you make' solar cells different from existing techniques. Discoveries and advances of the field are described using citation data from the 'backbone' of publications in this particular field of dye-sensitized solar cells. The analysis of this complex field can contribute to a better understanding of how new areas of inquiry can be traced using citation data, as shown in the example of the organic dyes. This particular type of solar cells is in the crossroads of chemistry, physics and material sciences and can thereby also provide further insight into the dynamics of interdisciplinary technology.

Introduction

In solar cell technology, there are a number of parallel technical solutions. This study is focusing on a special type of solar cells that are imitating the photosynthesis in how it is designed to capture the light from the sun by using dyes (in a similar way as plants use chlorophyll). This solar cell technology is therefore often called Dye-sensitized solar cells and had some key breakthroughs in the late 1980s and early 1990s. Some features of these solar cells are that they offer other design opportunities in lightweight solutions, transparency and multi-color options facilitating building integration and design for consumer products (Hagfeldt et al. 2010). As a field of solar cell technology it is characterized by interactions across disciplines (Larsen 2008) with a breakthroughs following efficiency improvements reported in scientific publications in late 1980s and early 1990s (O'Regan & Gratzel 1991). This empirical case of solar cell technology can thereby improve our understanding of emerging fields in renewable energy technology. Previous

1 This work was supported by the Swedish Foundation for International Cooperation in Research and Higher Education (STINT).

analyses of technological systems for renewable energy (wind- and solar energy) discuss mechanisms of cumulative causation (Jacobsson & Bergek 2004) and lock-in for photovoltaic technology and design (Sandén 2005). However, these studies were not specifically analysing the emergence of novel technology using citation data for analysis of distributed innovation processes. One example is that the discovery processes is distributed across university departments in analysis of the dyes (pigments used to harvest sunlight) made by Physics departments (Hallin 2010) and similarly the importance of Chemistry departments in creating these dyes and elaborate on variations for improving the efficiency of the solar cell. This particular type of solar cells is in the crossroads of chemistry, physics and material sciences and can thereby also provide further insight into the dynamics of interdisciplinary technology.

Objective and contribution of study

The broader objective is to study how an emerging field in solar cell technology evolve over time in search of more efficient and robust solutions. The study aims to contribute to an understanding of emerging technology and improvement sequences by analysis of the empirical case of solar cell technology. This is done by citation network analysis over a time period of more than 20 years. Methodologically, the work is expanding on studies of network analysis of dye-sensitized solar cells (Larsen & Calero 2012; Larsen 2007) to analyse knowledge domains innovation sequences in research and development of novel solar cell technology. Thereby, this case of solar cell technology also contribute to methodology development made in previous studies of the micro-foundations of innovation in other areas of application oriented research and development, including studies of medical sciences of (Mina el al. 2007, Ramlogan et al. 2007). Previous work about discovery processes has examined milestone events using citation data for DNA (Garfield et al. 1964) but analysis of discovery processes over time using citation data in renewable energy is less examined and further analysis is warranted.

Data and methods

The data collection was established as follows. We started with all publications citing the O'Regan & Gratzel (1991) paper in Nature. This collection is called Set 0. From this Set 0 we created a list of most highly cited papers on the reference lists of Set 0. From this list we extracted the top 10 and by coincidence none of them where previous to 1991. But because we needed to have an insight of the knowledge previous to the O'Regan & Gratzel (1991) paper so we went to the next 15 most highly publications and from there we extracted the papers published before 1991. We added them to the top 10 and we refer to this list as L1. Subsequently, we expanded Set 0 with all publications citing at least one publication from list L1. This yields Set 1. The final selection of papers to represent the fields is established by all publications in Set 1 plus the publications from list L1. In current analysis, we have a citation matrix of about 12800 papers up until 2008.

This study is combining quantitative (citation and co-authorship networks) and qualitative (interviews with researchers and secondary data) methods to uncover mechanisms though which knowledge networks and science domains evolve over time in the quest for more efficient, cheaper and robust solutions in developing a novel type of solar cell technology. To achieve that, this study is combining methodology earlier applied to study the spread of ideas (Calero & Noyons 2008) and identification of research groups (Calero et al. 2006) with the identification of topical "islands" of knowledge domains (Batagelj & Zaversnik 2004) with studies of knowledge networks in emerging fields of solar cell technology (Larsen 2008).

First the citation network enables us to measure the change over time of the connectedness of the system. In the evolution of knowledge, phases of consolidation of past results coexist with exploration of new approaches. One of the techniques is called main path analysis (Calero & Noyons 2008). In the work presented here, the analysis of main path and main component of articles refers to the understanding that if knowledge flows through citations, a citation that is needed in paths between many articles is more important than a citation that hardly plays any role linking articles. For analysis of the "chains" of citations from the most recent records to the oldest, the network algorithm computes the paths that are most frequently encountered (the backbones of the research tradition). Finally we have shown part of the results of our analysis to experts in the field, enriching the analysis with a better understanding of collaborative research ventures.

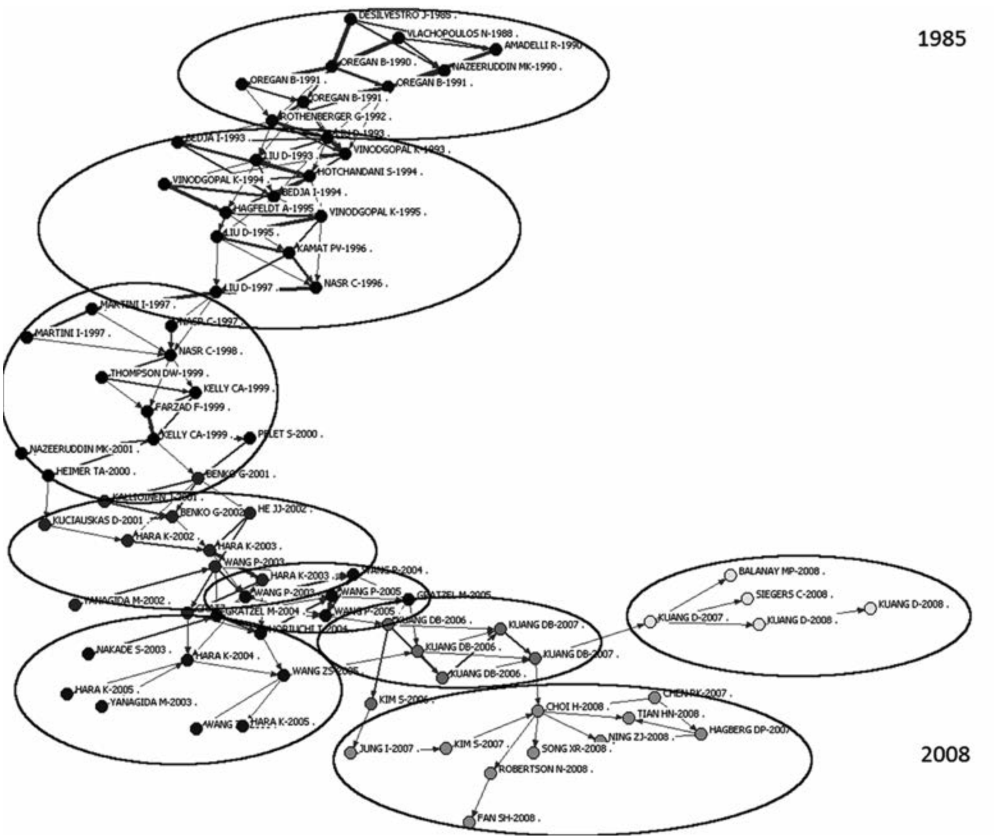
Results

Some preliminary results from the citation analysis, shown in Figure 1, is the main component with the islands identified based on the data until year 2008. The main path shows "chains" of citations from the most recent papers to the oldest, the network algorithm computes the paths that are most frequently encountered (the backbones of the research tradition). The main sub-network contains not only the main path but also other important branches from the citation network providing rich information about the developments of the field. Finally taking the main sub-network as starting point we identify connected groups of nodes (islands) that locally dominate according to the values of lines. They are sub-networks trying to embrace single theme clusters (nine of them), see Figure 1. Each island is represented in a different colour (grey-scale). The results from the analysis presented are based on Pajek (Batagelj & Mrvar 2006).

The islands are connected sub-networks (groups of vertices) that locally dominate according to the values of vertices or lines. The main component (mentioned above) network can be seen as some kind of a landscape, where the papers with the largest value have the highest peak. Eliminating the papers (and the corresponding lines) with height lower than t, we obtain a group of internally connected sub-networks – islands called a *vertex cut at cutting level t*. We are usually interested in sub-networks with specific number of vertices – not smaller than k and not larger than K – trying to embrace single theme clusters. To identify such islands we have to determine vertex cuts at all possible levels and select only those islands of the selected size. Zaversnik &

Batagelj (2004) developed an efficient algorithm for determining such islands. In Figure 1, we have selected the islands with a size between 5 and 15 papers. We made this decision based on reaching an amount of papers that would be manageable and interesting to be analyzed and interpreted together as part of the island.

Figure 1. Main component of citations. The nodes represent the papers, the colour of the nodes (grey-scale and highlighted with circles) the island (nine overall), the arrows "cited by", the vertical dimension represents approximately publications year. Time period 1985–2008, starting with Desilvestro et al. 1985 (in the first topical island at the top) and ending with the White island (2007–2008) and Light-grey island (2007–2008) in the lower end of figure.



Since the islands are representing highly related areas of inquiry as measured by citation network analysis this can be used for analysis of new problem areas, as shown in earlier studies (Mina et al 2007). Figure 1 is highlighting the island in the early time period (1985–1991) and in the period following directly after. In the early period, some key phenomenon was observed sparking the

ideas further about how to solve also practical concerns with stability of components (such as dyes) and the device used in the experiments. In the early phase of the development in the field the stability of the use of dyes in the solar cell device was a concern, although results demonstrated high surface area (Desilvestro et al. 1985).

In 1991, the findings with the publication in Nature (O'Regan & Gratzel 1991) demonstrated higher efficiency of this type of solar cells (around 7–8%) followed by development of pigments used increasing efficiencies (Hagfeldt et al. 2010). The period following after, which can be described as the establishment of the DSC field of research in the period 1991–1995, are in the citation network represented by publications with original contributions in addition to review-papers in journals such as Chemical reviews (Hagfeldt & Gratzel 1995). In the following years "a relatively simple picture of how DSCs operate" developed as described in the review articles (Hagfeldt et al. 2010, page 6596) while analyses reported in scientific publications (in period 2000–2005) of basic concepts and electrochemical reactions have increased the awareness of the chemical complexity of the solar cell device.

Other examples illustrate the continuous search for improvements in emerging areas of inquiry in discovery processes of organic dyes in more recent periods. This search for improved organic dyes is based on arguments relating to low cost, easily synthesized and also work with more environmentally friendly substrates as compared to noble metals used in other options for dyes examined (Hagfeldt et al. 2010). In Figure 1, publications in this area of inquiry are represented in recent time periods of 2007–2008.

Conclusion

One contribution from the study relates to the nature of distributed innovation in this particular case of solar cells. In previous work, improvement sequences are analysed using citation networks in testing of clinical devices (Mina et al. 2007). Application in devices is one aspect of solar cells – another is the interdisciplinary interaction that is central for the development of this type of solar cells. This requires that research labs are well networked in terms of sharing knowledge (about function of new dyes and measurement techniques) with researchers in other disciplines.

The methodology development in this study contributes with an interesting new empirical setting for citation network analysis applied to solar cell technology. In particular, this area represents analysis of innovation processes in a complex setting where changes in one part of the solar cell system (dyes or electrolyte for example) generate new questions for the researchers. The main component analysed here is capturing key aspects of the field and how it is evolving over time (the backbone of the research tradition). The analysis of citations covers a time period over more than 20 years. In the time period studied, starting in 1985, there were some important advances for establishing the field. In later periods, the citation networks reflect intensification of search for new dyes, including organic dyes, for capturing the sunlight. Further analysis draws on analysis of

the topical islands combined with interviews of researchers in the field to address early advances in the emerging field. Thereby contributing to a further understanding of discovery processes in complex fields in the quest for improvements and innovation in solar cell devices.

References

Batagelj, V. & Mrvar, A. (2006). Pajek: Program Package for Large Network Analysis, University of Ljubljana, Slovenia. Retrieved via: http://vlado.fmf.uni-lj.si/pub/networks/pajek/

Calero, C.M., Buter, R., Cabello Valdes, C. & Noyons, E.C.M. (2006). How identify research groups using publication analysis: an example in the field of nanotechnology. *Scientometrics*, 66, 365–376.

Calero, C.M. & Noyons, E.C.M. (2008). Combining mapping and citation network analysis for a better understanding of the scientific development: The case of the absorptive capacity field. *Journal of Informetrics*, 2, 272–279.

Desilvestro, J., Gratzel, M., Kavan, L., Moser, J. & Augustynski, J. (1985). Highly efficient sensitization of titanium dioxide. J. American Chemical Society, 107, 2988–2990.

Garfield, E., Sher, I.H. & Torpie, R.J. (1964). The use of citation data in writing the history of science. Philadelphia: Institute for Scientific Information.

Hagfeldt, A., Boschloo, G., Lincheng, S., Kloo, L. & Pettersson, H. (2010). Dye-sensitized solar cells. *Chemical Reviews*, 110, 6595–6653.

Hagfeldt, A. & Gratzel, M. (1995). Light-induced reactions in nanocrystalline systems. *Chemical Reviews*, 95, 49–68.

Hallin, M. (2010). Electronic and molecular surface structures of dye-sensitized TiO2 interfaces. Doctoral Thesis from Uppsala University, Sweden. Department of Physics and Astronomy, Molecular and condensed matter physics.

Jacobsson, S. & Bergek, A. (2004). Transforming the energy sector: the evolution of technological systems in renewable energy systems. Industrial and Corporate Change, 13(5), 815–849.

Larsen, K. (2008). Knowledge network hubs and measures of research impact, science structure, and publication output in nanostructured solar cell research. *Scientometrics*, 74(1), 123–142.

Larsen, K. (2007). Interdisciplinarity in environmental technology applications – examining knowledge interaction between physics and chemistry research teams. Paper presented at the 11th International Conference on Scientometrics and Informetrics, Madrid, June 2007.

Larsen, K. & Calero C.M. (2012). Citation networks over time: distributed innovation and dyes in solar cell technology. Sunbelt conference, arranged by INSNA – International Network for Social Network Analysis, Redondo Beach, CA, US. March 2012.

Mina, A., Ramlogan, R., Tampubolon, G. & Metcalfe J.S. (2007). Mapping evolutionary trajectories: applications to the growth and transformation of medical knowledge. *Research Policy*, 36, 789–806.

O'Regan, B. & Gratzel, M. (1991). A low-cost, high-efficiency solar cell based on dye-sensitized TiO2 films. *Nature*, 353, 737–740.

Ramlogan R., Mina, A., Tampubolon, G. & Metcalfe, J.S. (2007). Networks of knowledge: the distributed nature of medical innovation. *Scientometrics*, 70(2), 459–489.

Sandén, B. (2005). The economic and institutional rationale of PV subsidies. *Solar Energy*, 78, 137–146.

Zaversnik, M. & Batagelj, V. (2004). Islands. Presented at Sunbelt XXIV Conference, Portoroz, Slovenia, May 2004.

What Differentiates Top Regions in the Field of Biotechnology? An empirical Study of the Texture Characteristics of 101 Biotech Regions in North-America, Europe and Asia-Pacific

Catherine Lecocq — Cathy.Lecocq@kuleuven.be / VIVES, Steunpunt Ondernemen en Regionale Economie, KU Leuven, Naamsestraat 61 bus 3550, 3000 Leuven (Belgium)

Bart Van Looy — Bart.Vanlooy@kuleuven.be / Department of Managerial Economics, Strategy and Innovation, Faculty of Business and Economics, KU Leuven; Research Division INCENTIM, KU Leuven; School of Management and Governance, University of Twente; Expertisecentrum O&O Monitoring (ECOOM), Naamsestraat 69 bus 3535, 3000 Leuven (Belgium)

Abstract

Over the last decade, the cluster phenomenon has triggered the attention of researchers and policy makers. While several case studies provide valuable insights on the dynamics within (individual) biotech clusters (e.g. Breznitz et al., 2008; Chiarone and Chiesa, 2006; Bartholomew, 1997), large-scale quantitative studies addressing the texture characteristics of clusters are lacking. Building on patent and publication-based indicators, our analyses encompass the texture characteristics of 101 regions in North-America, Europe and Asia-Pacific that developed substantial technological activities in the field of biotechnology over the period 1992–1997 (the first growth phase of the biotech industry). Our findings signal two distinctive types of regions that are able to obtain the status of 'top region': "concentrated" regions in which technology development is mainly situated within private firms and "distributed" regions where technology development is more equally distributed between private firms and public knowledge institutes. Using logit regression models, we investigate which texture characteristics (industrial texture, and presence of entrepreneurial-orientated knowledge institutes)) differentiate top regions from other biotech regions. Our findings reveal that both types of regions differ to some extent in terms of antecedents of growth.

Introduction

Biotechnology is often considered as one of the promising technologies that will bring economic growth and welfare to a region. Evidence indicates that regions that have developed into successful biotech clusters are very limited (Audretsch & Feldman, 1996; Feldman & Florida, 1994). Thriving clusters such as the San Francisco Bay Area, San Diego and Boston have therefore been widely studied by researchers and policy makers in order to identify the main factors behind the success of those biotech clusters. General consensus exists that well developed biotech regions, so-called

clusters or hot spots, are characterized by the presence of world-class scientific research, high levels of entrepreneurial activity (both academic spin-offs and industrial ventures), high labour mobility and dense social networks, and the presence of venture capital and a dedicated support infrastructure (e.g. Casper, 2007; Cooke, 2001; Owen-Smith, Riccabonni, Pammolli & Powell, 2002). About the respective role and importance of public knowledge institutes and private firms for the emergence and early development of biotech regions different perspectives are being advanced. Case study research provides evidence that universities and knowledge generating institutes have played a central and active role in the creation of biotech clusters in the region of Boston (Breznitz, O'Shea, & Allen, 2008) and San Francisco Bay area (Chiarone & Chiesa, 2006). In contrast, private firms have played a prominent role in the development of biotech activities in the regions of Milano, Italy and Uppsala, Sweden (Chiarone & Chiesa, 2006) as well as in Japan (Bartholomew, 1997).

While case study research provided valuable insights on the characteristics and the dynamics within individual (biotech) clusters, so far large-scale empirical studies addressing the texture characteristics of regions are absent. As (industrial) biotechnology is entering a growth phase (Lecocq & Van Looy, 2009), the question whether regions can evolve into leading clusters by relying on a distributed texture or whether the presence and/or emergence of an anchor tenant firm (Agrawal & Cockburn, 2003) is a prerequisite in this respect becomes pertinent, both for practitioners and policy makers engaged in regional economic development.

Building on patent and publication-based indicators, we engage in such a study in the field of biotechnology. Our analyses cover 101 regions from North-America, Europe and Asia-Pacific that developed substantial technological activities in the field of biotechnology over the period 1992–1997. The period of analysis corresponds with an era of rapid growth in biotech industry. First, we identify the worldwide leading regions in terms of biotech technology development in the period 1992–1997. Next, we analyze the texture characteristics of those 'top' regions in terms of industrial base and the presence of entrepreneurial-orientated public knowledge institutes. In the last part of the analyses, we study which texture characteristics are important to become a top region in the field of biotechnology.

The field of modern biotechnology

Modern biotechnology is a complex, knowledge-intensive field that has generated important breakthroughs in the industry, in particular the pharmaceutical industry (Arora & Gambardella, 1990; Zucker & Darby, 1997), by enabling the creation of entirely new organic materials and profoundly changing the process of (drug) discovery and product development (Powell et al. 1996). Of crucial importance for the origin of modern biotechnology was the discovery of the double helix structure of DNA (1953) by Watson and Crick in the laboratories of the University of Cambridge (UK). The foundation for the modern biotech *industry* was laid in 1973, when professors Cohen (Stanford University, US) and Boyer (University of California, US) discovered

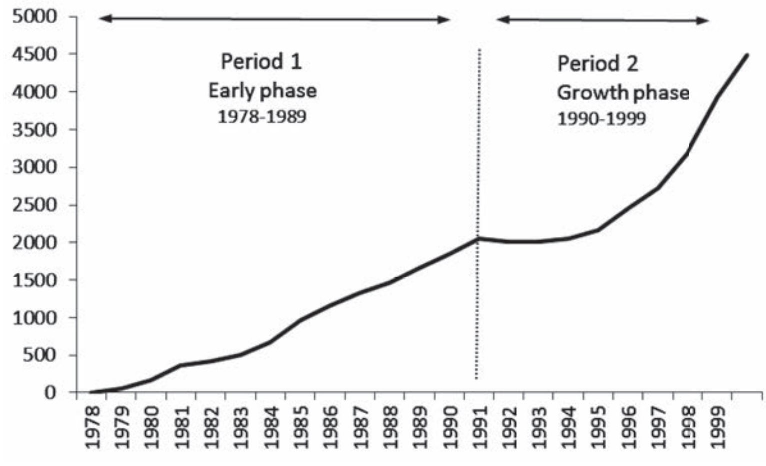
the recombinant DNA technique, which allowed to transfer the basic science of molecular biology into useful knowledge for a wide range of industrial applications (Feldman, 2003).

Following the discovery of the recombinant DNA technique, the second half of the 1970s and the 1980s was marked by the creation of the first companies dedicated to modern biotechnology, the so-called *New Dedicated Biotech Firms* (NDBFs). These new biotech companies were often cofounded by, or maintained strong linkages with academic researchers (Zucker & Darby, 1996). They focused on exploring new technological and scientific research results and translating them into the commercial domain (Acharya, 1999; Galambos, 2006). As new, scientific knowledge is often characterized by a substantial amount of tacit knowledge, developing an idea from science most often requires close links with the academic inventor(s) (Zucker & Darby, 1996; Rosenberg & Nelson, 1994) and NDBFs were therefore most often established in close vicinity of universities or research centres.

By the late 1980's, large pharmaceutical firms started to display interest in the field of biotechnology. From the early 1990s onwards, they entered into the field by setting up strategic alliances with and/or acquiring small biotech firms. A large strand of literature shows that further technology advances in the biotech industry relies to a large extent on interorganisational collaborations between organizations with complementary resources, including universities and research centers, new dedicated biotech firms and mature pharmaceutical and chemical firms (Arora & Gambardella; Gertler & Vinodrai, 1996; Powell et al. 1996).

Figure 1 presents the evolution of biotech technology development, measured by the number of EPO patent applications (worldwide), over the period 1978–1999. The figure shows a steady, linear increase in the number of patent applications in the early phase of the biotech industry (period 1978–1990), followed by an exponential growth in the number of patent applications from the early 1990s onwards (Lecocq & Van Looy, 2009). This study focusses on the period 1992–1997 characterized by rapid growth of biotech technology development.

Figure 1. Evolution of patenting in the field of biotechnology (EPO Patents 1978–1999, worldwide)



Towards hypotheses

It is clear that in the early days of the biotechnology industry, basic research and entrepreneurial activities have played an important role for the development of biotech technology activities. In the 1990's, when the first technologies (products) are being commercialised and large established players enter the field, industrial capabilities are evidently becoming more important. In order to profit from the take-off of economic activities in the growth phase of biotech, regions may therefore benefit from a different configuration in terms of presence of (entrepreneurial) research universities and industry composition (presence of new dedicated biotech firms and more established firms) than in the early days of the industry.

In this research, we look at the texture characteristics of regions in relation to the technological performance of regions during the growth phase of biotech (period 1992–1997). Specific attention is given to industrial texture characteristics and the entrepreneurial orientation of scientific actors in the regions.

Industrial texture characteristics

By their nature and core raison d'être', firms are best placed to identify market needs, translate technological opportunities into prototypes and commercial products, and bring these new products to the market. Even in science-intensive fields such as biotechnology, private firms remain the major player on the market place. In regions with a critical mass of activities directed towards market exploitation and commercialization, firms have more opportunities to interact and learn from high-quality suppliers, demanding (industrial) customers and other innovative

firms producing similar or complementary goods and services (Porter, 2000) resulting in enhanced innovation dynamics in the region.

The concentration of innovative activities within larger, R&D intensive firms might be of particular relevance for the development of a new industry because of their scale and access to larger financial resources as compared to new and / or small firms (Gray & Parker, 1998). By creating local niches and/or intermediary markets, larger firms may also encourage entrepreneurial activity in the region and attract high-quality suppliers which would not be present or of lower quality in the absence of the anchor firm (Agrawal & Cockburn, 2003).
Therefore, we propose that:

Hypothesis 1a: *Regions in which technology development activities are to a larger extent situated within firms, are more likely to become a leading biotech region in the growth phase of biotech.*

Hypothesis 1b: *Regions with higher levels of concentration of regional biotech technology development activities (by private firms) within an anchor tenant firm are more likely to become a leading biotech region in the growth phase of biotech.*

Entrepreneurial-orientated knowledge institutes

For firms active in complex, science-intensive fields such as biotechnology searching for and acquiring highly-specialized scientific knowledge from outside the boundaries of the organization is essential (Powell, Koput & Smith-Doerr, 1996). As the field of biotech further develops, relevant scientific knowledge is also spreading on a more global scale. The diffusion of knowledge is further enhanced by the numerous international collaborations between public knowledge institutes and private firms (e.g. Cooke, 2001; Zeller, 2001; Lecocq & Van Looy 2009). Therefore, the presence of a strong local science-base and entrepreneurial-oriented knowledge institutes in the region may become less important in later stages of the technology life cycle. This leads to the following two hypotheses:

Hypothesis 2a: *The science-intensity of a region, measured by the number of publications per population, is no longer instrumental for becoming a leading biotech region during the growth phase of biotech.*

Hypothesis 2b: *Likewise, the entrepreneurial orientation of scientific actors, measured by their involvement in technology, is no longer instrumental for becoming a leading biotech region during the growth phase of biotech.*

Data

To identify the worldwide leading clusters in terms of biotech technology development and study the texture characteristics of biotech regions in a quantitative way, we draw on the dataset with EPO patent applications and Web of Science publications in the field of biotechnology created by Glänzel et al. (2004). We focus on the time frame 1992–1997, the period of rapid growth of the biotech industry.

In a first step, all patents and publications have been allocated to their respective regions based on the address information of applicants (patents) and authors (publications) following the "*patent allocation methodology*" developed by Lecocq, Van Looy & Vereyen (2011). Table 1 shows, for every country, the regional level of analysis selected in this study in order to provide comparable units of analysis in terms of population. Only those regions that developed a substantial amount of biotech activity over the time period 1992–1997 (minimum 18 EPO patent applications, i.e. on average three patents/year) are retained for the analyses[i].

Table 1. Regional level of analysis[ii]

Australia	States (n=6) and major mainland territories (n=2)
Canada	Provinces (n=10) and territories (n=3)
Europe (EU-15 + Switzer-land)	Nuts1/2 regions (n=197)
Japan	Prefectures (n=47)
United States	States (n=51)

The "*sector allocation methodology*" developed by Du Plessis et al. (2011) allows to identify which type of actor (private firms, public universities and research centres, research hospitals and/or persons) applied for the patent.

Based on the "*name harmonizing method*" of Magerman et al. (2011), we identify the firm and/or other actor with the largest number of patents in the region. In the study, we refer to those firms and other actors as the "lead company" and the "lead actor" in the region.

The "lead company" in the region is further classified as "New, Dedicated Biotech Firm" (NDBF), "Established Firm (EF) or "Other firm" according to the definitions in Table 2. This classification of firms relies on information on the industry(ies) in which the firm is (primarily) active, it's year of establishment and the location of the headquarter[iii] retrieved from company websites and other sources[iv] such as reports on merger and acquisition activities and new product and technologies in the field of life science, market research reports and company profiles.

Table 3 provides an overview of the texture variables used in the regression analyses of the paper.

Table 2. Refinement of the typology of firms

New Dedicated Biotech Firms (NDBF)	Firm primarily active in the field of biotechnology and established after 1974.
Established Firm (EF)	Firm primarily active in other fields than biotechnology (e.g. pharmaceutical, chemical, food and other industries) and established before 1974.
Other Firm	Firm active in the field of biotechnology but not as a product or research company (e.g. regional technology transfer offices, venture capitalist, regional industrial agency)

Table 3. Texture variables

Variable	Description
Number of firms	Number of companies active in biotech patent applications in the region.
Company concentration index	Ratio of the number of biotech patents of the leading firm in the region and the total number of company biotech patents in the region.
Science-intensity of the region	Number of biotech publications in the region per 1000 population.
Entrepreneurial orientation of knowledge institutes	Ratio of the total number of biotech patents applied by public knowledge generating institutes in the region and the total number of biotech publications in the region
International collaboration with firms	Number of biotech co-patents in the region with a firm from outside the country
International collaboration with knowledge institutes	Number of biotech co-patents in the region with a knowledge generating institute from outside the country

Analyses

Top regions in biotech

Regions with the highest count of biotech patents (based on assignee addresses) are considered as leading regions in biotech. Table 4 shows the 15 leading regions in biotech over the period 1992–1997. Most top regions are located in the US, e.g. North-California (San Francisco region), Massachusetts (Boston) and South-California (San Diego region). Japan has two top regions in biotech: Tokyo and Osaka. The three largest biotech regions in Europe are Île de France (Paris region, France), Denmark and London (United Kingdom). Biotech technology development activities

are highly concentrated in a few regions or clusters worldwide: together the 15 leading regions in terms of biotech technology development account for 56% of all biotech patent activity.

Table 4. Leading biotech regions (period 1992–1997)

Rank	Regio, land	patents 1992–97
1	North California, US	1,083
2	Tokyo-TO, Japan	921
3	Massachusetts, US	824
4	South California, US	711
5	New Jersey, US	650
6	New York, US	626
7	Maryland, US	576
8	Île-de-France, France	563
9	Osaka-FU, Japan	477
10	Pennsylvania, US	449
11	Denmark, Denmark	376
12	Inner London, UK	328
13	Illinois, US	305
14	Karlsruhe, Germany	288
15	Nordwestschweiz, Switzerland	280

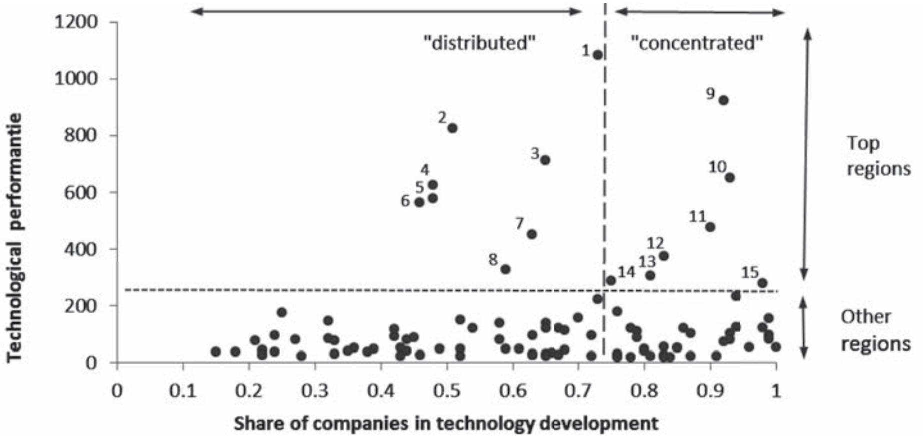
Towards a typology of (leading) biotech regions

The history of the biotech industry illustrates that different types of actors ranging from private firms (new dedicated biotech firms and established firms) to public knowledge institutes and research hospitals are involved in biotech technology development. In this part of the analysis, we investigate whether during the growth phase of biotech, technology development activities in 'top' regions is to a larger extent driven by firms as compared to the other biotech regions.

Figure 2 presents the technological performance of regions and the share of biotechnology development activity undertaken by private firms for the 101 regions in our study. The figure again confirms the strong geographical concentration of biotech technology development. Overall, no obvious, linear relationship can be discerned between the performance of regions and the share of technology development undertaken by private firms. On the one hand, one observes no 'top regions' when the share of companies (in terms of technology development) is situated below 40%. On the other hand, when looking at the leading regions only, we notice that in some regions technology development activities are highly concentrated within firms (share of company patents above 75%), while in other regions technology development is much more distributed over private firms and other types of actors (with the share of company-owned patent situated

around 50%). These results indicate that to become a leading biotech region in the growth phase of biotech, regional technology development activities do not need to be primarily driven by private firms.

Figure 2. The technological performance of regions and the share of biotech technology development activities undertaken by private firms



The top 15 leading regions in biotech technology development over the period 1992–1997 are **1**. North California, US; **2**. Massachusetts, US; **3**. South California, US; **4**. New York, US; **5**. Maryland, US; **6**. Île de France, France; **7**. Pennsylvania, US; **8**. Inner London, UK ; **9**. Tokyo-TO, Japan; **10**. New Jersey, US; **11**. Osaka-FU, Japan; **12**. Denmark, Denmark; **13**. Illinois, US; **14**. Karlsruhe, Germany; and **15**. Nordwestschweiz, Switzerland.

Annex 1 shows for each of the 15 main biotech regions, the lead actor(s) in the region, where lead actor is defined as the organization with largest number of biotech patent applications in the years 1992 to 1997. For the leading biotech regions where technology development is highly concentrated within private firms (regions 9 to 15 in Figure 2), the leading organization in the region is always an established firm, mostly primarily active in pharmaceuticals. In the leading biotech regions where technological activity is much more distributed over private firms and other actors (Regions 1 to 8 in Figure 2), the leading organisations in the region imply a combination of public research institutes (university, research center or research hospital) and private firms (New Dedicated Biotech Firm or Established Firm).

The analysis of the texture characteristics of the top biotech regions thus provides evidence for the presence of two types of regions: regions in which technology development is mainly situated or concentrated within private firms – hereafter called "concentrated regions" – and regions where technology development is more equally distributed between private firms, and entrepreneurial universities and/or research centres/hospitals, hereafter referred to as "distributed regions". Figure 2 shows that both a distributed and a concentrated texture can give rise to a

leading technology cluster in biotech. The T-test statistics on the refined texture variables in Table 5, further reveal some distinct features of "distributed" versus "concentrated" biotech regions. "Concentrated" regions are characterized by a higher share of technology development activities by private firms. Technology development activities by private firms is also much more concentrated into the leading firm in the region than in the "distributed" regions. "Concentrated" regions also engage more into international technology development collaborations with knowledge generating institutes. "Distributed" regions are characterized by a higher science-intensity of the region, measured by the number of publications per population, as well as the presence of universities and research centres with a more entrepreneurial orientation.

Table 5. T-test statistics on leading "distributed" versus "concentrated" regions (yearly figures, period 1992–1997)

	Top "distributed" regions (n = 8)	Top "concentrated" regions (n = 7)	diff	t-test	
Share of company patents	0,58	0,89	-0,32	-14,1	***
International collaboration with knowledge institutes	0,65	1,38	-0,74	-2,7	***
International collaboration with firms	0,88	0,95	-0,08	-0,3	
Entrepreneurial orientation of knowledge institutes	0,014	0,004	0,010	11,9	***
Company concentration index	0,28	0,42	-0,14	-3,2	***
Number of firms	23,7	20,3	3,40	1,0	
Science-intensity of the region	0,31	0,24	0,07	2,3	**

What differentiates leading regions?

During the growth phase of biotech, we find evidence of regions catching up and regions falling back over time in the ranking of leading biotech regions. In regression models that follow, panel data for the 101 biotech regions are used to analyze which texture characteristics differentiates leading regions from other biotech regions, by means of logit models with the following functional form:

$$P(y_{it} = 1 \;/\; x_{it}) \text{ with } t = 1\text{–}7, x_{it} \text{ contains the explanatory and the control variables}$$

The dependent variable in the logit regression models is a dummy which takes the value 1 if the region is among the top 15 regions in year t, or else the value 0. Random effects take into account and control for the panel structure of the data. The explanatory variables are the refined texture variables presented in the data section (Table 3). We further include the size of the region measured by its population, and time-specific effects. A US dummy variable is used to control for a possible "first mover"-effect of US regions in the field of biotechnology.

Table 6 shows the results of the logit regression models. In Model 1, analyses are run for all 101 biotech regions. As previous results showed that leading regions have different texture characteristics, separate analyses are also run for the regions with a "distributed" texture (n= 64, Model 2) and the regions with a "concentrated" texture (n=37, Model 3), where the latter have been defined as those regions in which technology development activities is predominantly situated within private firms (share of company patents >= 0.75) and the leading player in the region (period 1992–1997) is an established firm.

Table 6. Random Effect Logit models

	Model 1 All Regions	Model 2 Regions with distributed texture	Model 3 Regions with concentrated texture
Science-intensity of the region	17.56** (7.89)	86.18*** (25.07)	52.39*** (17.34)
Number of firms	0.69*** (0.18)	2.17** (0.87)	0.92** (0.41)
Company concentration index	6.14** (3.02)	-22.77 (23.41)	25.34*** (9.56)
Entrepreneurial orientation of knowledge institutes	201.14** (102.21)	1479.37*** (557.68)	504.07 (329.23)
International collaboration with knowledge institutes	1.28** (0.60)	3.36 (5.24)	4.32*** (1.25)
International collaboration with firms	-0.41 (0.57)	-1.98 (3.08)	-0.31 (0.99)
Population	0.0005 (0.0003)	0.0007 (0.0006)	0.0032*** (0.0012)
US dummy	-0.23 (1.96)	14.24* (7.57)	-6.07 (6.38)
Time	-0.78*** (0.29)	-0.87 (1.16)	-2.52*** (0.76)
Constant	-19.06*** (5.03)	-75.49*** (18.35)	-51.66*** (10.34)
Observations	606	384	222
Loglikelihood	-525.07	-9.86	-27.23
P	0.0086	0.0021	0.0004

The results in Table 6 (Model 1) reveal that higher levels of science-intensity as well as increasing numbers of firms active in biotech technology development contribute to achieving the status of 'top region'. These results hold for both "distributed" (Model 2) and "concentrated" regions (Model 3) and indicate that in science-intensive industries such as biotechnology, the continuous development of a strong science base remains important, also in the growth phase of the tech-

nology. At the same time, the results indicate that the creation or attraction of companies active in biotech technology development is instrumental as well in this respect .

The regression analyses (Model 3) further reveal that leading biotech regions with a "concentrated" texture not only benefit from increasing numbers of firms active in biotech technology development; higher levels of concentration within an 'anchor tenant' firm are instrumental for a leading position as well.

Next, the regressions also show that "distributed" regions (Model 2) benefit from a stronger entrepreneurial orientation of the knowledge institutes in the region, while no significant impact is found for the "concentrated" regions (Model 3).

Finally, the analyses reveal that "concentrated" regions (Model 3), in which a positive impact of entrepreneurial-oriented institutes is largely absent, do benefit from international technology collaborations with knowledge institutes. For the "distributed" regions (Model 2), no similar effect is found in terms of international collaboration. The results also reveal no significant impact from international technology collaborations with firms.

Discussion and conclusions

Biotech technology development activities are highly concentrated in a limited number of top regions or clusters worldwide (Audretsch & Feldman, 1996; Feldman & Florida, 1994). In this paper, we study the texture characteristics of regions (industry composition, presence of entre preneurial-orientated scientific actors) that are instrumental for becoming a top region during the first growth phase of the biotech industry (period 1992–1997). Random Effect Logit models are used to analyze which texture variables differentiates leading biotech regions from other regions.

Our results provide evidence for the presence of *two* types of leading biotech regions: "concentrated" regions in which technology development is mainly situated within private firms and "distributed" regions where technology development is more equally shouldered by private firms, entrepreneurial universities and/or research centres/hospitals. These results indicate that to become a leading biotech region in the growth phase of biotech, regional technology development activities do not necessarily need to be primarily driven by private firms.

Further, our analyses indicate that regions with "concentrated" texture characteristics benefit, in terms of overall technological activity, from increased levels of concentration of technology development activities within a leading firm, thereby supporting the anchor-tenant hypothesis proposed by Agrawal & Cockburn (2003). Further research reveals that the "anchor" firm(s) in the leading "concentrated" regions are large, R&D intensive firms primarily active in the pharmaceutical, chemical, food and other industries, and established well before the creation of the first

dedicated biotech firms in the second half of the 1970s. Our analyses suggest that these large established firms, which have extensive industry experience and important access to (internal) financial resources, have been of particular importance for the development of regional biotech technology activities in the first growth phase of the biotech industry. Following Agrawal & Cockburn (2003), such large, R&D intensive firms may have played an important role in breeding regional entrepreneurial initiatives in the field of biotechnology. Our results also indicate that regions with a "concentrated" texture benefit from engaging in international technology collaborations with scientific actors. In science-based industries such as biotechnology, developing relevant and highly-specialized scientific knowledge within the region also remains essential (see also Anderssen, 2001; Glänzel et al., 2004).

While the role of science and entrepreneurial-orientated universities and research centres is widely acknowledged for the early, incubation phase of new, science-based technologies, our study shows that in the growth phase of the biotech industry, the orientation and contribution of scientific actors in terms of technology development is positively influencing whether or not regions with more "distributed" texture characteristics evolve to become top regions. Indeed, our results show that top "distributed" regions benefit, along with an excellent science base, from a more entrepreneurial orientation of their knowledge institutes. To become a leading region, regions with a "distributed" texture also have to create sufficient industrial activities in the field of biotechnology, by generating new entrepreneurial activities in the field of biotechnology or attracting new firms in the region. Also the continuous investment in a strong science base remains important in the growth phase of science-based industries.

References

Acharya, R. 1999. *The emergence and growth of biotechnology. Experiences in industrialized and developing countries*. Edward Elgar.

Agrawal, A. & Cockburn, I. 2003. The anchor tenant hypothesis: exploring the role of large, local, R&D-intensive firms in regional innovation systems. *International Journal of Industrial Organization* 21, 1227–1253.

Andersen, B. 2001. *Technological Change and the Evolution of Corporate Innovation. The structure of Patenting* 1890–1990. Edward Elgar.

Arora, A. & Gambardella, A. 1990. Complementarity and External Linkages: The Strategies of the Large Firms in Biotechnology. *The Journal of Industrial Economics*, 38(4), 361–379.

Audretsch, D. & Feldman, M. 1996. R&D spillovers and the geography of innovation and production. *The American Economic Review* 863, 630–640.

Bartholomew, S. 1997. National systems of biotechnology innovation: complex interdependence in the global system. *Journal of International Business Studies*, vol 282, 241–266.

Breznitz, S.M., O'Shea, R.P. & Allen, T.J. 2008. University commercialization strategies in the development of regional bioclusters. *Journal of product innovation management* 25, 129–142.

Casper, S. 2007. How do technology clusters emerge and become sustainable? Social network formation and inter-firm mobility within the San Diego biotechnology cluster. *Research Policy* 36, 438–455.

Chiaroni, D. & Chiesa, V. 2006. Forms of creation of industrial clusters in biotechnology. *Technovation* 26, 1064–1076.

Cooke, P. 2001. Regional innovation systems, clusters and the knowledge economy. *Industrial and Corporate change* 104, 945–974.

Du Plessis, M.; Van Looy, B.; Song, X. & Magerman, T. 2011. Sector allocation. In: *Patent Statistics at Eurostat: Methods for Regionalisation, Sector Allocation and Name Harmonisation*. Eurostat Methodologies and Working Papers.

Feldman, M. & Florida, R. 1994. The geographic sources of innovation: Technological infrastructure and product innovation in the United States. *Annals of the Association of American Geographers* 842, 210–229.

Feldman, M. 2003. The Locational Dynamics of the US biotech industry: knowledge externalities and the anchor hypothesis. *Industry and Innovation* 10 (3), 311–328.

Galambos, L 2006. *Innovation and industry evolution: a comment. In knowledge accumulation and industry evolution*. The case of pharma-biotech. Edited by Mazzucato, M. and Dosi, G.. Cambridge University Press.

Gertler, M.S. & Vinodrai, T. 2009. Life sciences and regional innovation: One path or many? *European Planning Studies* 17 (2), 235–261.

Glänzel W., Meyer M., Schlemmer B., Du Plessis M., Thijs B., Magerman T. & Debackere K. 2004. *Domain Study "Biotechnology" – An Analysis Based On Publications and Patents*. Report Published By ECOOM. www.ecoom.be

Gray, M. and Parker, E. 1998. Industrial change and regional development: the case of the US biotechnology and pharmaceutical industries. *Environment and planning* 30 (10), pp, 1757–1774.

Klepper, S. 2010. The origin and growth of industry clusters: the making of Silicon Valley and Detroit. *Journal of Urban Economics* 67, 15–32.

Lecocq, C. & Van Looy, B. 2009. The impact of collaboration on the technological performance of regions: time invariant or driven by life cycle dynamics? An explorative investigation of European regions in the field of Biotechnology. *Scientometrics*, Vol. 80, No. 3, 847–867.

Lecocq, C.; Van Looy, B. & Vereyen, C. 2011. Regionalisation of Patent Data. In: *Patent Statistics at Eurostat: Methods for Regionalisation, Sector Allocation and Name Harmonisation*. Eurostat Methodologies and Working Papers.

Magerman, T.; Peeters, B.; Song, X.; Grouwels, J.; Callaert, J & Van Looy, B 2011. Name harmonisation. In: *Patent Statistics at Eurostat: Methods for Regionalisation, Sector Allocation and Name Harmonisation*. Eurostat Methodologies and Working Papers.

Owen-Smith, J., Riccaboni, M., Pammolli, F. & Powell W.W. 2002. A comparison of US and European university-industry relations in the life sciences. *Management Science* 48 (1), 24–43.

Powell, W.W., Koput, K.W. & Smith-Doerr, L. 1996. Interorganizational collaboration and the locus of innovation in biotechnology. *Administrative Science Quarterly*, vol 41 (1), 116–145.

Rosenberg, N. & Nelson, R.R. 1994. American universities and technical advance in industry. *Research Policy* 23 (3), pp, 323–348.

Zeller, C. 2001. Clustering biotech: a recipe for success? Spatial patterns of growth of biotechnology in Munich, Rhineland and Hamburg. *Small Business Economics* 17, 123–141.

Zucker, L.G. & Darby, M.R. 1996. Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 93 (23), 12709–12716.

Zucker, L.G. & Darby, M.R. 1997. Present at the biotechnological revolution: transformation of technological identity for a large incumbent pharmaceutical firm. *Research Policy* 26(4–5), 429–446.

Annex

Annex 1. Leading organisations per regions (period 1992–1997)

	Organisation name	**Organisation type**
1. North California, US	Genentech Inc. Incyte University of California	New Dedicated Biotech Firm New Dedicated Biotech Firm University
2. Tokyo-TO, Japan	Ajinomoto Co., Inc. Kyowa Hakko Kogyo Co., Ltd.	Established Pharmaceutical Firm Established Firm
3. Massachusetts, US	General Hospital Coporation Genetics Institute	Hospital New Dedicated Biotech Firm
4. South California, US	Amgen Gen-Probe Incorporated Scripps Research Institute	New Dedicated Biotech Firm New Dedicated Biotech Firm Research Center
5. New Jersey, US	Becton Dickinson & Co. Merck	Established Pharmaceutical Firm Established Pharmaceutical Firm
6. New York, US	Bristol Myers Squibb Co. Johnson & Johnson Ludwig Institute for Cancer Research New York University	Established Pharmaceutical Firm Established Pharmaceutical Firm Research Center University
7. Maryland, US	Department of Health and Human Services Human Genome Sciences, Inc.	Research Center New Dedicated Biotech Firm
8. Île de France, France	Institut National de la Sante et de la Recherche Medicale (INSERM) Institut Pasteur Rhone-Poulenc AG	Research Center Research Center Established Firm
9. Osaka-FU, Japan	Ono Pharmaceutical Co., Ltd. Sumitomo Electric Industries, Ltd. Suntory Limited Takeda Chemical Industries, Ltd.	Established Pharmaceutical Firm Established Firm Established Firm Established Firm
10. Pennsylvania, US	BAYER AG Smithkline Beecham University of Pennsylvania	Established Firm Established Pharmaceutical Firm University
11. Denmark	Novo Group	Established Pharmaceutical Firm

12. Inner London, UK	Cancer Research Campaign Technology Limited Medical Research Council Unilever Zeneca	Other Firm Research Center Established Firm Established Pharmaceutical Firm
13. Illinois, US	Abbott Laboratories	Established Pharmaceutical Firm
14. Karlsruhe, Germany	Roche Diagnostics	Established Pharmaceutical Firm
15. Nordwestschweiz, Switzerland	F. Hoffmann-La Roche AG Novartis	Established Pharmaceutical Firm Established Pharmaceutical Firm

i The state of California (US) was split in North and South California as the state covers 2 large and distinct biotech clusters. Three outlier regions have been removed.

ii Nuts1 level was selected for the smaller European countries (Austria, Belgium, Greece, and Ireland), while nuts2 level is used for the other countries of the EU-15 and Switzerland.

iii Information on the (headquarter) location was matched with the address information on the patent to ensure the information retrieved via web sources corresponds with the assignee (company) of the patent application.

iv Since the 1990s have been characterized by a lot of merger and acquisition activities in the field of biotechnology, but also because of the high failure rates of new (biotech) companies, we had to rely on exhaustive web searches to find company information, especially for the companies that no longer exist today, exist under a different name or have been acquired in the meantime.

Interactive Overlay Maps for US Patent (USPTO) Data Based on International Patent Classifications (IPC)[1, 2]

Loet Leydesdorff loet@leydesdorff.net / Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Kloveniersburgwal 48, 1012 CX Amsterdam (The Netherlands)

Duncan Kushnir duncan.kushnir@chalmers.se / Environmental Systems Analysis, Chalmers University of Technology, Göteborg (Sweden)

Ismael Rafols i.rafols@sussex.ac.uk / SPRU (Science and Technology Policy Research), University of Sussex, Falmer, Brighton, East Sussex BN1 9QE, (United Kingdom); Ingenio (CSIC-UPV), Universitat Politècnica de València, València (Spain)

Abstract

We report on the development of an interface to the US Patent and Trademark Office (USPTO) that allows for the mapping of patent portfolios as overlays to basemaps constructed from citation relations among all patents contained in this database during the period 1976–2011. Both the interface and the data are in the public domain; the freeware programs VOSViewer and/or Pajek can be used for the visualization. These basemaps and overlays can be generated at the 3-digit and/or 4-digit levels of the International Patent Classifications (IPC) of the World Intellectual Property Organization (WIPO). The basemaps provide a stable mental framework for analysts to follow developments over searches for different years, which can be animated. The full flexibility of the advanced search engines of USPTO are available for generating sets of patents and/or patent applications which can thus be visualized and compared. This instrument allows for addressing questions about technological distance, diversity in portfolios, and animating the developments of both technologies and technological capacities of organizations over time. Rao-Stirling diversity measures are provided on the basis of "technological distances" (Jaffe, 1989:88), that is, (1 – cosine).

Introduction

Using non-patent literature references, Narin et al. (1997) signaled a more intense and closer linkage between patenting and publishing in several fields of science and technology (Henderson et al., 2005). With increased awareness of the emergence of a knowledge-based economy, the patent system became further adapted to the publication system. With the 2011 "American Invents Act," for example, the USA brought its patenting system in line with the rest of the world by changing (as of 2013) from "first to invent" to "first to file" as the basis for granting patents.

1 The full paper is forthcoming in *Scientometrics*, and already available at doi: 10.1007/s11192-012-0923-2.

2 We acknowledge support by the ESRC project 'Mapping the Dynamics of Emergent Technologies' (RES-360-25-0076). We are grateful to Nils Newman and Antoine Schoen for previous comments and communications.

Patents can also be considered as output of the technosciences, and as indicators of input into the economy. Grilliches (1984) focused on patents as such latter indicators, and noted the different and sometimes incompatible organization of various statistics (Grilliches, 1994:14). Jaffe & Trajtenberg (2002) then used three million patents and 16 million citations in the USPTO database in a comprehensive study of what these authors called "a window on the knowledge-based economy." However, patents are indicators of invention; innovation presumes the introduction of inventions into a market. Patents are thus developed in relation to two social contexts: the sciences and markets (Klavans & Boyack, 2008).

Accordingly, patents are classified in terms of technologies and not by products or industries (Jaffe, 1986)[i]. However, patenting is an indicator of industrial activity more than of academic production (Shelton & Leydesdorff, 2012). Whereas scholarly literature is mainly organized into journals, patents are organized into patent classification systems. There are two major classification systems: that used by the US Patent and Trade Office (USPTO), and the International Patent Classifications (IPC) developed by the World Intellectual Property Organization (WIPO) in Geneva. The latter was first developed for international patenting under the Patent Cooperation Treaty (PCT) that has been signed by most countries of the world since its inception in 1970.

Like publications, patents can also be distinguished in terms of their numbers of citations, but citations in the case of patenting mean something different from citation in scholarly literature. In addition to inventor citations, the examiner can attach citations to the front page of the patent in order to ensure coverage of prior art because of the possibility to challenge patents in court (Criscuolo & Verspagen, 2008). Since 2001, the full texts of patents allows one to distinguishes between applicant and examiner citations in US patents, while the latter are asterisked on the front pages of the patents (Alcácer et al., 2009).[ii] The patenting system is thus regulated more formally than the publication system.

In this study, we map USPTO data in terms of aggregated citations among IPC classes. In an earlier attempt, Leydesdorff (2008) explored mapping WIPO data in terms of IPC co-classifications, but noted that the hierarchical structure introduced by the thesaurus made it difficult to map the co-classification structure at the aggregated level. Indexer effects are generated, for example, when classes are split (or otherwise changed) because they grow too large. Such effects can have an uncontrolled impact on co-classifications. Classifications make discrete cuts, whereas the network of citation relations can vary in density within and across clusters (Kay et al., 2012). In other words, the *citation network* among the IPC classes is heterarchical, whereas the IPC provides a hierarchical representation (with one less degree of freedom). Because of the additional degree of freedom in networks when compared with hierarchies, one can expect that the citation network is less sensitive to misclassifications than the co-classification network (Rafols et al., 2010: 1887).

Co-classifications have more often been used for measuring the "technological distances" between patenting units such as firms or nations (e.g., Breschi et al., 2003; Dolfsma & Leydes-

dorff, 2011; Jaffe, 1986, 1989). More recently, several academic groups have proposed organizing patent data in terms of aggregated citation structures among IPC classes (Kay et al., 2012; Schoen, 2011). Schoen et al. (2012) use a database derived from the comprehensive PatStat database, but selected by the Corporate Invention Board (CIB) and used by the Institute of Prospective Technology Studies of the European Commission. Furthermore, these authors construct a classification of their own. Kay et al. (2012) used EPO data from 2000–2006, but varied the number of digits in the IPC hierarchy in order to optimize the sizes of the categories for the sake of mapping. In our opinion, one can normalize for size differences among distributions using, for example, the cosine. Our approach is closest to that of Boyack and Klavans (2008:181), who developed a USPTO patent map based on co-classifications of 290 IPC classes.

Methods and materials

USPTO data were collected by one of us for the years 1976–2011 on February 12, 2012. The set contains 4,597,127 patents ranging from 75,544 patents granted in 1976 to 247,727 in 2011. These are the so-called technical patents; design patents and genetic sequences were excluded. References in technical patents to design patents or genetic sequences are also not included in this data. The approximately 39 million citation relations were organized in terms of IPC classes. The aggregated citation matrices at both the 3-digit and 4-digit level of IPC were normalized using the cosine as a similarity measure among citation distributions in different classes (Ahlgren et al., 2003). Jaffe (1986, 1989:88) defined "technological proximity" and "technological distances" in terms of the cosine measure that we also use standardly in other maps. Our purpose is to make available an interactive basemap comparable to the previously constructed basemaps for journals, journal categories, and MeSH categories, and to leave the user as much flexibility as possible for further adjustments. All necessary programs and data needed are in the public domain, including the programs VOSViewer and Pajek that are used for the visualization (De Nooy et al., 2011; Van Eck & Nees, 2010).

USPTO data has been available online in html format since 1976.[iii] At the time of this study (October 2012), 2011 was the last complete year. All valid citations from the period 1976–2011 among IPC classes at the 3-digit and 4-digit levels are used for the two respective basemaps. A routine is available online (at http://www.leydesdorff.net/ipcmaps) to assist the user in downloading sets based on specific searchstrings from the USPTO database[iv] and organizing each set as an overlay to the basemaps. As with our previous maps, we use the "citing" side of the cited/citing matrix among the classes for the analysis, because in the "cited" direction older patents may be prevalent, whereas the analyst is chiefly interested in the current state of a unit of analysis (e.g., a country or a technology) publishing and citing patents in its knowledge base. In a later state, we envisage extending the system to include the "being cited" counts, as was done – using the top-quartiles with different colors – in case of the overlays of patent statistics to Google Maps (Leydesdorff & Bornmann, 2012).

The number of IPC-classes per patent can range from one to more than twenty in exceptional cases. However, only the primary classes of patents were used for generating the basemaps. Among these patent classes, citations were counted at both 3-digit and 4-digit levels and organized in an (asymmetrical) citation matrix (that is, "citing" versus "cited"). Table 1 provides the descriptive statistics. The total number of citations in the matrix is on the order of 39 million, independently of whether this is measured at the 3-digit or 4-digit level.

Table 1. Descriptive statistics of IPC classes and their citation relations in the USPTO data (1976–2011) at three- and four-digit levels.

	3 digits		**4 digits**	
	USPTO data (including data-base mistakes)	**After correction**	**USPTO data (including data-base mistakes)**	**After correction**
IPC Classes	817	124 (129 in IPC)	4,126	630 (637 in IPC)
N of edges	28,599	13,541 (47.4%)	253,049	176,972 (72.8%)
N of citations	39,278,933	39,124,366 (99.6%)	39,286,577	38,824,390 (98.8%)

Note that the correction for misspellings of IPC in the database has a considerable effect (52.6% and 27.2%, respectively) on the number of edges, but not on the sum totals of citations. This latter error is negligible at the 3-digit level (0.4%) and still relatively small (1.2%) at the 4-digit level. As noted above, errors in the classifications can often be considered as disturbances in terms of the bibliographic links at the network level. Furthermore, the IPC contains several classes such as "C99" (at the 3-digit level) which are labeled as "subject matter not otherwise provided for in this section," but these classes are not used by USPTO. Thus, we use 124 of the 129 available IPC at the 3-digit level, and 630 or the 637 at the 4-digit level. In sum, we analyze the data listed in the right columns of 3-digit and 4-digit data in Table 1, after deletion of a large number of misspelled classifications.

The two citation matrices (for the 3- and 4-digit levels; available from the website) are input into SPSS v.20, so that cosine-normalized matrices can then be computed and exported. The cosine is a non-parametric similarity measure; it has the advantage of being insensitive to the large numbers of zeros in the vectors of sparse matrices. The cosine can also be considered as a non-parametrized Pearson correlation coefficient. Jaffe (1989:89; cf. Jaffe, 1986:986) proposed using the cosine to measure technological distances, but he used co-classifications. As noted, cosine values were computed from the citing side of the matrices because "citing" is the active variable, whereas in the "cited" dimension the complete archive can be represented.

In previous mappings, we generated basemaps from cosine matrices using the spring-embedded algorithm of Kamada & Kawai (1989) in Pajek. Since cosine values tend to be larger than zero in (almost) all cells of the matrix, a threshold had to be set in order to make a grouping of the categories visible; otherwise, almost all vertices are connected by a single (largest) component. Using the MDS-like solution of the program VOSViewer, however, a threshold is not needed, since the algorithm uses all quantitative information for the mapping (Van Nees et al., 2010). Leydesdorff & Rafols (2012) noted that further normalization within VOSViewer did not disturb the maps on the basis of cosine-normalized matrices.

Because of the additionally available network statistics in Pajek (and similar programs for network analysis), we shall extend the options by providing base maps and overlay files in the Pajek-format in a later section. Other programs such as UCINet and Gephi can read Pajek files. VOSViewer and Gephi have solved the problem of the cluttering of labels in the visualization (Leydesdorff et al., 2011); Pajek and VOSViewer are most convenient for making overlay maps, but differ in terms of the possible visualizations and clustering algorithms available (Leydesdorff & Rafols, 2012). For the sole purpose of visualization, VOSViewer is probably the first option.

VOSViewer provides its own clustering algorithm based on modularity optimization (Blondel et al., 2008; Newman, 2004; Waltman et al., 2010). We use this algorithm as the default. Both in the 3- and 4-digit case, the results are conveniently organized in five main clusters and therefore colors; all IPC classes (124 and 630, respectively) are included in both cases. The user is able to replace the clustering and colors (see Leydesdorff & Rafols [2012] for more detailed instruction). For example, one may wish to color the maps according to the eight top-level categories of IPC (A to H). The text files that contain the mapping information (available at the website) can be edited. VOSViewer (at http://vosviewer.com) can be either web-started or downloaded from the internet and used locally.

The labels at the 4-digit level are sometimes long, and this may affect the readability of the maps. VOSViewer uses 30 characters as a default, but this can be adapted within the program interactively. One can also change the color selection interactively or turn off the prevention of blurring of the labels that is provided as a default. In the case of long labels, we follow the common practice of using the IPC (sub)headings by cutting off at a maximum of 75 characters, with three dots after the right-most space in the string. The user may edit the files differently if so wished. A default of 30 characters works without problems in the 3-digit case, but may require some editing in the 4-digit map. Note that the length of the strings can affect the visibility of individual labels because the program optimizes readability.

The maps provide a representation of distances between categories and the diversity among categories in each specifically downloaded set of patents. In addition to the distance on the visible map, $\|x_i - x_j\|$, the (technological) distance in the data between each two classes can also be denoted analytically as $d_{ij} = (1 - cosine_{ij})$. The mapping program projects the multivariate space – spanned by the citing patents as vectors of the matrix – into the two dimensions of the map by minimizing a function such as Kruskall's (1964) stress.[v]

The technological distances between categories (d_{ij}) can also be summed and normalized in accordance with the weight of the respective categories using their proportions in the distribution. This leads us to the Rao-Stirling measure of diversity (Δ) that is sometimes also called "quadratic entropy" (e.g., Izsák & Papp, 1995; cf. Rao, 1982):

$$\Delta = \sum_{ij} p_i p_j d_{ij} \qquad (1)$$

Stirling (2007:712) proposed this measure as a summary statistics for analyzing diversity and interdisciplinarity in studies of science, technology, and society (Leydesdorff & Rafols, 2011a; Rafols and Meyer, 2010). The routine writes the values of this parameter locally to a text-file ("ipc_rao.txt") at both the 3-digit and 4-digit level.

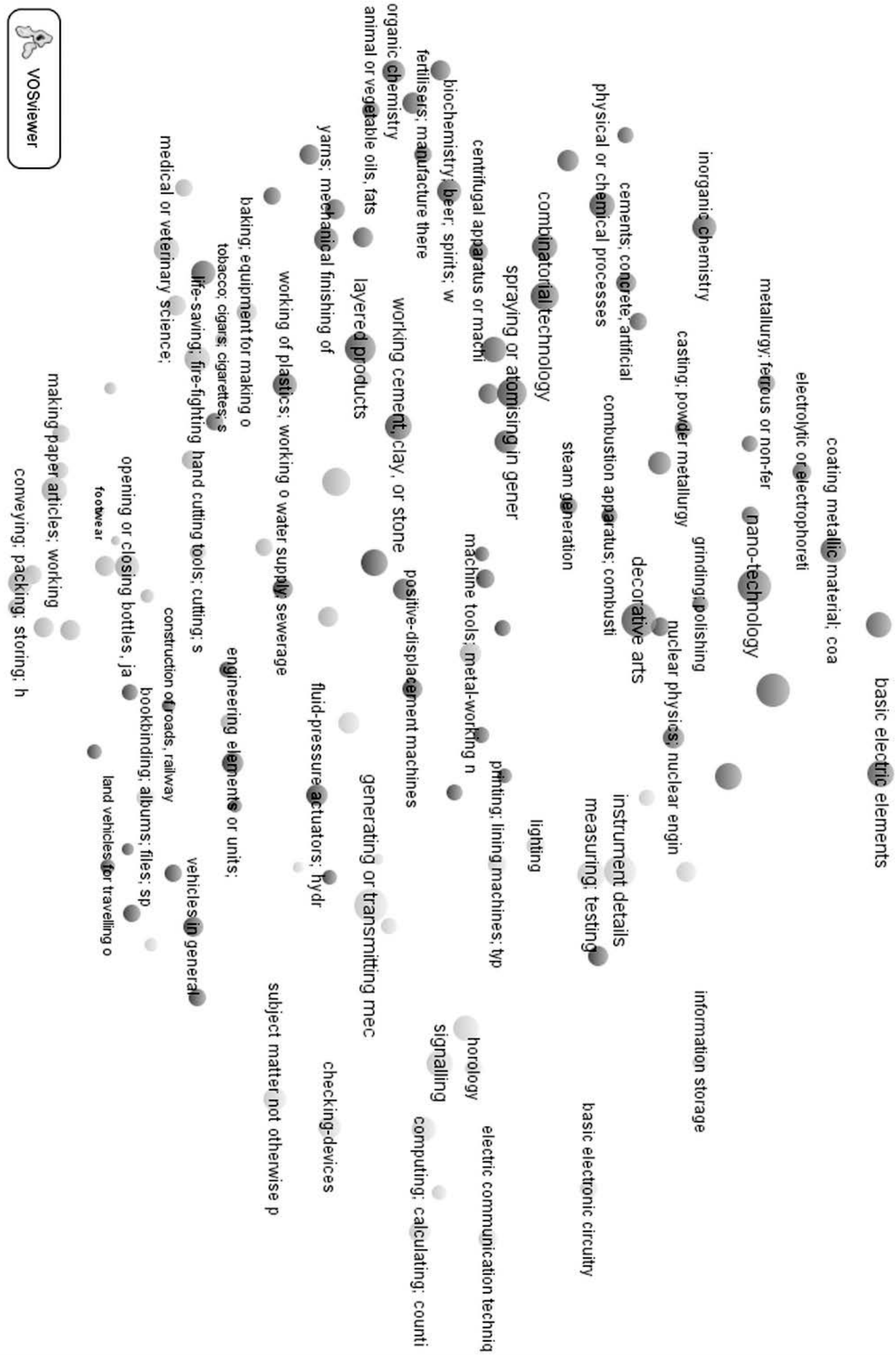
In summary, we operationalize the technological distance between patent classes as the complement of the cosine between the distributions of citations, and are then able to compute diversity in terms of Equation 1. The cosine matrices for both 3-level and 4-level IPC are available at the webpage, and if these files are present, the routine writes in each run a file "ipc_rao.txt" that contains the diversity measures given the two matrices at the respective digit-levels. Note that there is no hierarchical relation between the two maps because a patent may contain a single category at the 3-digit level, under which many different 4-digit categories can be subsumed that co-vary with other 4-digit categories at this level. In the case of the overlays, IPC classes will be used proportionally, that is, as a proportion of the total number of IPC classes attributed to each patent, so that each patent contributes a sum value of unity to the overlays. Both the numbers of patents and the fractional counts are made available in a file "vos.dbf" at the occasion of each run. Note that this file is overwritten by a next run!

Results

Basemaps at the three- and four-digit level

The generation of the basemaps from the cosine matrices is straightforward. Figure 1 shows the basemap of 124 IPC classes at the 3-digit level, and Figure 2 (not displayed here, but to be web-started interactively at http://www.vosviewer.com/vosviewer.php?map=http://www.leydesdorff.net/ipcmaps/ipc4.txt) shows 630 classes at the 4-digit level. In the latter case we used the heat or density map as an alternative representation.

Figure 1. Basemap of 124 IPC categories at the 3-digit level, using approximately 39 million citations among 4.2M US patents (1976–2011); cosine-normalized "citing"; VOSViewer used for the visualization. This map can be accessed interactively at http://www.vosviewer.com/vosviewer.php?map=http://www.leydesdorff.net/ipcmaps/ipc3.txt .



The 3-digit labels seem sufficiently precise for high-level analysis (e.g. policy analysis of countries), whereas 4-digit labels may be needed for specific technology analysis. As might be expected, the resolution among the labels at the 3-digit level is higher than at the 4-digit level. Note that the 3-digit and 4-digit maps exhibit the same structural dimensions but flipped both horizontally and vertically; each solution may freely be rotated and translated in the MDS domain.

Overlays to the base maps

One can generate the input for overlays by using, for example, the interface of USPTO Advanced Searching for (granted!) patents at http://patft.uspto.gov/netahtml/PTO/search-adv.htm . First, the search must be formulated so that a recall of more than fifty patents is generated. By clicking on "Next 50" one finds the information used by the dedicated programs ipc.exe and uspto1.exe (available at http://www.leydesdorff.net/ipcmaps) to automate further searches, and to organize input files for VOSViewer and Pajek thereafter. Any search string compatible with the USPTO database will do the job. We provide examples below.

In patents with a sequence number in the retrieval higher than 50, one copies (Control-C) the complete string in the browser after opening this patent (for an example and instruction see also at http://www.leydesdorff.net/ipcmaps). All information about the search is embedded in this string and can be used by the program uspto1.exe to download the set. The routine uspto1.exe is optionally called by ipc.exe.

The patents are downloaded and further organized in the same folder of the disk, and two output files (named "vos3.txt" and "vos4.txt," respectively) can be used directly by VOSViewer as input for generating overlays. In addition to "ipc.exe", the user should first download also the files "ipc.dbf" and "uspto1.exe" from the same website into the same folder because these files are also required. If the files "cos_ipc3.dbf" and "cos_ipc4.dbf" are also downloaded from the website, the routine writes a local file "ipc_rao.txt" containing the Rao-Stirling diversity values for three and four digits, respectively.

The program ipc.exe opens with the option to download patents in this run or use patents downloaded in a previous run. Note that patents from a previous run are overwritten in the case of a new download; one should save them elsewhere for future use. The USPTO limits the number of downloads to one thousand, but one can begin subsequent downloads at 1001, 2001, etc. The program accepts subsequent numberings. In this case, one should use uspto1.exe directly, because ipc.exe overwrites results from previous runs.

In the output of the routine ("vos3.txt" and "vos4.txt") the IPC classes assigned by the USPTO to patents in the download are organized into the map files of VOSViewer on the basis of counting each IPC attribution in proportion to the total number of classes attributed to the patent at the level in question. In the "label view" of VOSViewer, the empty classes are made visible as little grey dots for the orientation of the user, but these classes are not labeled. The classes in use are

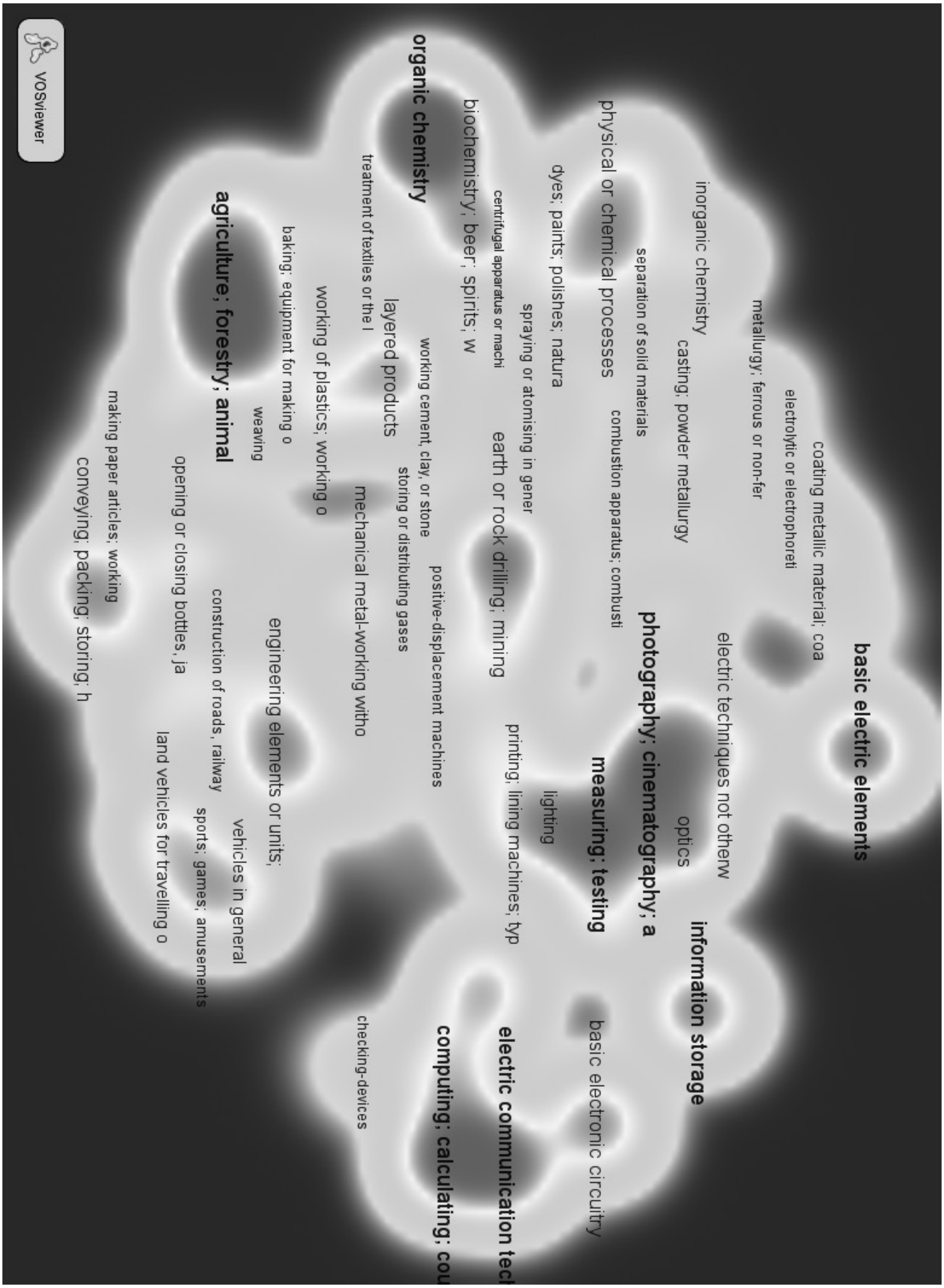
normalized proportionally to the logarithm of the number +1. The "+1" prevents single occurrences or fractions smaller than unity from disappearing.[vi]

Figure 3 shows the density map of all patents granted in 2007 with a Dutch address for the inventor. This same data was used for an overlay to Google Maps by Leydesdorff & Bornmann (2012, at pp. 1446 ff.). One can edit the input files or use the interactive facilities of VOSViewer to further enhance the representation.

VOSViewer shows the labels of the most prominent classes in the sets under study. (One can also turn this off.) As noted, the user can modify the clustering classification; our labels correspond to the labels in the (most recent) IPC 2012 version. The quantitative information about the results of each run are stored both at the 3-digit and 4-digit levels in a file "vos.dbf" for further (e.g., statistical) analysis.

Figure 3. Heat map of 1,908 US patents with an inventor in the Netherlands, and publication date in 2007 at the 3-digit level; 3018 IPC classes; Rao-Stirling diversity $\Delta = 0.869$.

This map can be accessed interactively at http://www.vosviewer.com/vosviewer.php?map=http://www.leydesdorff.net/ipcmaps/vos3.txt&view=2



The USPTO interface offers a host of possibilities for combining search terms such as inventor names and addresses, applicants, titles, abstract words, issue dates, etc., that can be combined with AND and OR operators. Furthermore, a similar interface is available at http://appft1.uspto.gov/netahtml/PTO/search-adv.html for searching patent *applications*. For example, 3,898 applications with a Dutch inventor address can be retrieved by using "icn/nl and apd/2007$$" as a search string in this database. (Unlike "isd" for "issue data" above, "apd" is the field tag for "application date".)

This second routine named "appl_ipc.exe" operates similarly to "ipc.exe", but parses downloads from this interface of patent *applications*. (The format is somewhat different from granted patents.) As above, any webpage-address can be harvested at the 51st patent application, for example, and used for the download (which is strictly similar to the one using ipc.exe). One analytical advantage of using applications is that applications follow the research front, whereas the process of examining and granting a patent can take several years.

By mapping the same searches for different years, for example, and copying the results into a program such as PowerPoint, one can animate the visualizations. The positions will be stable, but the size of the nodes and the most prominent labels will change. We use the option "normalized weights" as default in vos3.txt and vos4.txt so that the user can compare results across years and otherwise. However, for esthetic optimization by VOSViewer, one may wish to turn this option into "weight" in a single-case study. As noted, the colors and sizes of nodes that one wishes to highlight can be adapted by editing these lines in the input files to VOSViewer. Thus, new developments can be made visible in animations. For example, one can add exclamation marks to the labels.

Baseline maps in Pajek format

In our opinion, Gephi and VOSViewer offer superior visualization techniques (Leydesdorff et al., 2011), but Gephi and Pajek/UCINet provide network statistics. The comparison made us realize that with little effort we could also make our outputs compatible with Pajek, and via Pajek also for Gephi (which reads Pajek files). This offers additional flexibilities such as using algorithms for community detection among a host of other network statistics which are available in Pajek and Gephi, but not in VOSViewer.

Baseline maps for Pajek in the 3-digit and 4-digit format can be retrieved at http://www.leydesdorff.net/ipcmaps/ipc3.paj and http://www.leydesdorff.net/ipcmaps/ipc4.paj, respectively. The available baselines were in this case partitioned – for didactic purposes – using Blondel et al.'s (2008) algorithm for community finding. Without a threshold for the cosine, the visualization is not informative. Using a threshold of cosine > 0.2, the largest components contain 109 of the 124 classes at the 3-digit level (11 communities; $Q = 0.529$), and 605 of the 630 classes at the 4-digit level (21 communities; $Q = 0.681$). However, the experienced user should change these base maps and their partitioning or coloring using the options available in Pajek (or Gephi).[vii]

Both routines (ipc.exe and appl_ipc.exe) provide two vector files (ipc3.vec and ipc4.vec, respectively) with the same values – "weights" in VOSViewer – as used above; that is, the fractionally counted patents/IPC class. These vector files can be used for the visualization of the patent contributions to IPC classes in the downloaded sample(s) under study. Additionally, the so-called cluster files ipc3.cls and ipc4.cls (in the Pajek format) enable the user to label the sample under study exclusively in the Draw screen under Options > Mark Vertices Using > Mark Cluster only.

Conclusion

By providing basemaps for USPTO data in terms of the IPC and flexible tools for overlaying patent classifications on top of them both at the coarse-grained level of three digits and the fine-grained level of four digits, we have complemented our series of studies and tool development efforts to follow new development more systematically across databases. Grilliches (1994:14) already noted that the core problem in studying innovations has been the silos of data that are constructed for institutional reasons and then developed longitudinally over the years without sufficient cross-connections. For the study of innovations, one needs to map transversal translations from one context into another.

One common baseline among the different databases is provided by the institutional address that can be overlaid on Google Maps in order to show geographical diffusion (Leydesdorff & Rafols, 2011b). Additionally, one would like to move from one context to another in terms of the classifications and codifications that are proper to each database in cognitive terms. In the case of scholarly publishing, the prime unit of organization appears to have been the scientific journal. Using Web-of-Science or Scopus, citation relations can be accessed directly. In dedicated databases such as Medline for the bio-medical sciences, specific interfaces are needed to relate citation information to classifications (Leydesdorff & Opthof, 2013).

The international patent classification IPC provides a means to organize patents intellectually, with a trade-off between technological refinement and user-friendliness. The IPC provides a baseline in all major patent systems such as WIPO, USPTO, and EPO. This paper has described the possibilities for using the fields in the USPTO databases for mapping and animation purposes. As an "instrumentality" (Price, 1984), the exploitation of this interface enables us to address central questions of patent analysis, such as those formulated in a study of technological competencies (Patel & Pavitt, 1997, at p. 141) concisely, as follows:

(1) They [large firms] are typically *multi-field*, and becoming more so over time, with competencies ranging beyond their product range, in technical fields outside their 'distinctive core'.
(2) They are *highly stable* and *differentiated*, with both the *technology profile* and the *directions* of localised search strongly influenced by firms' *principal products*.
(3) The rate of search is influenced by both the firm's *principal products*, and the conditions in its *home country*. However, *considerable unexplained variance* suggests scope for managerial choice.

These conclusions identify a research program. The instruments provided here offer tools for addressing such a program of studies in quantitative terms (e.g., in terms of Rao-Stirling diversity) and for illustrating the results with animations of, for example, diffusion and diversification processes. Nowadays, the Internet enables us to upscale and use "big data" for performing these studies of science, technology, and innovation (Helbing & Balietti, 2011. In our opinion, the development of interfaces to access different databases ("big data") with flexibility, but with similar or equal search strings provides a strategy which may enable us to follow new developments in science and technology along trajectories and potentially developing into regimes (Leydesdorff et al., in press).

In most previous studies using OECD data, for example, analysis remained at the aggregated level and static analysis consequently prevailed (e.g., Jaffe, 1989; Patel & Pavit, 1997). These new instruments enable us to study individual firms, nations, and new technologies in considerable detail and dynamically by following the available retrieval options and tracing the various classifications at USPTO. In a follow-up, we envisage studying, for example, the new classes of "nanotechnology" developed in the IPC and available as B82$ in USPTO,[viii] and/or differently in the US Classification System as 977.[ix] Such a study would allow one to follow the changing position of "nano-patents" in terms of IPC classifications over time.

References

Ahlgren, P., Jarneving, B., & Rousseau, R. (2003). Requirements for a Cocitation Similarity Measure, with Special Reference to Pearson's Correlation Coefficient. *Journal of the American Society for Information Science and Technology*, 54(6), 550–560.

Alcácer, J., Gittelman, M., & Sampat, B. (2009). Applicant and examiner citations in U.S. patents: An overview and analysis. *Research Policy*, 38(2), 415–427.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 8(10), 10008.

Boyack, K. W., & Klavans, R. (2008). Measuring science–technology interaction using rare inventor–author names. *Journal of Informetrics*, 2(3), 173–182.

Breschi, S., Lissoni, F., & Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. *Research Policy*, 32(1), 69–87.

Criscuolo, P., & Verspagen, B. (2008). Does it matter where patent citations come from? Inventor vs. examiner citations in European patents. *Research Policy*, 37(10), 1892–1908.

de Nooy, W., Mrvar, A., & Batgelj, V. (2011). *Exploratory Social Network Analysis with Pajek (2nd Edition)*. New York, NY: Cambridge University Press.

Dolfsma, W., & Leydesdorff, L. (2011). Innovation systems as patent networks: The Netherlands, India and nanotech. *Innovation Management, Policy & Practice*, 13(3), 311–326.

Helbing, D., & Balietti, S. (2011). How to create an Innovation Accelerator. *The European Physical Journal-Special Topics*, 195(1), 101–136.

Henderson, R., Jaffe, A., & Trajtenberg, M. (2005). Patent citations and the geography of knowledge spillovers: A reassessment: Comment. *The American Economic Review*, 95(1), 461–464.

Izsák, J., & Papp, L. (1995). Application of the quadratic entropy indices for diversity studies of drosophilid assemblages. *Environmental and Ecological Statistics*, 2(3), 213–224.

Jaffe, A. B. (1986). Technological Opportunity and Spillovers of R&D: Evidence from Firm's Patents, Profits, and Market Value. *American Economic Review*, 76(5), 984–1001.

Jaffe, A. B. (1989). Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers. *Research Policy*, 18(2), 87–97.

Jaffe, A. B., & Trajtenberg, M. (2002). *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. Cambridge, MA/London: MIT Press.

Kay, L., Newman, N., Youtie, J., Porter, A. L., & Rafols, I. (2012). Patent Overlay Mapping: Visualizing Technological Distance. *Arxiv preprint arXiv*:1208.4380.

Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1), 7–15.

Klavans, R., & Boyack, K. (2009). Towards a Consensus Map of Science. *Journal of the American Society for Information Science and Technology*, 60(3), 455–476.

Leydesdorff, L. (2008). Patent Classifications as Indicators of Intellectual Organization. *Journal of the American Society for Information Science & Technology*, 59(10), 1582–1597.

Leydesdorff, L., & Bornmann, L. (2012). Mapping (USPTO) Patent Data using Overlays to Google Maps. *Journal of the American Society for Information Science and Technology*, 63(7), 1442–1458.

Leydesdorff, L., & Meyer, M. (2010). The Decline of University Patenting and the End of the Bayh-Dole Effect. *Scientometrics*, 83(2), 355–362.

Leydesdorff, L., & Opthof, T. (2013). Citation Analysis using the Medline Database at the Web of Knowledge: Searching "Times Cited" with Medical Subject Headings (MeSH). *Journal of the American Society for Information Science and Technology*, 64(5), 1076–1080.

Leydesdorff, L., & Rafols, I. (2011a). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics*, 5(1), 87–100.

Leydesdorff, L., & Rafols, I. (2011b). How Do Emerging Technologies Conquer the World? An Exploration of Patterns of Diffusion and Network Formation. *Journal of the American Society for Information Science and Technology*, 62(5), 846–860.

Leydesdorff, L., & Rafols, I. (2012). Interactive Overlays: A New Method for Generating Global Journal Maps from Web-of-Science Data. *Journal of Informetrics*, 6(3), 318–332.

Leydesdorff, L., Rotolo, D., & de Nooy, W. (in press). Innovation as a Nonlinear Process, the Scientometric Perspective, and the Specification of an "Innovation Opportunities Explorer". *Technology Analysis & Strategic Management*; available at http://arxiv.org/abs/1202.6235.

Leydesdorff, L., Rotolo, D., & Rafols, I. (2012). Bibliometric Perspectives on Medical Innovation using the Medical Subject Headings (MeSH) of PubMed. *Journal of the American Society for Information Science and Technology*, 63(11), 2239–2253.

Narin, F., Hamilton, K. S., & Olivastro, D. (1997). The increasing link between U.S. technology and public science. *Research Policy*, 26(3), 317–330.

Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6), 066133.

Patel, P., & Pavitt, K. (1997). The technological competencies of the world's largest firms: complex and path-dependent, but not much variety. *Research Policy*, 26(2), 141–156.

Price, D.J. de Solla (1984). The science/technology relationship, the craft of experimental science, and policy for the improvement of high technology innovation *Research Policy*, 13, 3–20.

Rafols, I., & Meyer, M. (2010). Diversity and Network Coherence as Indicators of Interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, 82(2), 263–287.

Rafols, I., Porter, A., & Leydesdorff, L. (2010). Science overlay maps: a new tool for research policy and library management. *Journal of the American Society for Information Science and Technology*, 61(9), 1871–1887.

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management. *Research Policy*, 41(7), 1262–1282.

Schmoch, U., Laville, F., Patel, P., & Frietsch, R. (2003). Linking Technology Areas to Industrial Sectors. *Final Report to the European Commission, DG Research*(available at ftp://ftp.cordis.europa.eu/pub/indicators/docs/ind_report_isi_ost_spru.pdf).

Schoen, A. (2011). *A Global Map of Technology*. Paper presented at the IPTS Patent Workshop, Seville, Spain, June 13–14.

Schoen, A., Villard, L., Laurens, P., Cointet, J.-P., Heimeriks, G., & Alkemade, F. (2012). *The Network Structure of Technological Developments; Technological Distance as a Walk on the Technology Map*. Paper presented at the Science & Technology Inidicators (STI) 2012 Montreal.

Shelton, R. D., & Leydesdorff, L. (2012). Publish or Patent: Bibliometric evidence for empirical trade-offs in national funding strategies *Journal of the American Society for Information Science and Technology*, 63(3), 498–511.

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, 4(15), 707–719.

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538.

van Eck, N. J., Waltman, L., Dekker, R., & van den Berg, J. (2010). A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology*, 61(12), 2405–2416.

Waltman, L., van Eck, N. J., & Noyons, E. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629–635.

---

i Schmoch et al. (2003) provides a concordance table between International Patent Classifications (IPC) and industrial classifications with the NACE codes ("Nomenclature statistique des activités économiques dans la Communauté européenne") used by the OECD.

ii The program uspto1.exe can be used for downloading the citing patents using the routines available at http://www.leydesdorff.net/indicators/lesson5.htm (Leydesdorff, 2004).

iii EPO data are available online, but there are limitations to the searches (above 500 documents) and some of this data is in the pdf format. PCT patents of WIPO are available online (Leydesdorff, 2008), but as noted of lower technological and market quality than USPTO patents (Shelton & Leydesdorff, 2012).

iv The two databases can be found at http://patft.uspto.gov/netahtml/PTO/search-adv.htm for granted patents, and http://appft1.uspto.gov/netahtml/PTO/search-adv.html for patent applications, respectively.

v Kruskall's formula (1964) is expressed as follows:

$$S = \sqrt{\frac{\sum_{i \neq j} (\|x_i - x_j\| - d_{ij})^2}{\sum_{i \neq j} d_{ij}^2}}$$

In addition to summing the differences between the visually available distances and the algorithmic distances, this stress value is normalized at the level of the system of distances. Regrettably, VOSViewer does hitherto not provide a stress value for the (deterministic) visualization.

vi This addition of unity is needed because log(1) is zero. Fractional counts may be smaller than unity.

vii The full cosine matrices are available at http://www.leydesdorff.net/ipcmaps/cos_ipc3.dbf and http://www.leydesdorff.net/ipcmaps/cos_ipc4.dbf for those users who wish to be able to work with cosine values lower than the current threshold of 0.2.

viii Since April 2009, class B82 in IPC replaces the previous class Y01 in ECLA.

ix A search with "icl/B82$" provided a recall of 344 patents on September 25, 2012, whereas a search with "ccl/977$)" provided a recall of 8,134 patents. Of these two sets 249 overlap (using an AND statement).

Societal Use of Research (SURe): stakeholder perceptions and data method development

Ingeborg Meijer — i.meijer@cwts.leidenuniv.nl / CWTS, Leiden University, P.O. Box 905, Leiden, 2300 AX (The Netherlands)

Clara Calero-Medina — clara@cwts.leidenuniv.nl / CWTS, Leiden University, P.O. Box 905, Leiden, 2300 AX (The Netherlands)

Rebecca Abma-Schouten — r.abma-schouten@hartstichting.nl / Dutch Heart Foundation, Prinses Catharina Amaliastraat 10, Den Haag, 2496 XD (The Netherlands)

Abstract

From the perspective of our knowledge-based society, it is felt that the focus of research not only should be on scientific excellence, but also on the societal benefit of research. In this work-in-progress report, the CWTS aims to contribute to the scientific-scholarly foundation of societal research-assessment methodologies in science. We define societal quality of research as the value that is created by connecting research to societal practice, and it is based on the notion that knowledge exchange between research and its related professional, public and economic domain strengthens innovation. Methodology is based upon a case study approach, and is developing along two lines of research:

(1) To explore communication and reporting of science with Web based tracking tools. In a case study for an international policy organisation, we investigate other ways of reporting and the channels of communication by researchers, as well as the stakeholders for a set of 'grey literature' reports.

(2) To explore the conditions (criteria and method) to allow societal stakeholders to stimulate societal use of research. For a charity foundation in healthcare, we investigate the role of societal criteria and the further development of the criteria by an end-users panel in and early phase of research: assessing research proposals.

Introduction

Questions regarding the socio-economic and cultural relevance of scientific research have been on the science policy agenda for decades. In recent years, discussion increased about the knowledge economy, the concepts of 'valorisation', the 'knowledge paradox' and societal challenges. From the perspective of our knowledge-based society, it is felt that the focus of research not only should be on scientific excellence, but also on the societal benefit of research. Policy makers stress the importance of the added value of science, which is different for different stakeholders and

therefore science should be seen in context. Thus, evaluation of research should focus on societal (social, cultural and economic) elements as well (LERU, 2012). In many research communities and policy institutions there is an apparent need to develop new evaluation methods and indicators that do more justice to the variety of goals and activities of researchers (Piwowar, 2013). In this work-in-progress report, the CWTS aims to contribute to the scientific-scholarly foundation of societal research-assessment methodologies in science.

Conceptual framework

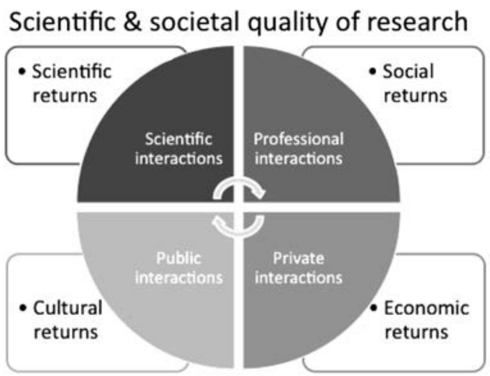
We define societal quality of research as the value that is created by connecting research to societal practice, and it is based on the notion that knowledge exchange between research and its related professional, public and economic domain strengthens innovation (see Figure 1.). It requires researchers to interact with stakeholders in these different societal domains (Mostert et al. 2010; Spaapen & van Drooge, 2011; Meijer, 2012). Interaction depends on 'sending' (by scientists) and 'receiving' (by stakeholders), and vice versa. Communicating about science requires a different approach for each of the stakeholder domains; likewise, each stakeholder has different perceptions on the societal relevance of science. Our method builds upon investigating the sending and receiving in more detail.

Methods and data

Methodology is based upon a case study approach, and is developing along two lines of research:

1) To explore communication and reporting of science with Web based tracking tools
2) To explore the conditions (criteria and method) to allow societal stakeholders to stimulate societal use of research.

Figure 1. Conceptual framework



Ad 1) When the results of science are to be received by stakeholders other than their scientific peers, plain publications may not be the first line of reporting and the scientific journals may not be used as the first channel of communication. Instead other means of reporting and communication could be used, e.g. based on social media. Tools like Mendeley (http://www.mendeley.com/), Impact Story (http://impactstory.org/) and Social Mention (http://socialmention.com/) aim to track these types of communication. In a case study for an international policy organisation, we investigate these other ways of reporting and the channels of communication by researchers, as well as the stakeholders for a set of 'grey literature' reports. This is necessary because there is a lack of data on other types of scientific output (than journal publications).

Ad 2) The societal relevance of funded research depends – in part – on the (fast) transfer of knowledge and results to potential users. For a charity foundation in healthcare, we investigate the role of societal criteria and the further development of the criteria by an end-users panel in and early phase of research: assessing research proposals. The goal is to define criteria that are suitable to review the societal relevance of research proposals. And the end-users panel is one strategy to involve users and other stakeholders from the start of research, anticipating that research can be more societal relevant when it is planned, executed and evaluated with the involvement of users and other stakeholders (te Kulve & Rip, 2011).

Results so far

In this section the results of sending and receiving are described in more detail:

- To have more insight into the relevance of research communication outside the scientific domain via reports (grey literature), we explored the usefulness of new type of tracking tools, such as Impact Story and Mendeley with the aim of detecting other type of audiences in socio-economic (professionals or private organisation) or cultural (lay public) domains. Impact Story primarily detects Facebook shares and Tweets, whereas with Mendeley professional readership can be traced (albeit overlapping with scientific use). Even though the penetration of this type of outlets is not very widespread as yet, it is possible to detect different types of visibility profiles for different types of reports, depending on the amount of counts they catch with the different tracking tools. The value of this type of altmetrics will be explored by investigating the stakeholders in more detail as well as the mutual relations (Priem, Piwowar & Hemminger, 2011).

- To have more insight in the perceptions of stakeholders on the societal relevance of research, we investigate the role of end users and a new set of criteria to enhance societal use in the ex ante evaluation of cardiovascular research proposals of the Dutch Heart Foundation with the aim of improving healthcare solutions in cardiovascular disease. The end-user panel was

carefully put together and includes medical specialists (in cardiovascular medicine, but NOT in research), general practitioners, nurses, nurse practitioners, patients having a cardiovascular condition and lay people, so as to take into account as many different social and medical perspectives. In order to keep the balance in the panel the patients involved were relatively highly educated. The end user panel started with a predefined set of criteria to review research proposals in a pilot study (in parallel to the scientific review process), which were further developed by the end-user panel. The societal criteria contain three groups of issues: one on the relevance of the problem and the potential result of the approach choses, one on the actuals activities and interactions with stakeholders to ensure 'the next step', and one on other potential hindrances and risks (mainly ELSI related). The formal procedure also includes an interaction with the principal investigators. The subsequent steps were independently observed to investigate internal values of end users with regard to societal relevance, interactions between end-users to come to consent, and usefulness of the criteria. The observations are complemented with the results of individual interviews before and after the pilot project. The usability of the societal criteria and the organisation of the process will be evaluated.

In conclusion, we acknowledge that the complex and intricate societal interactions and results are both long term and difficult to attribute to specific scientific research. Therefore we connect them in an early phase so that adequate activities can be set up to ensure the 'next step'. And by engaging in altmetrics types of analysis in order to detect different audiences for different types of scientific communication, we anticipate that this will contribute to building up alternative datasets. Both could be suitable ways to overcome the criticism that societal relevance (and impact) is more often postulated than demonstrated (Bornmann 2012).

References

Bornmann, L. (2012). *Measuring the societal impact of research*. EMBO reports 13(8): 673–676.

Meijer, I. (2012). *Societal returns of scientific research. How can we measure it?* Working paper (CWTS-WP-2012-014) presented to "New perspectives on enduring research questions in university-society interaction?" INTERACT-UNI EU-SPRI Conference, University of Twente, the Netherlands, 9th–11th May 2012.

Mostert, S., Ellenbroek, S.P., Meijer, I., van Ark, G. and Klasen, E.C. (2010). *Societal output and use of research performed by health research groups*. Health Research and Policy Systems 8:30.

Phillips, M., & Maes, K. (2012). *Research Universities and research assessment. League of European Research Universities* (LERU), May 2012.

Piwowar, H. (2013). *Value all research products*. Nature 493: 159.

Priem, J., Piwowar, H. A., & Hemminger, B. H. (2011). *Altmetrics in the wild : An exploratory study of impact metrics based on social media*. Metrics: Symposium on Informetric and Scientometric Research; 2011. New Orleans, LA, USA. Retrieved from http://jasonpriem.org/self-archived/PLoS-altmetrics-sigmetrics11-abstract.pdf

Te Kulve, H., & Rip, A. (2011). *Constructing Productive Engagement: Pre-engagement Tools for Emerging Technologies*. Sci. Eng. Ethics 17: 699–714.

Spaapen, J., & van Drooge, L. (2011). *Introducing 'productive interactions' in social impact assessment*. Research Evaluation 20 (3): 211–218.

Weighting of indicators[1, 2]

Ülle Must ulle.must@etag.ee / Department of International Research Cooperation, Estonian Research Council, Soola 8, Tartu 51013 (Estonia)

Oskar Otsus oskar.otsus@etag.ee / Department of International Research Cooperation, Estonian Research Council, Soola 8, Tartu, 51013 (Estonia)

Introduction

The quality of research output is measured with different S&T indicators. They measure the various components of research activity, including inputs, process, outputs, outcomes and impact and benefits. Some research assessments assign different weightings or values to the various indicators. In this way, some components of research activity are valued more highly than other activities. There is the eternal question of whether the indicators really measure what they are intended to measure because people adjust to the indicator value system by optimizing their indicator rather than their performance. Thus, no matter how sophisticated an indicator may be, it will never be proof against manipulation (Retzer and Jurasinski, 2009).

In the current paper we focused mainly on the questionnaire survey results which were conducted with the aim of receiving more insight into researchers' attitudes towards the ways in which quality, success, excellence and impact of scientific production are measured and evaluated.

Data and methods

A structured web based (Lime Survey) questionnaire was prepared for the survey. The survey was divided into four blocks, one of them was directed to S&T indicators (the rating of the most appropriate indicators in assessing the work of researchers; should the various indicators be weighed differently; is there a need to have different weightings for the various indicators in different subject areas; is there a need to have various indicators or different weighting for the various indicators at different career stages).
The survey was open for two months (November 2011–January 2012). The information and call to participate in the survey was sent out via different communication channels, in total to 4,280 persons, 2,114 of whom responded.

Respondents belong to 79 countries by nationality; their places of affiliation were in 66 countries. Among respondents, 1,337 were men, and 652 were women, 125 did not specify their gender. By

1 This work was supported by the FP7 project ACUMEN (http://research-acumen.eu/).
2 The current paper is supported by the FP7 project ACUMEN (http://research-acumen.eu/).

academic position, postdoctoral research fellows (643) and full professors (422) constituted the largest share of respondents, followed by lecturers/assistant professors (325), associated professors (323). The largest numbers of respondents belong to natural sciences (1096).

Results

In the survey we asked respondents to assess on a five-point scale the different indicator used in research assessments. In the selection of S&T indicators the proposed indicators of the Expert Group on Assessment of University based Research (Mustajoki. et al, 2010) were used. On an average, the highest rating was given to the following indicators: high ranked publications (4.5), citations (3.9), research collaborations and partnership (3.8), reputation and esteem – position as journal editors, membership of editorial boards and scientific committees and membership in learned academies (3.6), and number of prestigious national and international awards and prizes (3.6). In ratings there were no differences between men and women. But there were significant differences by fields (see Table 1).

Table 1. The rating of the S&T indicators by fields (1 = lowest rating, 5 = highest rating)

Indicators	Agricultural sciences	Humanities	Medical Sciences	Social Sciences	Engineering/ Technology	Natural Sciences
Effect indicators						
Citations	4.3	2.9	4.1	3.8	3.8	4
H index	3.8	2.4	3.6	3.4	3.5	3.6
Number of prestigious national and international awards and prizes	3.7	3.4	3.4	3.6	3.7	3.7
Employability of PhD graduates (in private sector)	3	2.1	2.8	2.4	3.3	2.8
Input indicators						
Recruitment of PhD students	3.2	2.9	3	2.6	3.2	3
External funding	3.7	3.2	3.6	3.4	3.5	3.5
Structure indicators						
Number of PhD students	3.2	2.6	3	2.6	3.2	2.9
Research collaborations and partnership	4.1	3.7	3.8	3.7	4	3.7
Reputation and esteem	3.9	3.6	3.5	3.8	3.7	3.5
Output indicators						
Publications	4.5	4.6	4.6	4.5	4.3	4.5
Non-bibliographical outputs	3.1	3	2.8	2.8	2.9	2.8
Number of PhD graduates and completion rates for graduates	3.5	2.7	3.3	3	3.3	3.3

... Continuation Table 1

Indicators	Agricultural sciences	Humanities	Medical Sciences	Social Sciences	Engineering/ Technology	Natural Sciences
Output indicators						
Patent development	3.1	2.4	2.9	2.7	3.2	2.8
Public outreach	3.8	3	3.4	3.5	3.4	3.1
Social indicators						
Relevance to citizens' concerns	3.3	2.7	3.5	3.4	3.1	2.8
Relevance to global societal challenges	3.7	2.9	3.6	3.6	3.3	3
Usefulness to policy decision makers	3.7	2.5	3.3	3.3	2.9	2.7
Contributing to science education	3.9	3.3	3.7	3.5	3.5	3.5
Relevance to science communication initiatives	3.8	2.8	3.5	3.3	3.2	3.1
Process indicators						
Seminar and conference activity	3.2	3.6	3.1	3.3	3.4	3.2
Invited keynotes	3.7	3.3	3.4	3.3	3.6	3.5
International visiting research appointments	3.4	3.6	3.2	3.3	3.4	3.2

This applies in particular to humanities. Although publications received the highest scores, it is clear that understanding what a high ranked publication is varies between fields. Such indicators as citations, h-index and patents are not relevant to the humanities. The h-index is the most widely used criterion (Bornmann and Daniel, 2007) but while correlating closely to peer judgments is more effective evaluating basic science applications and less effective in specific (Rinia, et al, 1998) and small research fields that have a smaller number of citations (van Raan, 2006). Looking at the signers list of the San Francisco Declaration on Research Assessment (http://am.ascb.org/dora/) we see that the impact based indicators are not only humanities concern.

Estimates of indicators related to the PhD (number of PhD graduates and completion rates for graduates, number of PhD students, and employability of PhD graduates in private sector) are also remarkably low in humanities. Indicators related to societal impact (relevance to citizens' concerns, relevance to global societal challenges; usefulness to policy decision makers, contributing to science education, relevance to science communication initiatives) were more highly rated in Medical and Agricultural Sciences, and the lowest in Humanities.

Views vary also at different career stages (Table 2). On an average, the size of rating is directly correlated to the position. Particularly different are the preferences of students and professors.

While relevance to global societal challenges, public outreach, contributing to science education, usefulness to policy decision makers, and relevance to citizens' concerns received the highest ranks from students, professors ranked the same indicators the lowest. And vice versa, while high ranked publications, citations, number of prestigious national and international awards and prizes, h-index received the highest ranks from professors, students ranked the same indicators the lowest.

Table 2. The rating of the S&T indicators by position (1 = lowest rating, 5 = highest rating)

Indicators	Student	Postdoctoral research fellow	Lecturer / Assistant professor	Associate professor / Reader / Senior lecturer	Full professor
Effect indicators					
Citations	3.6	3.9	3.9	3.9	3.9
H index	2.9	3.5	3.5	3.5	3.5
Number of prestigious national and international awards and prizes	3.1	3.7	3.6	3.6	3.6
Employability of PhD graduates (in private sector)	3.2	3	2.7	2.8	2.6
Input indicators					
Recruitment of PhD students	2.9	3	3	2.9	2.9
External funding	3.1	3.7	3.5	3.4	3.2
Structure indicators					
Number of PhD students	2.7	2.9	2.9	3	3
Research collaborations and partnership	3.9	3.9	3.7	3.6	3.5
Reputation and esteem	3.3	3.6	3.5	3.5	3.5
Output indicators					
Publications	4.3	4.4	4.5	4.6	4.6
Non-bibliographical outputs	3.1	3	2.9	2.7	2.6
Number of PhD graduates and completion rates for graduates	3.4	3.3	3.1	3.2	3.1
Patent development	3.1	3	3	2.7	2.5
Public outreach	4.3	3.4	3.3	3	2.9
Social indicators					
Relevance to citizens' concerns	3.8	3.1	3	2.8	2.7
Relevance to global societal challenges	4.4	3.3	3.3	3	2.8

… Continuation Table 2

Indicators	Student	Postdoctoral research fellow	Lecturer / Assistant professor	Associate professor / Reader / Senior lecturer	Full professor
Effect indicators					
Usefulness to policy decision makers	3,8	3	2.9	2.9	2.5
Contributing to science education	3.9	3.8	3.5	3.5	3.3
Relevance to science communication initiatives	3.6	3.4	3.2	3.1	2.9
Process indicators					
Seminar and conference activity	3.5	3.4	3.3	3.2	3.1
Invited keynotes	3.5	3.5	3.4	3.5	3.4
International visiting research appointments	3.2	3.4	3.2	3.3	3.1

The overall position was that various indicators should be weighted differently (agreed by 66.4%) and that there is a need to have different weightings for the various indicators in different subject areas (agreed by 68.5%) as well as a need to have various indicators or different weighting for the various indicators at different career stages (agreed by 69.1%). In responses we can see slight differences between areas; it seems that in natural sciences (80.5%) and medical sciences (77.5%) the need to have different weightings for the various indicators in different subject areas is much more urgent. Fairly predictable is the need to have various indicators or different weighting for the various indictors at different career stages in case of postdoctoral research fellows (71.9%), and lecturers / assistant professors (76%).

Despite the relatively high scores given to the indicators, the respondents were generally rather sceptical. For several times they quoted the phrase of Albert Einstein "*Not everything that can be counted counts, and not everything that counts can be counted.*" The overall position was that indicators should be used as an additional tool. This position was well summarised by a medical scientist from Denmark: "*Peer-reviewing can never be reduced to something quantitative and something that is only scored. It is all about (ir) relevance, comprehensiveness, and quality, which also have to be assessed qualitatively. Furthermore, it is context related and therefore scores across areas/issues/methods might not assess the same. Such scores might encompass differential item functioning and might favour certain areas of research.*"

There were more negative positions, for example a physicist from France stated: "*Most of the indicators can be deliberately biased (h-index, citations) and the number of invited talks etc. is in fact extremely dependent on your networking capacity not necessarily based on your expertise… so if those indicators are good they can also be very biased and this should be kept in mind at all times…*" or a biologist from France: "*Why focusing on stupid indicators? They are meaningless; you just evaluate the ability to improve*

indicators, not the ability to improve the underlying science. Evaluations committees have to do their job, i.e. to evaluate the quality of science based on science (read the papers, interview the scientists, and judge the quality of projects and results based on scientific criteria). By evaluating from indicators, you will just select the most sneaky project coordinators, who do not mind about science but who understood how to cheat the system. Needless to say, these people are the worst scientists."

Conclusion

The most favoured indicator was high ranked publications (4.5), and this is the only one which was accepted by researchers from all fields. The rating is directly correlated to the respondents' position. Particularly different are the preferences of students and professors.

The lion's share of respondents agreed that various indicators should be weighted differently (agreed by 66.4%) and that there is a need to have different weightings for the various indicators in different subject areas (agreed by 68.5%) as well as a need to have various indicators or different weighting for the various indicators at different career stages (agreed by 69.1%). The overall position was that indicators should be used as an additional tool.

References

Bornmann, L; Daniel, HD (2007). Convergent validation of peer review decisions using the h index – Extent of and reasons for type I and type II errors. *Journal of Informetrics*, 1 (3), 204–213.

Mustajoki, A; Newby, H; Peyraube, A; Rowley, S; Salmi, J; Sanchez, P (2010). *Assessing Europe's University-Based Research Expert Group on Assessment*. Luxembourg: Publications Office of the European Union.

Retzer, V.; Jurasinski, G. (2009). Towards objectivity in research evaluation using bibliometric indicators – A protocol for incorporating complexity. *Basic and Applied Ecology*, 10(5), 393–400.

Rinia, E.J.; van Leeuwen, T.N.; van Vuren, H.G.; van Raan, A.F.J. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria – Evaluation of condensed matter physics in the Netherlands. *Research Policy*, 27 (1), 95–107.

van Raan, A.F.J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67 (3), 491–502.

Identifying emerging biorefinery technologies & potential markets: A technometric approach

Michael Novotny michael.novotny@indek.kth.se / Department of Industrial Economics & Management, Royal Institute of Technology, Lindstedtsvägen 30, Stockholm, 10044 (Sweden)

Ulf Sandström ulf.sandstrom@indek.kth.se / Department of Industrial Economics & Management, Royal Institute of Technology, Lindstedtsvägen 30, Stockholm, 10044 (Sweden)

Abstract

Biorefinery activities have the potential to partly substitute oil, fossil and metal industries in a wide range of sectors in a near future (Kamm et al 2010; Axegård 2010). This paper is about methodologies and mapping of "biorefinery" technologies, constellations and potential markets thereof. The methodology is characterized by a technometric approach. The aim is, firstly, to track science and technology frontiers in the biorefining field and, secondly, identify the key agents involved through bibliometrics of applied science and engineering – which then can be matched with patent data and tracking demonstration plants close to commercialization. This paper has been guided by the theory on technological paradigms and socio-technical communities Kuhn, 1962; Dosi 1982; Thagaard, 1986) epistemological communities (Haas, 1992) or learning communities (Wenger, 1998). On the basis of this theory we formulate a hypothesis predicting that traditional forest industry countries (high exports/share of GDP) may face a lock-in effect of concentration for research to traditional industries. Our results indicate that countries like Sweden, Finland, Austria, and Germany – which to a considerable extent have been the frontrunners of forest industry technologies both with regards to exports, production volumes, R&D new equipment – are at risk due to lock-in effects.

Introduction

Growing environmental concerns, emerging economies and strong uncertainties of future supply of fossil feedstock contribute to a renaissance for biomass industries. Biomass is being increasingly regarded as an important raw material for industries that traditionally have had other feedstocks, e.g. oil, (petro)-chemical, automotive, metal and material industries. The increasing struggle for finding solutions to possible oil shortage and environmental impact abatement in different sectors (energy, automotive, chemical, basically all large scale industries) is in turn resulting in a global scramble for biomass in particular in the EU (UNECE/FAO, 2009, 2011).

This paper focuses mainly on mapping of "biorefinery" technologies, R&D constellations and potential markets thereof. The methodology is characterized by a technometric approach. The aim is, firstly, to track science and technology frontiers (S&T communities) in the biorefining

field and, secondly, identify the key agents involved through scientometric methods – which then can be matched with patent data and tracking of emerging technologies expressed by demonstration plants close to commercial introduction. A bibliometric analysis is useful in order to process a large amount of information from database material; in this case it particularly facilitates the mapping and identification of the frontiers of emerging technology fields. This quantitative method will, in a second part of the study, be supported by qualitative input i.e. expert opinion in picking relevant clusters, identify S&T/industry constellations and develop analysis from selected key words/cluster labels (bibliometrics combined with Derwent Index patent database).

The aim of the paper is to analyze emerging technology frontiers, the S&T communities behind them, and potential new entrants in the field of biorefinery. The questions that arise are: where are R&D efforts (publications primarily) done globally and who is in the front seat; country- and sector/company-wise? Since emerging fields probably are at rather early phases of technology development the unit of analysis is more on S&T communities – so called fronts or clusters.

Technological paradigms and methodological issues

Technological paradigm shifts or even larger regime shifts (Geels, 2002) usually involve series of interconnected events of industrial activities or technical development (Dahmén, 1950). In an era of high uncertainties due to climate change, environmental concerns, and competition from the BRICs (emerging economies/markets) the transformative pressure on energy-intensive and natural resource based industries (NRBI:s) has increased (UNECE/FAO 2009/2011).

A technological paradigm is made by a pattern of the technology at stake and by the specific technological challenges caused by such pattern (e.g. increasing capacity and dimensions in a process industry). It is "a set of procedures, a definition of the 'relevant' problems and of the specific knowledge related to their solution" (Dosi, 1982, p 149). Therefore, technology is identified as a problem-solving activity in which the problems to be solved are selected by the paradigm itself and its engineering and scientific communities (Thagaard, 1986). In this sense, a technological paradigm involves a strong trend on the trajectory of technological change that is the direction toward which future technical progress will converge. Such gradual improvements along specific lines set by the paradigm are what represent technological trajectories and progress (Dosi, 1982; Teece, 2008).

Inspired by Kuhn (1962), Constant used the concept 'technological community' to examine the practitioners of the technological change, i.e. the professional engineers who have both common educational backgrounds and common career expectations. Thagaard (1986) did the same for the scientific communities which in the industrialized world have been strongly connected with technological development. Technological communities with relatively homogenous cognitive view and expectations on knowledge formation could therefore develop considerable lock-in.

Constant demonstrated that old engineering communities seldom initiate radically new innovations. In his example, engineers of piston aircraft engines did not play any role in the development of the turbojet aircraft. As a consequence a lock-in usually takes place and whenever a radical shift occurs the incumbents rarely are the initiators (Christensen, 2000).

The biorefinery field is not a completely new field of knowledge, its roots derive from organic chemistry of the first half of the 20th century (Kamm, Gruber, Kamm, 2010). However, it has to a large extent been used as an industrial multi-product concept rather than a coherent science field, since there are a few established examples out there already (ethanol, sulfite and pine oil plants expressed by DuPont, Arizona Chemicals, Borregaard, Domsjö). However, by defining biorefineries as innovative activities rather than established production units we are able to overcome the academy-industry problematique.

There are several steps in a technometric delineation of an area. Firstly, the definition of the research field via keywords selection (a dozen keywords from the biorefinery field have been picked from recent conferences and academia – e.g. the definition of lignocellulosic biorefinery in Kamm et al 2010 and key concepts from conferences, e.g. Biorefinica and Nordic Wood Biorefinery Conferences). By selecting frequently used topics in the field such search terms aroused.

Our study design is based on findings from the most cited academic publications in applied science journals of biofuels, biochemicals, biomaterials and biomass fields which most frequently have used the term "biorefinery" and/or "biorefining" indexed by Thomson Reuters in Web of Science (WoS). In turn WoS was centered on the data base commonly known as SCI-EXPANDED.

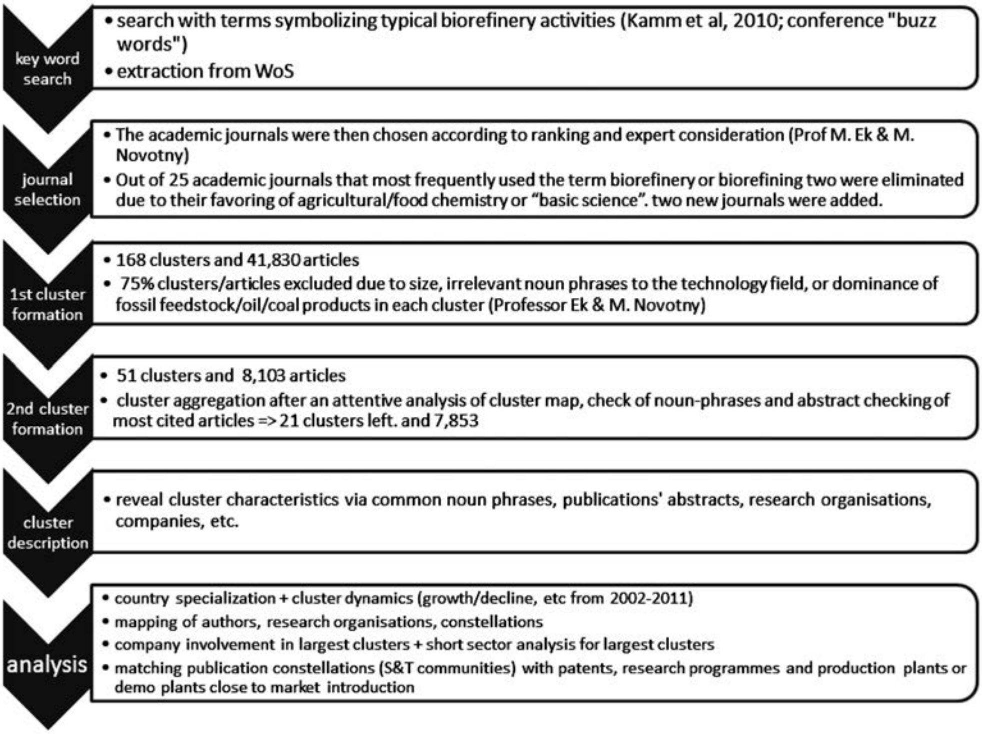
Following this keyword selection process, out of a first draft, 41,830 scientific publications were picked. A clustering procedure was then performed based on shared references between articles – i.e. bibliographical coupling (Kessler, 1963). For each research front a matrix is constructed which consists of a set of publications (connected to researchers). We receive a "similarity matrix" – a necessary input for cluster analysis. For the clustering we have used the MAM-method, a program that uses a clustering algorithm, which also suggests the optimal number of clusters (Zhang et al., 2010). Each front is a sub-set of publications on the basis of reference-similarity profiles and was mapped as a cluster or technological community. The idea is that researchers and engineers (connected to universities, institutes and companies) that use similar references establish communities with a common terminology and problem-solving issues, similar to those who Dosi labeled "technological paradigms" (Dosi, 1982).

Specifically, out of 168 clusters based on 41,830 articles, 46 clusters (and roughly 10,000 publications) were chosen as relevant. Three parameters were used in this selection process; a) the 20 most frequent noun phrases of each cluster indicating one or more of typical biorefinery activities (biomaterials, biochemicals, biofuels); b) expert advice on cluster description; c) most cited publications in a dozen "borderline clusters" with few relevant noun phrases (in short: when very few noun phrases indicated technologies or products based on biomass as a complement,

abstracts of the most cited articles indicated whether the research field was to be considered relevant or not).

A challenging step is the integration of mapping and performance. It helps us to position researchers from universities, research organizations and firms on the worldwide map of their field and to seize their influence in relation to impact-levels of the different clusters and research fields/communities. In that way we aim to map applied science, technology development and the agents from R&D and industry in the field. Individuation of researchers, research teams can then be connected country-wise or to corporation's R&D divisions. Here, a more qualitative approach comes at stage in order to identify links, R&D constellations, etc.

Figure 1. The methodological process of the study.



After the second clustering process we ended up with 21 S&T frontiers:

Table 1. Cluster description (label, characteristics, number of publications, prominent organizations and companies).

Cluster	Label	Characteristics	Publication count 2002-2011
1	Pyrolys	The biofuels cluster per excellence. Bio-oil (gasification and pyrolysis technologies dominating). Most frequent universities/research org: China Acad Sci and main Chinese Uni, VTT, Maine Uni, Zaragoza Uni. Companies: Foster Wheeler, Shell, BTG, Fortum	844
2	Biodiesel	Biofuels with emphasis on bioethanol and biodiesel based biomass, in particular microalgae and switchgrass. Most frequent universities/research org: US Doe, Beijing Chem. Uni, many US uni,, Comp: DuPont, Shell and small US firms	283
3	Nanowhisker	Nanocomposites, nanocellulose, nanocrystals, advanced bioplastics Org: USDA Ars, Giorgia Tech, Univ Saskatchewan, Wuhan Uni, Luleå Uni, Pais Vasco Uni, Helsinki Uni, VTT, Åbo Akademi, Innventia/STFI, BIM Kemi, Mitsubishi, BASF Inc,	203
4	Nanofiber	Nanocellulose, nanowhisker Org: Cornell, Penn State, USDA, Tianjin Uni, Donghua Uni	105
5	MFC	Microfibrillar cellulose, nanocellulose Org: Kyoto/Tokyo Uni; Innventia/STFI, USDA; Georgia Tech, KCL, Åbo, Donghua,	253
6	Biochemicals	Biorefinery products (compounds/platform chemicals) based on biomass among others: stover, algae, cellulose pulp. Lignin, saccharification, hemicellulose derivatives, xylanase, cellulase, enzymatic treatment, arabinoxylan Organisations: Wisconsin, Maine Uni, vTi, Vigo/Huelva Uni, Lappeenranta, Georgia Tech, Beijing Forest Uni, Exxon, Chevron, UPM, Ciba, Biopulping, Mead Westvaco, Lenzing.	923
7	Fractionation	Separation/fractionation of product streams in pulp mills: lignin, (dissolving) pulp, delignification, kraft. lignin-carbohydrate, black liquor, xylan Org: British Columbia Uni, Royal Inst. Tech (KTH), Georgia Tech, Domtar, Södra, MeadWestVaco.	370
8	Lyocell	Advanced wood based cellulose textiles substituting cotton and oil polyesters. Org: Lenzing AG; Innsbruck Uni, Vienna Uni	107
9	Pulping	New applications for cellulose pulp and optimization of by-products. Company-intensive cluster. Org: Paprican/FPI, Georgia Tech, Tianjin Uni,Helsinki Uni, Stora Enso, International Paper, Tembec, Bowater, Metso, Andritz, Noss, Eka (traditional p&p research organizations/firms). Exceptions: Mitsubishi, Specialty Minerals	584
10	Wood composites/ morphology	Focus on mostly woody compounds (e.g. fiber strength) of softwoods and some lignins (phenols/resins). Org: Kyoto Uni, Georgia Tech, BC Uni, Helsinki Uni, VTT, FPI, Beijing Uni, Canfor	218
11	Hydrogels	Advanced biomaterials; Chitosan-based hydrogels and films. State financed cluster. Org: Mostly Chinese Uni and R&D inst, Montreal Uni,Birmingham Uni, Unilever, Birla	841
12	Ionic liquids	Ionic liquids, levulinic acid Org: Beijing Uni. Chem Tech, Colorado Uni, DuPont	301
13	Jet fuel	USA-dominated cluster. Org: Ford Motor, National Institute of Standards & Technology, US DOE	74
14	Biogas	Anaerobic digester; sludge; biogas; methane Org: State-dominated, many US and Spanish R&D org; Dublin/Cork Uni, Lund Uni, Beijing Uni Chem Tech, NRC, Corp: Domtar Veolia, GIRO, Birla	265
15	CMC	Carboxymethyl cellulose, e.g. additive for food and medicine app's. Org: Nanyang Tech Uni, Karlstad Uni, Jena Uni, Sao Paolo Uni, Astra Zeneca, Pfizer, Danone, Unilever	190
16	PLA	PLA, bioplastic. foam, starch filmOrg: Chinese & US Uni's, Sao Paolo Uni, INRA, VTT Firmenich, Birla	538
17	Cellulose film	Bioplast from MFC, nano, pulp,ionic liquid,chitosan Royal Inst Tech, Innventia, Georgia, Kyoto, Donghua, Åbo/VTT, Corp: Kemira, BIM, SCA, Domsjö/Processum, Lenzing, Biorefinery GmbH, BASF, KFA, KAO, small US firms	681
18	Wood extracts	Wood by-products (e.g. lignan, flavonoid, bark). Org: Åbo Akademi, SLU, Maine Univ. Max Planck,FPI, Lenzing, Afocel, Meadwestvaco, Suzano, Andritz, Forintek, M-real,	334
19	bioplast (lipase)	Sugars, films based on algae and fast growing biomass feedstock. Org: Kyoto, Polytech NY, Innventia	197
20	SCWO	Supercritical water oxidation, supercritical water gasification. Org: Japanese, German & US Universities,	266
21	Biomethanol	Methanol, DME, biodiesel, research org mainly from Spain, China, USA. Org: Pais Vasco Uni, Utah Uni, National Petrochem Corp, Korea Research Institute of Chemical Technology	137

Analysis from scientometrics

The analysis has, hitherto, been performed based on four types of data processing in order to answer our research question: 1) cluster dynamics (growth/decline, etc from 2002–2011); 2) country specialization; 3) mapping of authors, research organizations, constellations in clusters; 4) company involvement in the largest clusters plus a short sector analysis for largest clusters. Later on, a fifth analysis will be performed; i.e. matching publication constellations (S&T communities) with patents and demonstration plants with technologies/products close to commercialization.

Firstly, which clusters are growing and which countries are positioned in the fast growing ones? We analyzed publications per year, and cluster and their progressions (table 2). Five clusters had a strong progression the last couple of years – hydrogels, cellulose film (nanoplastics), biochemical, biodiesel and pyrolysis clusters took off around 2005, more or less when "biorefinery" was becoming a recurrent term in research programs. More forest industry related clusters are somewhat slow growing – with the exception of the biochemicals and fractionation clusters. This is probably due to the older research and technology community and non-expansive public and industrial (private) R&D funding.

Table 2. S&T clusters in biorefinery activities: number of published papers per year

AREA-CLUSTER#	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	Growth RANK	DELTA 2005-2011
Hydrogels chitosan-11	57	53	70	80	151	148	191	220	284	529	1	60,2
Biochemicals-6	99	88	91	84	85	143	184	321	373	374	2	58,0
Cellulose film-17	45	45	38	75	86	91	129	150	165	470	3	50,1
Biodiesel-2	1	2	0	1	4	8	27	133	171	237	4	41,7
Pyrolys-1	43	68	60	70	128	118	216	227	301	125	5	22,1
PLA-16	35	49	70	54	89	98	144	163	176	153	6	19,1
Biogas-14	5	20	10	16	26	51	73	78	119	94	7	16,0
Biomethanol-21	5	2	6	8	2	9	27	21	24	111	8	13,0
Ionic liquid-12	2	8	25	32	29	61	113	93	88	69	9	9,3
Nanowhisker-3	6	12	10	14	25	73	52	86	150	5	10	8,4
MFC-5	35	40	35	44	62	42	55	57	75	101	11	7,6
Nanofiber-4	0	2	7	1	11	32	31	41	88	16	12	7,4
Wood composite-10	25	14	42	29	30	32	35	52	86	43	13	6,2
Jet fuel-13	0	0	3	5	3	8	11	19	26	15	14	3,1
Fractionation-7	75	53	37	63	71	73	91	100	67	84	15	2,9
Lyocell-8	8	3	16	17	24	17	32	57	29	19	16	2,0
SCWO-20	46	44	45	36	49	48	38	51	43	44	17	0,5
Bioplast (lipase)-19	7	21	11	8	13	12	16	12	16	3	18	-0,3
Pulping (TMP)-9	104	87	103	126	114	90	89	127	152	72	19	-1,8
CMC-15	34	38	21	44	51	39	28	34	45	23	20	-2,9
Wood extracts-18	66	87	58	91	80	42	70	70	85	31	21	-5,1

Secondly, what about country specialization? In order to perform such an analysis countries with less than 70 publications were excluded and considered obsolete. The mapping procedure resulted in 21 clusters based on 7,853 articles. These clusters were analyzed country-wise, mainly

the largest forest industry based nations plus emerging countries, according to their relative comparative advantages (RCA) per research front/cluster see table 3 below (Balassa, 1965). RCA shows the relative specialization of each country over clusters. For this analysis the 18 countries with the highest number of publications were selected. Therefore, the analysis using RCA is based on 6,167 articles. Traditional forest rich countries (largest exporters, production volumes or industry's share of GDP) such as USA, Canada, Sweden, Germany and Finland are frontrunners in some of the clusters. USA is at the first position by article production rate (#1228, 20% of publications in relation to considered 18 countries). China is at second place and probably first within a few years (#1121). Japan is third (#530), followed by Canada (#433), then France (#312) and Spain (not renowned for possessing a strong biomass/forest export sector) on a surprising sixth place (#298). Germany (#293), Sweden (#283) and Finland (#257) were not far behind.[i]

Depending on our theoretical hypothesis we have grouped countries in three structures: old countries, new world countries (BRIC and Tiger Economies). In between, we propose that there are the new European biomass countries. Traditional forest industry nations (called "old forest nations", see table 3), such as Sweden, Finland, Germany, Japan, Canada, USA are present in many clusters, but often in slow-growing clusters.. In the old world large incumbents demonstrate strong interaction with the S&T communities which may confirm the trend that they are related to established clusters where many firms are involved in research publication. Small firms, in many cases spin-offs from Universities, are also particularly present in USA.

Traditional forest industry nations are particularly prominent in clusters where wood compounds (cellulose, lignin, hemicellulose) or separation technologies are the main focus (e.g."fractionation", "wood extracts", "wood composite", i.e. cluster 7, 10, 18). To some extent this is also the case of Brazil we placed in the "new World section" in table 3 – a forest rich nation that has had a strong industrial development since the 1980s. The old forest nations are positioned in many slow-growing clusters which were established early around 2002–2003 (cluster 34, 18, 77) and in few occasions in fast-growing clusters ("cellulose film"). Old forest nations such as Canada, USA and Japanhave slightly better positions than Scandinavia in "hot" clusters, but lag behind "New European Biomass" (Spain, Portugal, etc) and "New World" (Asia and Brazil).

US organizations are a prominent players in many areas, but not unexpectedly they often demonstrate a stronger specialization in different forms of bioenergy/fuels (policy and energy security reasons), while Asia, i.e. China, Taiwan, South Korea, India are strong in fast growing clusters – e.g. in particular hydrogels chitosan based, PLA films and more than average in most biofuels clusters (not in US dominated jet fuel). Fast growing clusters may be labeled as per se a function of China's fast publication growth.

Table 3. Revealed Comparative Advantage, RCA (Balassa, 1965) per country and research cluster (1,0 = expected value). Rank follows figures in brackets, higher number means high publication production in areas with low growth, i.e. added five areas' placement with strongest specialization, in red) Instead, low figures (in bold) indicate activity in growing (hot) areas. RCA=the share of country y in cluster x in relation to the total of cluster x divided by country y's total publication production in relation to the grand total of publications (#7583). Growth trend in each cluster is due to the regression coefficient and demonstrate long term trend over ten years, see rank column

Region	New World					New European Biomass					Old Forest Nations										
CLUSTER																			Expect	1stTerm	RANK
11	1,7	2,3	2,9	4,5	1,2	0,6	0,7	0,7	1,0	0,6	0,5	0,4	0,1	0,8	0,1	0,4	0,0	0,2	1	hydrogels chitosan	1
6	1,5	0,9	0,3	0,8	1,3	1,1	1,0	0,4	1,8	1,6	0,6	1,0	1,0	0,8	0,8	1,1	0,7	0,7	1	biochemicals	2
17	0,7	1,3	1,0	0,4	0,7	1,5	1,4	0,2	0,9	0,6	0,7	0,7	1,3	1,2	1,4	0,8	0,9	1,3	1	cellulose film	3
2	0,9	0,8	0,4	1,3	1,9	0,7	0,4	0,8	2,3	0,5	1,9	0,4	0,5	0,5	0,4	1,9	0,1	1,0	1	biodiesel	4
1	0,9	1,3	0,6	0,4	0,1	0,3	0,2	5,3	1,4	1,6	0,9	0,8	0,5	0,9	0,4	0,9	0,3	0,3	1	pyrolys	5
16	1,7	1,2	1,4	0,8	3,2	1,8	0,5	0,3	1,2	0,6	1,2	0,7	0,5	0,5	0,6	1,0	0,6	0,6	1	PLA	6
14	2,6	1,6	0,8	2,8	0,8	1,1	0,4	1,3	0,6	3,4	0,6	1,4	0,1	0,2	0,6	0,2	0,0	0,7	1	biogas	7
12	0,9	1,3	0,5	0,8	0,6	0,4	2,0	0,1	1,0	1,8	0,1	0,4	0,2	0,4	0,0	1,8	0,1	1,6	1	ionic liquid	8
3	0,5	0,3	0,0	0,0	1,3	3,3	0,2	0,0	1,3	0,6	0,0	1,3	1,3	2,0	1,8	1,1	0,3	0,4	1	nanowhisker	9
21	0,7	1,2	2,4	0,5	0,0	0,6	0,0	1,0	0,3	3,8	0,2	1,2	0,2	0,7	0,0	1,2	0,5	0,9	1	biomethanol	10
4	1,5	1,1	5,2	1,4	0,4	0,0	0,9	0,0	0,0	0,3	1,8	0,3	0,2	1,7	0,7	1,3	0,0	1,3	1	nanofiber	11
5	0,1	0,5	0,9	0,0	0,3	1,8	0,4	0,4	1,3	0,4	0,3	0,4	0,9	2,4	2,4	1,0	1,7	2,0	1	MFC	12
10	0,7	0,3	0,2	0,7	0,1	2,4	0,5	1,7	0,2	0,0	1,7	2,6	1,9	1,3	1,1	0,5	1,2	2,8	1	wood composit	13
7	1,1	0,3	0,4	0,2	1,2	1,1	3,0	0,3	0,2	0,8	0,1	0,8	2,2	0,9	3,6	1,1	1,1	0,7	1	fractionation	14
8	0,9	0,5	0,2	0,0	0,1	0,7	0,0	0,0	1,5	0,1	1,4	0,1	0,5	0,3	0,1	0,5	29,6	1,8	1	lyocell	15
13	0,0	0,4	0,6	0,0	0,0	0,0	0,6	0,0	0,9	0,0	0,0	0,2	0,0	0,0	0,0	4,1	0,8	0,5	1	jet fuel	16
9	0,1	0,4	0,3	0,1	0,6	0,7	2,7	0,2	0,3	0,8	1,9	3,6	3,1	0,4	1,3	1,0	0,4	0,5	1	pulping (TMP)	17
20	0,2	0,5	2,0	0,3	0,0	0,5	0,0	0,5	0,8	0,8	1,0	0,2	0,1	3,5	0,1	1,6	0,0	1,8	1	SCWO	18
19	1,3	0,3	0,7	0,9	0,0	0,8	0,0	0,0	0,0	0,0	0,0	0,4	0,4	4,2	1,8	1,5	0,0	0,8	1	bioplast (lipase)	19
15	1,2	0,6	0,3	0,0	2,4	0,7	0,3	0,4	0,8	0,8	3,7	0,3	0,2	1,4	2,7	0,2	1,1	5,4	1	CMC	20
18	0,7	0,3	0,9	1,3	1,9	0,8	1,6	0,8	0,4	0,4	3,1	1,3	2,8	0,5	2,1	0,9	2,7	1,3	1	wood extracts	21
EXPECT	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
Five Rank	####	####	46	44	#####	####	63	39	####	####	73	67	####	####	86	65	81	####	Rank Max = 95; Rank Min = 15		
Avg Rank	40					43					74								Avg Rank All = 56		

China and USA, the two giants, are similar in total publication volume, but specialization is rarely in the same S&T clusters – only in the "ionic liquid" cluster. Specialization occurs in different clusters due to different policy prioritizing and sector/company involvement.

For analytical purposes we rank the clusters according to the progression for each cluster, i.e. the growth of publications over time. From that follows our ranking of clusters which is presented in Table 3 (see last column to the right). The sum of ranking number for the five clusters where each country has its relatively highest activity will give the "five rank" number (see row at bottom of Table 3). Accordingly, if a country (or region) has a lower five rank figure, then activities are more concentrated toward fast growing areas and vice versa. China (PRC) has the lowest figure of all countries and the group of new world countries has a markedly higher rank number. New European biomass countries seem to be at the same level as the new world.

Since we have identified company presence in most publication clusters we combined product and company search in the patent analysis. Industry involvement in both bibliometrics and patents is clear in – Austria's remarkable specialization in lyocell research (table 3 with score 29,6) – Lenzing, an Austrian company, has strong positions in the lyocell technology (wood based textiles), cluster (8) which reveals a strong correlation between publication counts, patents and specialization.

In our patent analysis we will focus on the largest publication clusters where the presence of Western and in particular Japanese corporations is prominent. Clusters with high company involvement in the publication analysis tend to be prominent also in the patent analysis i.e. these will be analyzed by using basically identical key words as in the bibliometric analysis. . Our subsequent study will analyze which companies that have accomplished patents with the most prominent researchers, i.e.to consider "linear" or circular innovation processes. In completely new clusters (i.e. advanced nanomaterials, bioplastics) on a first glance they tend to have similar characteristics as R&D biomedicine (patents/innovations connected to high-performing universities (this hypothesis will hopefully be discussed thoroughly in Berlin next fall).

Contributions and further research

This paper is guided by the theory on technological paradigms and sciento-technical communities (Kuhn, 1962; Dosi 1982; Thagaard, 1986). On the basis of this theory we formulated a hypothesis predicting that traditional biomass rich countries, in particular forest industry ones, will face a lock-in effect of concentration for research to traditional industries. Our results, so far, indicate that traditional European forest industry nations like Sweden, Finland, Austria, and Germany – that to a considerable extent have been the frontrunners of forest industry technologies during several decades both with regards to R&D, patents and machine supply (Laestadius 1998) – are risking lock-in effects. However, the analysis is two folded. Countries from New World and New European Biomass demonstrate a more conclusive national strategy within

biorefinery research. They have fewer lock-in effects due to lower involvement from national industry. These countries have in common an absence from "national industry" which in turn could be understood as a lack of institutionalization of the research establishment.

However, the inclination for partnering with industry (and national champions) of S&T clusters in traditional forest industry nations is not necessarily a lock-in, but could be a "positive" incremental path dependency characterized by a "Dosian" technological community (engineers from university, research institutes and industry) and with greater impact on the economy than the more homogenous, fast growing science communities in the New World. It may be argued that the greater involvement of S&T communities with incumbents in traditional sectors in the Old World, such as the forest or chemical industry, have a larger economic impact since they are already rooted in the Old World where industry have a relatively larger share of GDP/employment/exports/patents, etc. For future research longer time spans than a decade may enable for higher validity for research on S&T frontier formations and trajectories.

References

Balassa, B. (1965). Trade Liberalisation and "Revealed" Comparative Advantage. *The Manchester School*, 33, 99–123.

Constant, Edward W. (1980) *The origins of the Turbojet Revolution*. Baltimore: John Hopkins University Press.

Christensen, C. M. (2000) *The Innovators Dilemma*. Cambridge UP.

Dahmén, E. (1950) *Svensk industriell företagarverksamhet. Kausalanalys av den industriella utvecklingen 1919–1939*. [*Swedish industrial and entrepreneurial activities. Causal analysis of the industrial development 1919–1939*]. PhD diss. Stockholm Business School.

Dosi, G. (1982) Technological paradigms and technological trajectories. A suggested interpretation of the determinants and directions of technical change, *Research Policy*, 11, 147–162. Geels, F.W. (2002) Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study, *Research Policy*, 31, 1257–1274.

Kamm B., Gruber P.R., Kamm M. ed. (2010) *Biorefineries – Industrial Processes and Products*. Weinheim: Wiley-VCH Verlag.

Kessler M. M. (1963), "Bibliographic coupling between scientific papers." "American Documentation", 14(1), 10–25.

Haas, PM. "Epistemic Communities and International Policy Coordination." International Organization. Vol. 46. No. 1. Winter. MIT Press, 1992, 1–35.

Kuhn, T. (1962) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Pavitt, K. (1984) Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy*, 13, 343–374.

Teece, D.J. (2008) Dosi's technological paradigms and trajectories: insights for economics and management", *Industrial and Corporate Change*, 17, 507–512.

Thagaard, T. (1986) *Scientific communities*. Oslo: University of Oslo (diss.)

UNECE/FAO (2011) *European Forest Sector Outlook Study*, Geneva: Geneva Timber and Forest Study.

Van Raan, A.F.J. (2004), Measuring Science in Moed H.F. (ed.) *Handbook of Quantitative Science and Technology Research*, 19–50. New York: Kluwer Academic Publisher.

Zhang, L., X.H. Liu, F. Janssens, L.M. Liang, & W. Glänzel (2010). Subject clustering analysis based on ISI category classification, *Journal of Informetrics* 4, 185–193.

Conceptualization and development of benchmark bibliometric studies for research assessment

Ed Noyons noyons@cwts.leidenuniv.nl / Centre for Science and Technology Studies, Leiden University, PO Box 905, Leiden, 2300 AX (The Netherlands)

Rodrigo Costas rcostas@cwts.leidenuniv.nl / Centre for Science and Technology Studies, Leiden University, PO Box 905, Leiden, 2300 AX (The Netherlands)

Abstract

Benchmark analyses are a common tool in research evaluation. In this study we present a conceptualization of the different benchmark possibilities that are available with research evaluation purposes. We develop a bibliometric methodology that can contribute to automate and objectivise the detection and selection of benchmarks for any given focal unit of analysis, and we show it for the case of Leiden University. This methodology can contribute to expand the benchmark possibilities currently available in science policy.

Introduction

Benchmarking is seen as a tool for enhancing performance evaluation achieved through comparison (of a focal unit) with other organizations recognized as the best (Andersen & Pettersen, 1995) and it is regarded as a good management practice (Moriarty, 2008). The idea of benchmarking is common in bibliometric studies (Murphy, 1995; Thijs & Glanzel, 2009; Tijssen, Visser, & Van Leeuwen, 2002), where research organizations (and even researchers) wish to compare their performance with that of other units that could represent a good parameter or example for the improvement of the focal research organization. The basic idea of benchmarking is to compare your performance with the performance of others. The choice of those 'others' is crucial, of course. Often evaluated institutions suggest their benchmarks themselves. They decide with whom they should be compared. It is not always clear on what grounds these benchmarks are chosen.

University Rankings (e.g. WUR, Leiden Ranking – (Waltman et al., 2012) are good examples of benchmarking where there is a limited amount of criteria to be entered (among which: being a university). Such benchmarking has become very important both as a management and as a marketing tool. In spite of this important role of benchmarking in research evaluation and science policy, to the best of our knowledge, the definition of a proper benchmark has never been developed. For example, in the university rankings, does it make sense to compare a mega university with ditto resources with a small, specialized university? And, in spite of advanced and sophisticated normalizations, it seems inappropriate to compare the performance of a specialized *Delft University of Technology* with the performance of *Harvard* or *Cambridge*. Not because the

difference of reputation but primarily because of the different (research) scopes. In this paper we will present an approach to identify benchmark candidates of any focal research entity (e.g. university, institute, group or person) on the basis of their research portfolio. The proposed approach in this paper to identify benchmarks does not imply that the results of a comparison can be used in a strategic discussion, as it often is in economics (Feelders & Daniels, 2001). It merely evaluates the performance in the past of a unit and positions it in relation to other, similar units. Strategic decisions need more than looking at the past.

Objectives

The main objective of this paper is to make a first attempt for a general conceptualization of bibliometric benchmark studies and secondly, to present a bibliometric application aiming at broadening the possibilities of bibliometric benchmark studies.

Conceptualizing bibliometric benchmarking

Defining benchmarks may be done from different perspectives. Finding benchmarks within a national context differs from benchmarking within the context of a certain funding system. There are many different parameters that may be at stake:

- Type of organization;
- Funding base;
- Geographical location;
- Size;
- Mission;
- Research portfolio.

In our conceptualization of benchmarking we discern two types of benchmark identification: qualitative and quantitative. The former involves a suggestion done by the focal unit itself while the latter involves an identification of candidate benchmarks based on bibliometric statistics. Bibliometric statistics may regard topical research focus, production or citation impact. In table 1 we conceptualize these types for bibliometric benchmark analysis.

Table 1. Main types for bibliometric benchmark analysis

	Qualitative	Quantitative
Approach	Based on a qualitative selection of benchmarks (based on perceptions, ideas, past, traditions, etc.), normally done by the focal unit.	Based on the performance and thematic profile of the focal unit compared to a set of candidate benchmarks.
Advantages	– Focal unit is aware of own interests and why the benchmarks are valuable for it – Perception of control over the assessment – Benchmarks (can) remain stable over time in the analysis (e.g. enabling trend	– More objective and based on the actual performance of the focal unit – No biases in the selection of benchmarks – Benchmarks may be changed over time, adapting better to the current situation of the focal unit
Limitations	– The input from a focal unit may not be realistic or objective – Danger of biases in the selection of benchmarks (too excellent/low benchmarks)	– Requires a substantial amount of biblio metric data standardized – Focal unit has less influence on selected candidates. – Benchmarks may change over time.

For the second type we can base the criteria for candidates on three aspects of research performance:

(1) Size, as represented by scientific output, for instance;
(2) Research portfolio, as represented by distribution of output over research fields or topics;
(3) Citation impact, as represented by average number of citations received per publication.

In addition to such performance characteristics, candidates may be selected on the basis of other relevant features, such as origin, organization type or funding base.

We will elaborate on the second characteristic (research portfolio) in this paper. We will discuss and demonstrate an approach to identify candidate benchmarks for different research entities (universities, institutes and even persons) using the distribution of their output over fields and topics. In Noyons, Moed and Luwel (1999) we applied this method for the first time using the INSPEC classification scheme. The approach presented here is an advanced elaboration of that idea and as such more generic and applicable for any research entity.

The basic idea behind the approach to identify candidate benchmarks with a similar research portfolio is as follows: there is a structure of scientific output worldwide, i.e., a classification of all output. For a focal unit, we distribute its output over the scheme and identify those classes with the most output of the unit. This will be the 'core' of its research portfolio. Within this core we identify other units with a substantial output. These units are candidate benchmarks. The best candidates will be selected by matching the distribution of their entire output with that of the focal unit.

A final issue regarding the proposed approach is the part of the oeuvre of both focal unit and candidates to be considered (for selection as well as for the performance measurement): 'core only' or 'complete oeuvre'. The list of candidates is based on the (topical) core. Hence it would make sense to base the final matching of profiles on the same topical core only. On the other hand, if we want to compare the focal unit entirely with benchmarks, it makes sense to use the complete oeuvres, including more peripheral activities. Similarly, to measure the performance of focal unit and benchmarks we may use their core research focus or their full portfolio (output). In this study we use the entire oeuvre of universities, as selected from the Leiden Ranking.

Methodology to select benchmark candidates quantitatively

In economics (c.f., Feelders & Daniels, 2001) benchmarking is a well established process used in strategic studies. In research evaluations, particularly those using bibliometric data, there are no standardized methodologies to objectively and systematically detect the best benchmarks for a given focal unit. Previous studies (Thijs & Glanzel, 2009) have indicated that one of the major problems regarding bibliometric benchmark analysis is due to the disparity on the profiles of research units, making a proper comparison a difficult task. In that study research organizations were clustered on the basis of their research profile (distribution over journal classes). In the present study we also want to detect the best benchmarks for a given focal unit but from a different perspective. In our case, we do not cluster research entities but identify the most similar benchmarks for any given unit (e.g. universities, research institutes, but also research groups or even individuals). Hence we allow all focal units to have their own appropriate benchmarks. We designed a simple methodology, explained below. It can be described in 6 steps:

Step 1. Data collection for the focal unit
Step 2. Determination of its research core within a publication classification
Step 3. Detection of the most active units the same core – benchmark candidates
Step 4. Collection of complete output of the benchmark candidates
Step 5. Determination of the similarity of the benchmark candidates with the focal unit (cosine similarity measure)
Step 6. Final selection of the benchmark units and performance of the evaluation (e.g. through bibliometric indicators).

There are at least two reasons why a quantitative method for identification of benchmarks is preferred over a qualitative approach. First of all, evaluated units are not always aware of who their benchmarks are. They know their competitors, but they sometimes are not aware of other peers because these remain out of their sight or are geographically too remote. A second reason involves the ranking systems. As stated above, it does not always make sense to compare the performance of a specialized university of technology with the performance of a broad university such as *Cambridge* in the context of bibliometric performance measurement. Their focus as well as their mission is too different. Hence, a ranking based on the full oeuvres of universities in general (e.g.,

the Leiden Ranking) has its flaws. This seems to be a major concern of both general universities and specialized ones.

In Thijs and Glanzel (2009), a classification of universities is proposed based on their research profile. Clusters of universities are identified of universities with a similar research profile. This seems a viable way but has the disadvantage of a rigid classification. As each university is unique, the most relevant candidate benchmarks for one university in a cluster may be different from the candidates for another in the same cluster.

In our approach we also use a unit's profile but allow the identification of benchmarks from each unit individually. The context of benchmarks is different for each unit. Our approach is dynamic and as such applicable at any level of a unit. In the remainder of this section we will demonstrate how this works.

Classification scheme

The basis of this approach is a facility recently developed at CWTS involving a publication-based classification of the entire Web of Science on three levels (Waltman & van Eck, 2012).
This hierarchical scheme structures the entire set of publications in the Web of Science (2010–2010) at three levels:

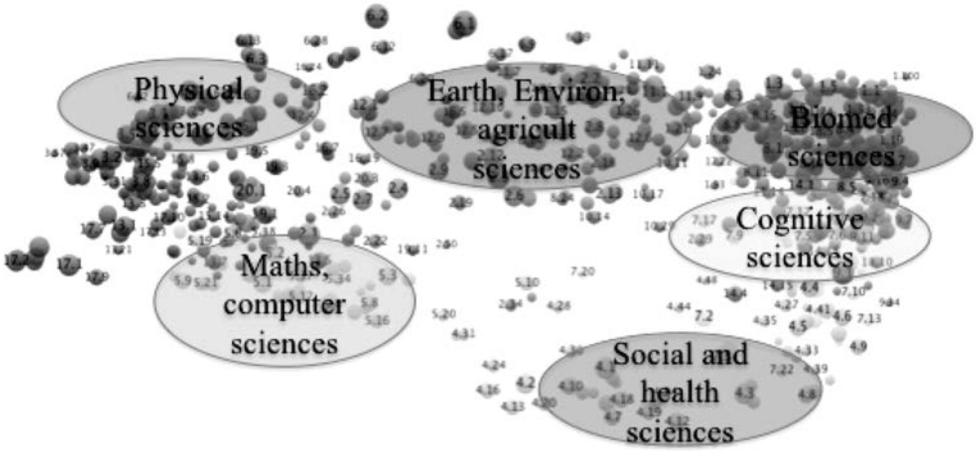
- top level of 20 clusters of publications
- intermediate level of 672 clusters of publications
- bottom level of around 20,000 clusters of publications.

This scheme is the result of a fully automated state-of-the-art clustering algorithm describing the structure of science.

This classification may be used on different levels for our approach, depending on the level of granularity needed. For individual persons a higher resolution is needed than for universities.

In the case we present, we will use the intermediate 'meso-classification' (672 clusters) to identify benchmarks for a university. The 672 fields together represent the entire structure of science, which can be visualized in a landscape as depicted below.

Figure 1. Overview map of all sciences, using 672 fields in from a publication based classification (Waltman & van Eck, 2012)

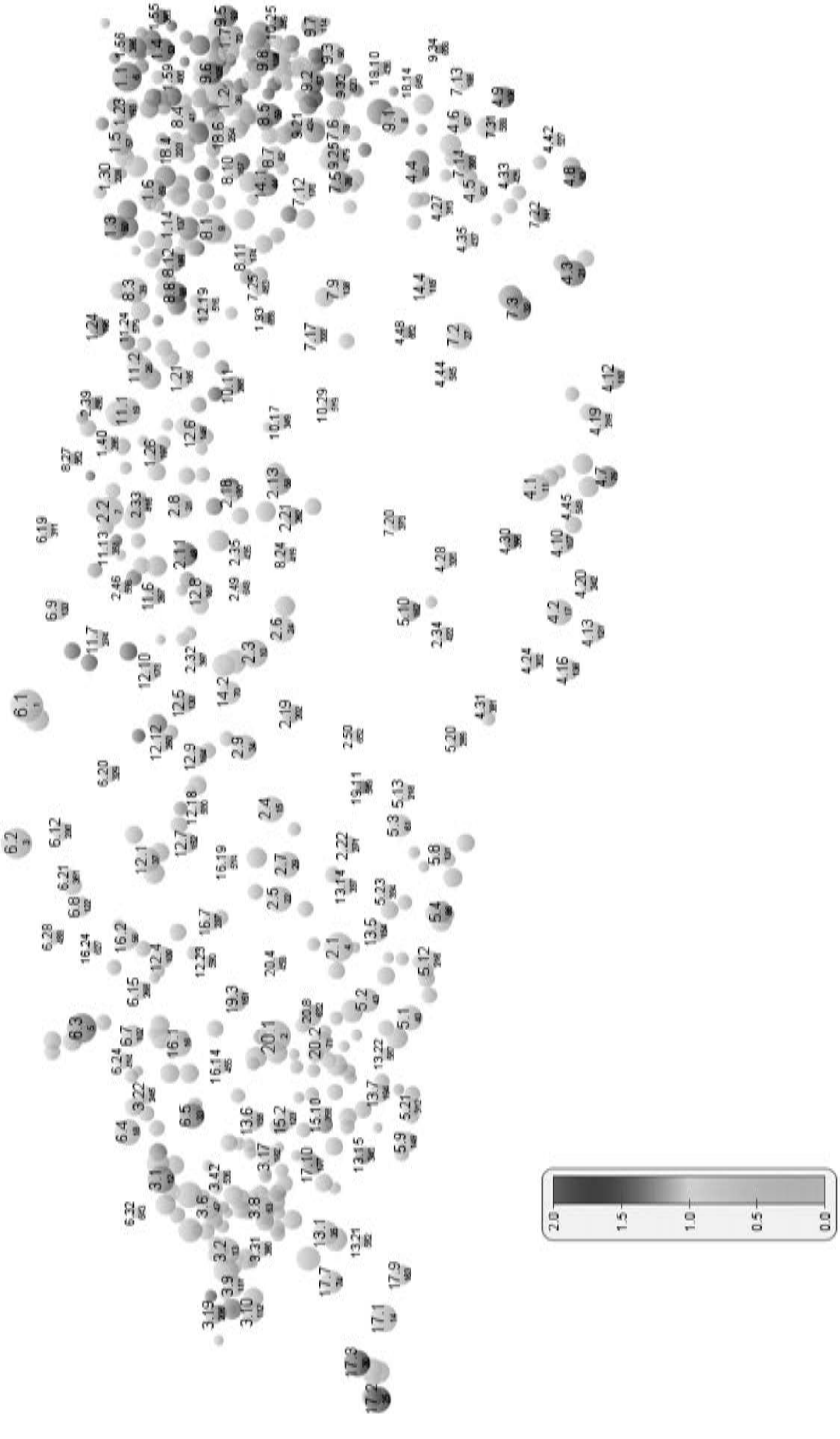


This structure of all sciences will be used to characterize the research profile (portfolio) of research units (universities).

Case of Leiden University

Leiden University is the focal unit in this case. This university has around 25,000 publications in the WoS in 2001–2010. The map below depicts its output distribution over the 672 fields. The color-coding indicates its relative contribution. The size of each circle represents the total size (number of publications) of a field. The relative contribution of a research unit depends on its overall contribution. If a units overall contribution in all sciences (all fields together) is 1%, a contribution of 2% in one particular field means a relatively high contribution. If the contribution is 0.5%, the relative contribution is low. The red areas represent fields in which this Leiden University output contribution is relatively high. In the green fields its contribution is normal (like overall), while in the grey areas its contribution is relatively low.

Figure 2. Output distribution of Leiden University over 672 research fields (2001–2010)



We apply a first filter to identify candidate benchmarks by defining the core of the Leiden University research fields. The core is defined by the fields in which Leiden has 50 papers or more (130 of the 672). We consider this the core of the Leiden research portfolio. Leiden has around 16,000 papers in this core (60%). The core represents mainly: biomedical sciences (30%), medical sciences (25%) cognitive sciences (10%), physical sciences (20%) and environmental and social sciences (15%).

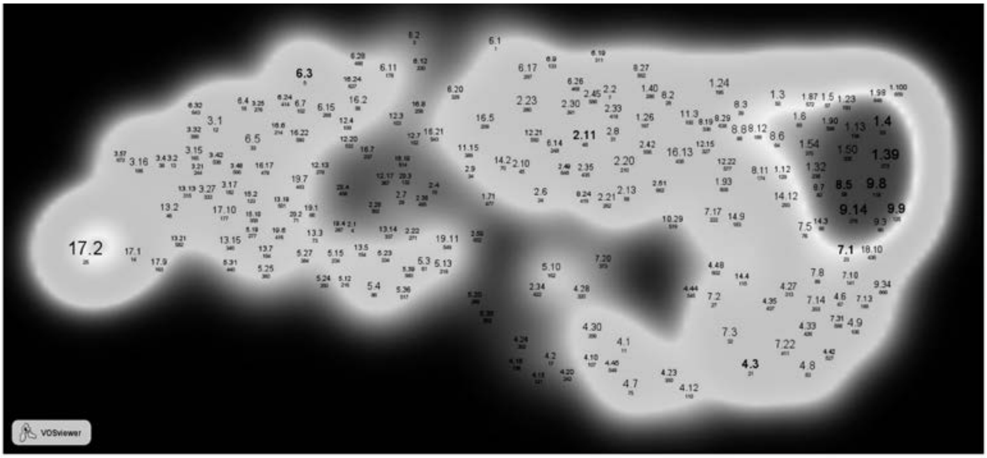
Subsequently, candidate benchmarks should have an output of more than 5000 papers (one third of Leiden) within the same core. This filter yields 180 universities from the 500 in the Leiden Ranking. Subsequently we used the complete distribution of these 180 over the 672 fields to match the distribution of Leiden. The distributions are compared among the 181 universities with the *cosine similarity* measure (c.f., Tan, Steinbach and Kumar, 2005). In table 2, we listed the most similar universities from the Leiden ranking as well as the least similar.

Table 2. Most similar and least similar universities to the Leiden University research profile

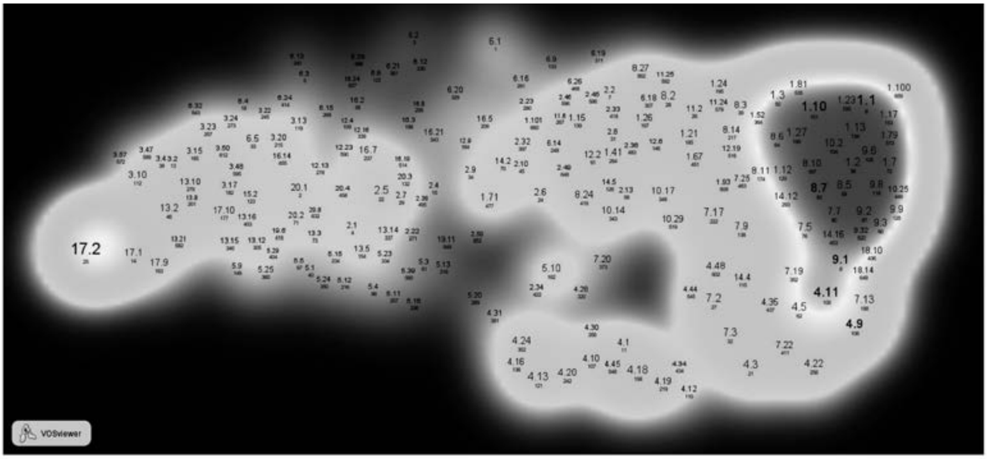
Similarity complete oeuvre	Similarity in Core	University	Npubs in Core	N pubs	Share in Core
1.00	1.00	Leiden univ	16,451	24,446	0.67
0.74	0.82	USA – Harvard Univ	51,399	115,901	0.44
0.70	0.80	NL – Univ Groningen	9,700	23,323	0.42
0.67	0.77	CANADA – Univ Toronto	24,787	64,207	0.39
####					
0.29	0.46	SG – Natl Univ Singapore	6,650	32,763	0.20
0.29	0.49	JP – Tohoku Univ	9,512	40,500	0.23
0.26	0.42	RUSSIA – Moscow Mv Lomonosov State Univ	5,106	25,032	0.20
0.25	0.37	CHINA – Zhejiang Univ	5,625	31,271	0.18

The top end of the table shows a strong similarity of *Harvard*, *Groningen* and *Toronto* with *Leiden* and a weak similarity at the bottom of the *Natl Univ Singapore*, *Moscow Mv Lomonosov State Univ* and *Zhejiang Univ*. The latter 4 universities have a substantial amount of output outside the core, as the Share-in-Core is much lower than for universities at the top end. The density view of *Leiden*, *Harvard* en *Zhejiang* illustrates the similarity and dissimilarity most clearly. In these maps we used the output in each field to draw the density maps.

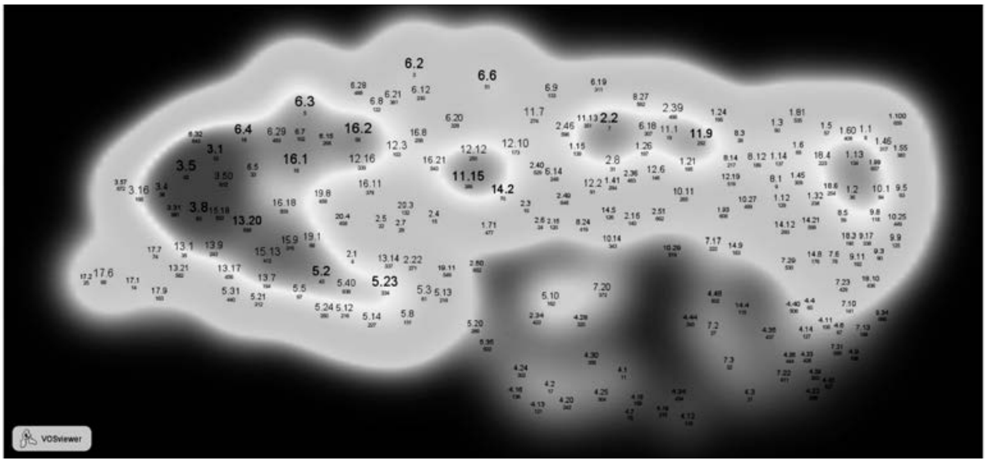
Figure 3. Density maps (VOSviewer) for Leiden, Harvard and Zhejiang universities



Leiden University



Harvard University



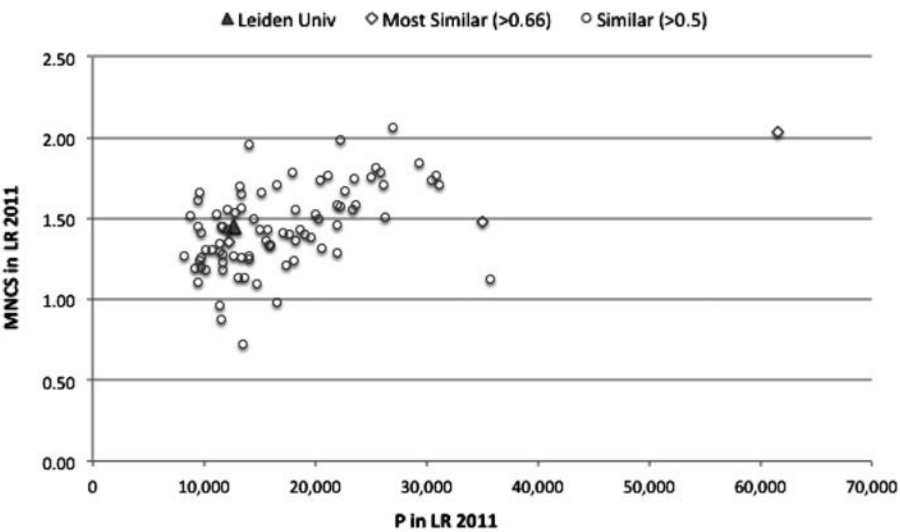
Zhejiang University

These pictures clearly illustrate how the profile of *Harvard* resembles the *Leiden* profile in Biomedical, medical, cognitive, social sciences and physical sciences. And on the other hand they illustrate how *Zhejiang* differs from Leiden with much less focus on cognitive and social sciences and more on the physical sciences.

Positioning Leiden University among its benchmarks

From the set of 180 universities we identified, having more than 5000 papers in the core fields of Leiden University, we found 86 with a cosine similarity higher than 0.5. For Leiden and these 86 we collected the performance measures from the Leiden Ranking (2011)[i]. Hence we were able to position the focal unit (Leiden University) among the most similar universities in the world. We summarized the results in Figure 4 using output (P) and field-normalized impact (MNCS).

Figure 4. Positioning output (P) and impact (MNCS) of Leiden Universities among 86 most similar universities worldwide.



In the overview we discern Leiden University, the most similar 3 universities (Harvard, Groningen and Toronto) and the other 83 universities with a similar research profile. The overview shows that similarity is not related to size or impact. Moreover, it shows that Leiden is at the lower end regarding output and in the middle regarding impact. Hence we are more accurately able to estimate its performance rather than just measuring the normalized impact related to the world average.

Discussion and further research

In this paper we conceptualized the different types of bibliometric benchmark studies, identifying the "performance-based benchmark selection" as a robust type of benchmark study with a lot of interest but still not fully methodologically developed. We have developed a bibliometric methodology that offers practical solutions expanding the possibilities of bibliometric benchmark analysis with relevance for research evaluation and science policy. In the full version of the paper we will develop further this conceptualization and present more practical examples of the development and application of this bibliometric methodology. Within the Leiden Ranking we have a wealth of data to further explore the potential of the approach. In the present study we focused on a university with a broad scope. In the full papers we will also include specialized universities.

References

Andersen, B., & Pettersen, P.-G. (1995). *Benchmarking Handbook*. Chapman & Hal.

Feelders, A.J., and Daniels, H.A.M. (2001), A general model for automated business diagnosis. *European Journal of Operational Research* 130, 623–637.

Moriarty, J. P. (2008). A Theory of Benchmarking.

Murphy, P. S. (1995). From bibliometric to research productivity: benchmarks and self-assessment. ISSI '95: *Proceedings of the fifth biennial international conference of the International Society for Scientometrics and Infometrics*. River Forest IL: ISSI.

Noyons, ECM; Moed, HF; Luwel, M (1999). Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric study. *Journal of the American Society for Information Science* 50(2). 115–131

Tan, P.-N.; Steinbach, M; & Kumar, V. (2005). *Introduction to Data Mining*. Addison-Wesley. ISBN 0-321-32136-7

Thijs, B., & Glanzel, W. (2009). A structural analysis of publication profiles for the classification of European research institutes. *Scientometrics*.

Tijssen, R. J. W., Visser, M. S., & Van Leeuwen, T. N. (2002). Benchmarking international scientific excellence: are highly cited research papers an appropiate frame of reference? *Scientometrics*, 54(3), 381–397.

Waltman, L., Calero-medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., Eck, N. J. Van, Leeuwen, T. N. Van, et al. (2012). The Leiden Ranking 2011 / 2012 : Data Collection , Indicators , and Interpretation, 63(12), 2419–2432. doi:10.1002/asi

---

i http://www.leidenranking.com. The measures are based on a limited number of publication years (2005–2009)

Measuring performance in complex science, technology and innovation strategies: A sociotechnical approach to indicator design[1]

Sebastian Pfotenhauer pfotenh@mit.edu / MIT Technology & Policy Program and MIT Portugal Program, Massachusetts Institute of Technology, 77 Massachusetts Ave, E40-221a, Cambridge, MA 02139 (United States) / Harvard Program on Science, Technology and Society, Harvard Kennedy School, Harvard University, 79 John F. Kennedy St., Cambridge, MA 02138 (United States)

Danielle Wood dwood26@jhu.edu / Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218 (United States)

Dava Newman dnewman@mit.edu / MIT Engineering Systems Division and Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, 77 Massachusetts Ave, E40-375, Cambridge, MA 02139 (United States)

Abstract

National science, technology and innovation (STI) strategies have become increasingly complex. Their goals and mechanisms often include domains as diverse as education, research, technology development, institution-building, regulatory frameworks and broader cultural change. Traditional STI indicators are often ill-suited to capture this complexity and comprehensiveness, and typically focus on the level of cross-national benchmarking indicators (accepting a trade-off between comparability and depth of contextualization) or program assessment (using primarily easily obtainable input and outcome indicators or one-time summative snapshots). In this paper, we offer lessons on indicator development from the study of complex STI initiatives. Using emerging constructs from engineering systems and STS, and drawing upon two sets of cases – international university partnerships and collaborative satellite development – we identify three areas of improvement for STI indicators to better capture sociotechnical and dynamic aspects of complex STI initiatives: systems architecture, multi-level technological learning and sociotechnical imaginaries. These indicators help address the long-standing measurement gap between program assessment and national benchmarking indicators. We further argue that in light of the increased complexity and sociotechnical embedding, indicator-based monitoring is no longer an administrative by-product, but a full-fledged research task on sociotechnical systems analysis and indicator development, where measurement instruments must be co-produced with the program architecture, an for which a researcher position should be made compulsory above a certain program size.

1 This work was supported by the MIT Leading Technology & Policy Initiative, the MIT Portugal Program, National Science Foundation Program on the Science of Science and Technology Policy (SciSIP), National Defense Science and Engineering Graduate Fellowship and the MIT Carroll Wilson Fellowship.

Introduction

Over the past decades, national and regional policy initiatives in science, technology, and innovation (STI) have continuously grown in complexity and comprehensiveness. Countries around the world are venturing boldly into technology sectors that were previously pursued by few advanced economies only, e.g. in space (Harvey, Smid, & Pirard, 2010) or nuclear technology (IAEA, 2010); they are designing regional clusters or whole inno-cities (Braczyk, Cooke, & Heidenreich, 2004; Breschi & Malerba, 2005; Marceau, 2008), strategies for research excellence (e.g. Fischer, 2010; Nature, 2006)), or deploying new technologies at scale in test-bed projects. This development towards complex initiatives is driven, on the one hand, by the increasingly systemic nature of technology itself (Weck, Roos, & Magee, 2011). On the other hand, it is owed to our increasingly systemic understanding of STI, acknowledging the multiple activities and required coordination between actors and institutions that go into successful STI strategies. Frequently, STI strategies include aspects of education, research, finance, institutional change, legal and policy frameworks, and – frequently – cultural idiosyncrasies (Fealing, Lane, Marburger, & Shipp, 2011; Lundvall & Borrás, 2006; Lundvall, Bengt-Åke, 1992; Nelson, Richard, 1993). Moreover, STI policies are facing pressures of globalization. As the world moves towards knowledge-based "learning economies," access to central knowledge hubs, the rapid diffusion of information, the circulation of skilled human capital, and cross-border collaboration in science and innovation, and benchmarking against global best practices have become more important than ever (Archibugi & Lundvall, 2001; Gibson & Heitor, 2005). STI strategies, then, take place on multiple levels, from institutional to international, and their spatial and temporal effects are not easily captured through indicators in a comprehensive way.

From a policy standpoint, our understanding of the effectiveness of such complex STI strategies is limited by a long-standing measurement gap. On the one hand, national STI achievements are measured primarily via systems-level benchmarking indictors, which inevitably accept a certain level of crudeness and homogenization for the sake of enabling comparison. On the other hand, individual STI initiatives are evaluated on the level of program assessment, which frequently focuses on easily obtainable input/output indicators or one-time post-hoc snapshots, thus missing dynamic and broader systemic aspects. Despite tremendous progress in benchmarking indicators (OECD, 2010) and evaluation research (Rossi, Lipsey, & Freeman, 2003), however, few indicators straddle different levels to connect programmatic and systemic aspects, thus casting light on complexity, sociotechnical embedding, and the process character of STI strategies. In short, contrary to the visible and steady growth in complexity and comprehensiveness in STI strategies, indicator development has not followed suit.

In this paper, we present three directions for developing indicators that link the programmatic and the systemic levels. The paper is part of a broader, on-going effort to understand, model, and assess complex strategies for STI capacity building. It draws upon our theoretical and empirical work on Complex International Innovation Partnerships (CIIPs), which will be developed in detail elsewhere. Here, we look specifically at the question of how sociotechnical complexity in STI strategies can inform indicator development and measurement.

Our approach acknowledges that STI strategies are thoroughly socio-technical – that is, their social and technical components cannot be easily disentangled for analytic or policy purposes. Research on sociotechnical systems has been most notably advanced by two fields. *Engineering Systems*, on the one hand, emphasizes how technology itself has become increasingly systemic, and how scientific, technical, social, organizational, and policy aspects need to combined in the design and pursuit of STI initiatives (Weck et al., 2011). *Science and Technology Studies*, on the other hand, has put much emphasis on the social construction of technological systems and how STI and social order are co-produced (Bijker, Hughes, & Pinch, 1989; Jasanoff, 2004).

Cases

Our paper draws upon research on complex STI initiatives in two sectors. The first sector is international university collaborations for innovation leverage. In these collaborations, governments enter into strategic partnerships with eminent international research universities to help build STI capacity in their domestic university systems by way of best-practice transfer. For our study, we chose four international partnerships of the Massachusetts Institute of Technology (MIT), which make use of different modes of collaboration. Collaboration modes include the creation of new universities (such as Masdar Institute of Science and Technology or the Singapore University of Technology and Design), the up-grading of existing universities (such as the MIT-Portugal Program or the Singapore MIT Alliance), and ecosystem enhancement of top-tier universities (such as the Cambridge-MIT Institute).

The second set of cases features collaborative satellite development projects in developing countries. In these projects, governments team with foreign firms or institutions to acquire the technological capability necessary to design, fabricate, test and operate satellites. We study in detail several collaborative satellite development projects from Africa, the Middle East and Southeast Asia. Expert partner organizations have come from the United Kingdom, China, France, South Korea and the Ukraine. The partners host customer engineers for months or years and provide training to customer engineers while developing earth observation satellites.

The current paper is too limited in space to discuss our cases in depth (for more details, see Pfotenhauer, Roos, & Newman, 2013, and Wood & Weigel, 2012). A few remarks must suffice to illustrate, in a nutshell, why we chose the cases and why we believe they are both important and fertile subjects of study from an indicator perspective. First, both sets of cases represent national flagship initiatives in STI policy, which are the very initiatives that governments are most interested in assessing. Second, they are sociotechnically complex and typically large-scale endeavors, involving a diverse mix of technologies, knowledge assets, and institutional structures. They pursue multiple objectives in science, technology, education, institution-building, and cultural change, involve hundreds of people, and typically cost on the order of ten to one hundred mission dollars over a lifespan of 5 to 20 years. They are collaborative initiatives between multiple institutions located in different countries. Third, they are driven by public sector interests

addressing economic, infrastructural and societal needs, and include the government as a major actor.

Indicators for complex international STI strategies – a socio-technical approach

This paper proposes a range of emerging constructs from Engineering Systems and STS as a valuable addition to traditional STI indicators in assessing complex STI initiatives. In the following, we will address indicators for three areas: systems architecture, technological learning, and socio-technical imaginaries.

Architectural Models

Utility of the indicator: Systems architecture is an approach that uses models to understand, design, and manage complex systems while remaining cognizant of their emergent properties. Architectural analysis makes explicit the strategic decisions within the evolution of a system. It shows how alternative approaches align or conflict with the objectives of stakeholders and highlights trade-offs between multiple objectives. Architectural models elucidate the influence of context and stakeholder objectives on a system. The models also identify the forms through which the system performs functions to meet stakeholder objectives. Finally, the model groups functions into categories that represent relevant stakeholder views of the system.

Application: We create architectural models summarizing the design and execution of complex STI initiatives. For the sake of brevity, only cases of collaborative satellite development projects will be discussed. The architectural models for six collaborative satellite development projects inductively define a set of twelve architectural views that capture the major decisions involved in these systems. Examples of Architectural Views include Organization, Supplier Selection, Training, Technical Approach, and Policy. Each Architectural View includes a set of related functions and the options among forms that can execute the function. In each case, the actors select one of the forms based on their context and stakeholder objectives. As decision makers in these systems learn more, they are better able to select the forms that align well with stakeholder objectives. This is a challenge, however, because the selection of various forms to execute one function may create conflicts or constraints with other forms.

Table 1 exemplarily displays three of the functions within the Supplier Selection Architectural View, capturing several decisions that are part of the larger activity of identifying a firm that will work as the primary contractor to provide the satellite and training program. The figure shows alternative forms for Choosing the Satellite Supplier, alternative attributes that execute the function of Differentiating among Suppliers and types of firms that were Competing for the Supplier Contract. The analysis reveals that the various nations took distinct approaches, choosing for example different forms for the function of Choosing a Satellite Supplier. Some nations relied on

formal selection processes, involving well-defined criteria, documentation and bureaucratic transparency. Others relied on informal relationships. Both selection strategies may be appropriate depending on contextual factors such as the level of political support for the satellite project and the technical requirements for the final product.

Table 1. Examples of System Functions within the Supplier Selection Architectural View

Supplier Selection View							
Generic Forms	**Function**	**Examples of Forms from Existing Projects**					
Supplier Selection Process	Choosing satellite supplier	Choose personal acquaintance	Join invitation for collaboration	Call for selective Tendering	Hire Consultant to Review	Open Call for Proposals	Travel to four international suppliers
Priority Supplier Attributes	Differentiating among suppliers	Technical performance and flexibility	Training package	Space heritage	Price	University Relationship	Schedule
Competing Suppliers	Competing for Supplier Contract	Government Space Agencies	Small Commercial Firm	Medium Commercial Firm	Large Commercial Firm	State owned Enterprise	

Multi-level technological learning

Utility of the indicator: Learning plays a key role in STI (Archibugi & Lundvall, 2001; Gibson & Heitor, 2005; Kim, 1997; Mani & Romijn, 2004). Yet, learning has proven difficult to quantify, and the research approach varies with the unit of analysis. We begin by distinguishing different types of learning relevant to STI strategies based on the *level* of learning:

- *Individual learning* is the most basic form of learning that refers to the acquisition of knowledge or skills that enable the execution of a certain task relevant to the STI project. These skill sets may for example include technical, analytic, managerial or inter-personal skills.

- *Group learning* occurs when small groups learn how to combine knowledge and skills to achieve tasks that require more than individual effort. Frequently, group learning deals with the transfer of "tacit knowledge," either by making this tacit knowledge explicit within the group or through observation and socialization (Edmondson, Winslow, Bohmer, & Pisano, 2003).

- *Organizational learning* (Nonaka, 1994), has two relevant sub-forms:
  - *Institutional learning* refers to the formation of operational routines, which are created and codified over time, representing a form of organizational memory and allowing a learning organization to gradually move to higher performance.
  - *Program learning* is related to what the educational literature describes as "formative assessment." The goal is to assess and adjust program performance while it is running. It strategically uses program data for real-time decision-making and effective governance.

- Systems learning describes how a regional or national system of innovation improves by identifying and refining cultural norms.

We further distinguish between short- and long-term learning efforts. At each level, short-term learning describes a specific learning episode, such as the process for an individual to learn a specific skill or for a cycle of program evaluation. Long-learning describes the impact over time as the learning process brings individuals, groups, institutions, programs and systems to new stages of capability or autonomy. Table 2 summarizes the different types of learning.

Table 2. Multi-level technological learning

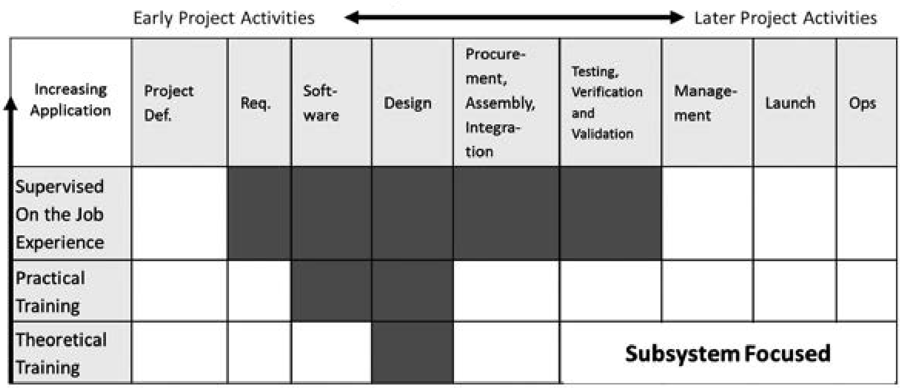
			Organizational		
Learning	**Individual**	**Group**	**Institutional**	**Programmatic**	**Systemic/National**
Short Term	Individuals learning new knowledge and skills	Individuals within a small group learning how to combine skills to achieve a task; Individuals converting tacit knowledge to explicit knowledge by communicating	Establishing a new institution or achieving new institutional milestones	During a program, observing progress toward goals; making changes to ensure progress toward goals	Recognizing the cultural norms of a system and adopting new norms in a narrow context.
Long Term	Individuals increasing autonomy in executing tasks based on knowledge	Individuals within a small group learning how to leverage contributions of each member	Individuals and groups within an institution communication knowledge across units	During a program, observing both goals and progress toward goals; changing goals as needed	Gradual cultural change at the level of a national system that leads to new norms

Application: All levels and timescales of learning are relevant to assess how STI initiatives can be applied to the acquisition and maintenance of STI capabilities. For the sake of brevity, we discuss three instances of measuring learning, taken from both sets of cases: short-term individual learning (satellites), short-term program learning (universities), and a case of long-term systemic learning (universities).

Example 1: Short-term individual learning. The case studies of collaborative satellite development projects develop frameworks to assess progress in technological learning by engineers that are trained within foreign firms. Table 3 is an example of a Capability Building Framework applied to

one engineer that illustrates short term, individual learning. The horizontal axis shows the phases of the satellite development lifecycle in which the engineer participated. The vertical axis shows the level of autonomy that the engineer achieved – ranging from theoretical training to supervised, on the job experience. This engineer experienced all the phases of the satellite lifecycle that occur between Requirements and Testing/Verification/Validation. The engineer received theoretical training that primarily addressed design aspects of satellite engineer and later worked in supervised, on the job training on a variety of topics. The engineer moved over time toward greater levels of autonomy. Because this engineer worked primarily on the details of one section of the satellite, their experience can be characterized as subsystem focused rather than system focused. A system focused engineer would have more activity on either end of the horizontal axis. Understanding how to specify and evaluate the learning experience in these concrete terms can help CIIP managers better define training programs.

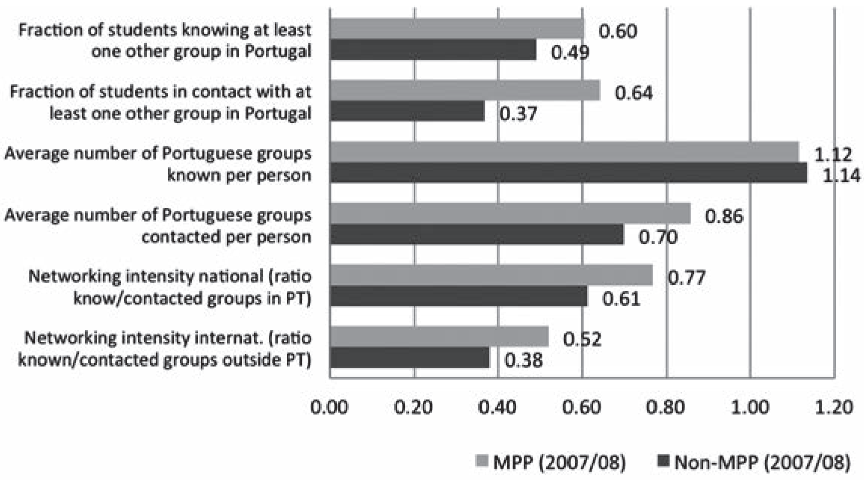
Table 3. Individual Capability Building, this figure shows one engineer's achievement during a project



Example 2: Short-term program learning. Short-term program learning is, in many ways, closely related to classical program assessment, particularly to times series approaches and what is know as "formative assessment." Figures 1 & 2 show some snapshot results of a student survey implemented in the MIT-Portugal Program, designed foster program learning in two ways. First, it was designed as a comparative survey measuring program performance against its closest peer programs, i.e. other Portuguese graduate programs in engineering, thus allowing the program leadership to identify the relative performance and impact of the program. Second, the approach was expanded into time series by conducting multiple iterations of the same survey of the course of the program duration, which allow tracking the improvement (or learning) of the program over time. For the sake of brevity, we only discuss the first (comparative) results here. For a more exhaustive treatment, cf. (Pfotenhauer, Jacobs, Pertuze, Newman, & Roos 2012).

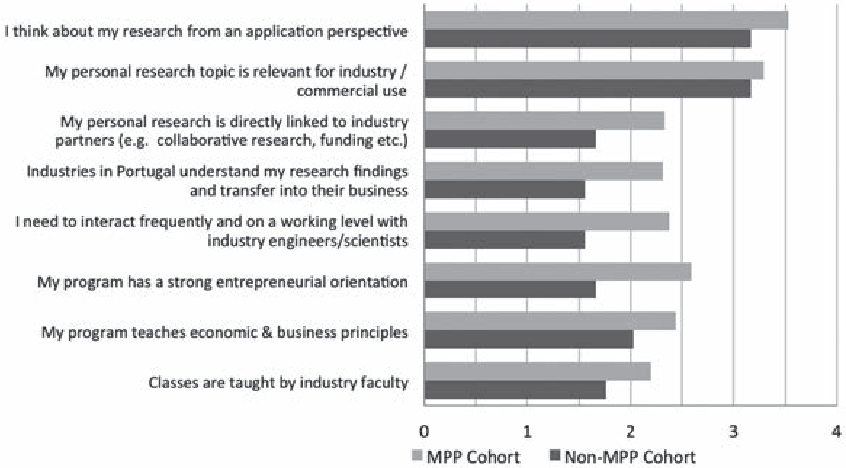
Figure 1 captures the connectivity of students. The data shows that MIT-Portugal students on average are more aware of, and in contact with, other research groups working on the same field both inside and outside Portugal. Figure 2 captures self-reported industry linkages of students. Here, MIT-Portugal students report among other things a higher industry linkages, more frequent interaction with industry partners, and a higher entrepreneurial orientation of their program. Importantly, the surveys also revealed program weaknesses, which resulted in important adjustments in the program implementation, including curriculum structure or the underperformance of one focus area compared to the others.

Figure 1. Connectivity of students



(Source: Pfotenhauer, Jacobs, Pertuze, Newman, & Roos, 2012)

Figure 2. Self-reported industry linkages of students



(Source: Pfotenhauer, Jacobs, Pertuze, Newman, & Roos, 2012)

Example 3: Long-term systems learning. Little work has been done on how STI initiatives affect their surrounding systems (i.e. those system components that are not directly involved in the STI initiative); yet, this information is crucial if one wants to understand whether and how a program is able to seeding larger-scale institutional or national change, or even cultural paradigm shifts. In order to understand this aspect, we conducted an analysis of spillover effects for university partnerships. Table 4 discusses spillover pathways for one collaboration (the MIT-Portugal Program) on four different levels: On an individual level, participants of the program may begin to apply their knowledge and experience outside the program, such as faculty teaching other courses, or administrators benefitting from newly acquired administrative capacities. On a programmatic level, MIT-Portugal curricula may serve as blueprints for newly developed graduate programs Portugal. On an institutional level, we found that MIT-Portugal has stimulated the self-formation of an energy research cluster at one university that was previously not part of the program, with the explicit goal to join the program. On a systemic level, program activities were designed that explicitly address gaps in the innovation system and are open to the whole region/country. For example, MIT-Portugal created an annual student Venture Competition, the first of it's kind in Portugal. Participants come not only from MIT Portugal students and local universities, but increasingly from national and international teams, thus creating a systemic effect.

Table 4. Channels of spillover across program boundaries

Level	Mechanism	Effect
Individual	Faculty teaching	– Faculty teaching practices influenced by the program; teach differently also outside the program – Non-program faculty approach program faculty
Individual	Administrative capacity	– Formation of novel capacities on cross-institutional and cross-sectoral networking, curricular modularization, internationalization, industry linkages, and mobility
Programmatic	Curricular blueprints	– Program courses and curricula serve as blueprints for the design of new educational programs in Portugal
Institutional	Clustering & self-formation	– Stimulation of self-formation of energy cluster in Portugal with the explicit goal to create critical mass and participate in the program.
Systematic	Fostering entrepreneurship	– Creation of first Venture Competition in Portugal – Significant growth in scale and scope over time, including non-MIT-Portugal participants (national and international)

Sociotechnical imaginaries

Utility of the indicator: From a sociotechnical perspective, the reduction of national performance to common benchmarking indicators always comes with a risk – the risk of missing important social, cultural, institutional, and historical specifics, and thus possibly the strongest explanatory factors. Over the past years, several advances from the field of Science and Technology Studies have been made to capture and formalize these idiosyncrasies through rigorous theoretical frameworks. One prominent approach is that of "sociotechnical imaginaries," which Jasanoff (2009) defines as "collectively imagined forms of social life and social order reflected in the design and fulfillment of nation-specific scientific and/or technological projects." It identifies persistent patterns in the decision-making surrounding national sociotechnical projects and traces them back to deeply ingrained social and institutional configurations. This analytic lens reveals the shared, higher-order assumptions and motives embedded in a society, expressed in how key actors imagine the purpose, realization and effects of STI.

Application: A sociotechnical imaginaries analysis was carried out for the case of international university collaborations. Again for the sake of brevity, we only present the high-level results in Table 5:

Table 5. Sociotechnical imaginaries of innovation embodied by innovation partnership

	Cambridge-MIT Institute	MIT-Portugal Program	Masdar Institute of Science and Technology	Singapore-MIT Alliance / Singapore University of Technology and Design
Structural particularity of MIT partnership	– Bilateral hub & spokes model: partnership between two "world-class" institutions of equal standing; involve affiliates – Focus on commercializing existing top research – Minimal institutional invasiveness	– Network model: consortium including 8 universities & 20 research centers; UT Austin, CMU – Focus on human resource development and institutional change – Create critical mass through network	– Institutions-building model: Build new graduate research university from scratch – Focus on "jump-starting" the ecosystem and solve fundamental challenges for national economic development	– Growth and expansion model: Gradually intensify scope of relationship and ambitions, from education partnerships to research centers to full university
University system	– Old, centered on elitist Oxbridge core – Strong university research	– Old "delayed" by dictatorship until 1974; no single eminent university	– Very young – Universities hitherto have played no role in the economy	– Young, ambitious, rapidly expanding – Core part of Singapore's identity and competitiveness strategy
Sociotechnical imaginary	– Maintaining global leadership – Societal transformation occurs through elite institution: "Upgrading" Cambridge will fix a national problem	– Catching-up through outward orientation and adaptation – Societal transformations occurs with social consensus and broad institutional base – Skepticism towards governmental leadership	– Economic, social and political interests overlap – Maintaining energy leadership – Commodity attitude: Buying the best education, research, innovation infrastructure	– Central hub of the region, thriving through extreme outward orientation – Highly technocratic and progressivist
Implicit innovation model	– Linear: Building the rear end of the innovation pipeline; create add-on to existing strong front end	– Networked "learning economy:" Connect Portuguese universities to leading innovation hubs and to one another	– "Condensed" triple helix: governmental, industrial and academic interests are identical	– Central technocratic industrial policy with high systems integration

Conclusion

In this paper, we have argued that a thorough sociotechnical and process-based understanding is required to capture the characteristics of current large-scale STI initiatives. Based on a study of international university partnerships and collaborative satellite development projects, we discussed three approaches that can help improve measurement to capture the systemic, socio-technical properties of complex STI initiatives: architectural views, multi-level technological learning, and sociotechnical imaginaries. Several more general lessons emerge:

- Adequate measurement for STI capacity building needs to straddle the gap between program assessment and national systems-level indicators. Neither decontextualized cross-country benchmarking nor the frequently static, post-hoc program assessment approaches capture the essence of STI initiatives today.

- Just like there are no general templates for constructing effective STI initiatives across countries, there are limits to one-size-fits-all indicators. Adequate metrics for complex STI strategies cannot be fully developed ahead of time, but have to be co-produced with the initiative itself based on its context, co-evolve with the program, and provide real-time feedback for program learning.

- Consequently, indicator development should be seen as an integral part of STI program design and execution. This exceeds standard assessment practices typically carried out by program administrators.

- We argue for a research-based definition of "program assessment" in STI policy. We recommend that a research position with focus on indicator development and program assessment be created as part of every innovation initiative above a certain size to account for the unique, quasi-experimental character of these initiatives. The researchers would be core members of the program team with full right to observe, interview, conduct surveys and review documents over the program lifetime. The strategic and material benefits of effective in-depth program assessment as discussed in this paper justify, in the opinion of the authors, the costs of such a research position manifold.

References

Archibugi, D., & Lundvall, B.-Å. (Eds.). (2001). *The Globalizing Learning Economy*. New York: Oxford University Press.

Bijker, W., Hughes, T. P., & Pinch, T. (1989). *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. MIT Press.

Braczyk, H.-J., Cooke, P., & Heidenreich, M. (Eds.). (2004). *Regional Innovation Systems: The Role of Governances in a Globalized World* (2nd ed.). Routledge.

Breschi, S., & Malerba, F. (2005). *Clusters, Networks, And Innovation*. Oxford University Press.

Edmondson, A. C., Winslow, A. B., Bohmer, R. M. J., & Pisano, G. P. (2003). Learning How and Learning What: Effects of Tacit and Codified Knowledge on Performance Improvement Following Technology Adoption. *Decision Sciences*, 34(2), 197.

Fealing, K., Lane, J., Marburger, J., & Shipp, S. (Eds.). (2011). *The Science of Science Policy: A Handbook* (First Edition.). Stanford Business Books.

Fischer, K. (2010, September 12). Portugal Aims to Modernize With Help From the U.S. The *Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/Portuguese-Universities-Turn/124364/

Gibson, D., & Heitor, M. (Eds.). (2005). *Learning and Knowledge for the Network Society*. Purdue University Press.

Harvey, B., Smid, H. H. F., & Pirard, T. (2010). *Emerging Space Powers: The New Space Programs of Asia, the Middle East and South-America* (2010th ed.). Praxis.

IAEA. (2010). *International Status and Prospects of Nuclear Power* (2010 Edition). Vienna: International Atomic Energy Agency. Retrieved from http://www.iaea.org/Publications/Booklets/NuclearPower/np10.pdf

Jasanoff, S. (2004). Ordering Knowledge, ordering society. In *States of Knowledge: The co-production of science and social order* (13–45). New York: Routledge.

Jasanoff, S., & Kim, S.-H. (2009). Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea. Minerva, 47(2), 119–146. doi:DOI: 10.1007/s11024-009-9124-4

Kim, L. (1997). *Imitation to Innovation: The Dynamics of Korea's Technological Learning*. Boston, MA: Harvard Business School Press.

Lundvall, B.-Å., & Borrás, S. (2006). Science, Technology and Innovation Policy. In J. Fagerberg, D. C. Mowery, & R. R. Nelson (Eds.), *The Oxford Handbook of Innovation* (1st ed., Vol. 1, 599–631). Oxford, England: Oxford University Press.

Lundvall, Bengt-Åke. (1992). *National Innovation Systems: Towards a Theory of Innovation and Interactive Learning*. London: Pinter.

Mani, S., & Romijn, H. (Eds.). (2004). *Innovation, Learning, and Technological Dynamism of Developing Countries*. United Nations.

Marceau, J. (2008). Introduction: Innovation in the city and innovative cities. *Innovation: Management, Policy & Practice*, 10(2–3), 136–145. doi:10.5172/impp.453.10.2-3.136

Nature. (2006). New competition in Germany. *Nature Neuroscience*, 9(6), 711–711. doi:10.1038/nn0606-711

Nelson, Richard (Ed.). (1993). *National Innovation Systems. A Comparative Analysis*. New York/Oxford: Oxford University Press.

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science*, 5(1), 14–37.

OECD. (2010). Measuring Innovation: A new perspective. Paris: OECD.

Pfotenhauer, S. M., Jacobs, J. S., Pertuze, J. A., Newman, D. J., & Roos, D. T. (2012). Seeding Change through International University Partnerships: The MIT-Portugal Program as a Driver of Internationalization, Networking, and Innovation. *Higher Education Policy*. doi:10.1057/hep.2012.28

Pfotenhauer, S., Roos, D., & Newman, D. (2013). International University Collaborations as an Innovation Strategy for Transitioning Countries: A study of MIT's major international partnerships. *forthcoming*

Rossi, P. H. (Henry), Lipsey, M. W., & Freeman, H. E. (2003). *Evaluation: A Systematic Approach* (7th ed.). Sage Publications, Inc.

Weck, O. L. de, Roos, D., & Magee, C. L. (2011). *Engineering Systems: Meeting Human Needs in a Complex Technological World*. The MIT Press.

Wood, D., & Weigel, A. (2012). Architectures of Small Satellite Programs in Developing Countries. Presented at the International Astronautical Congress, Naples, Italy.

Ranking universities: The challenge of affiliated institutes

Fleur Praal f.e.w.praal@cwts.leidenuniv.nl / Centre for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62-A, Leiden, 2333 AL (The Netherlands)

Joost Kosten kostenmjf@cwts.leidenuniv.nl / Centre for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62-A, Leiden, 2333 AL (The Netherlands)

Clara Calero-Medina clara@cwts.leidenuniv.nl / Centre for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62-A, Leiden, 2333 AL (The Netherlands)

Martijn S. Visser visser@cwts.leidenuniv.nl / Centre for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62-A, Leiden, 2333 AL (The Netherlands)

Abstract

On a theoretical level, the definition of a university has not been uniform, which has led university rankings such as the Leiden Ranking to include institutions with diverging missions, objectives and organizational structures. This issue cannot be solved, yet university rankings should aim to maximize the comparability of the institutions included. An important aspect that should be addressed in that perspective is the myriad of possible variations in relations between universities and other research institutions: this is usually regulated on national or regional level, but international systemic variation severely complicates worldwide comparison. This paper explores our efforts in dealing with university hospitals, as examples of organizational structures usually regulated nationally or regionally, varying greatly worldwide, and producing much research strongly represented in the Web of Science. We diagrammatically represent two types of hospital systems: a symbiotic and a networked system. In the consecutive case study, English universities and hospitals from their national networked system are subject to two methodologies of address unification: a strict, exclusive method and a more relaxed, inclusive one. The results demonstrate that the methodologies applied directly affect the Leiden Ranking. This emphasizes that a thought-out and transparent data editing methodology is necessary in compiling university rankings.

Introduction: Academic systems

Higher education systems worldwide are in a constant state of reorganization, partly in response to university rankings (Dill; 2006; Dill & Soo, 2005). These reorganizations in turn pose a challenge for university rankings, as the definition and delimitation of universities, which have not been universal, are now subject to constant change. In this paper, we will explore the challenges for university rankings posed by international variance in the organisation of academic systems. We will particularly focus on our current work with one type of institutions notoriously hard to delimitate: university hospitals (Barrett, 2008: 804). Data from a case study of English universi-

ties and hospitals will demonstrate the considerable scope of the issue, and therefore we must conclude that it merits more attention.

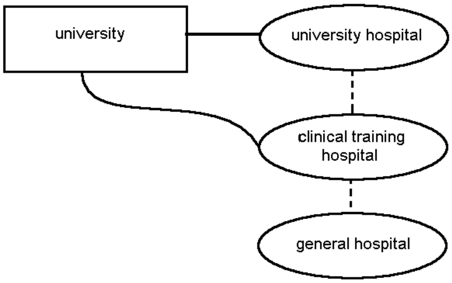
Relevant criteria defining a university include education, research intensity, activity in a broad range of fields, and doctorate-granting authority; but these are not universally applied and are currently changing (Carnegie Foundation, 2013; Birtwistle, 2003: 230–1)[i]. Based on a combination of criteria, university rankings currently include various types of institutes with different missions and tasks. It is necessary to create a classification of institute types for international comparison, or more pragmatically, to explicate the dimensions now implicitly marking the boundaries of organizations in scientometric studies (Hardeman, 2013). However fundamental, this issue is theoretical and beyond the scope of this practically-oriented paper.

The universities currently included in the rankings are therein represented as well demarcated, homogeneous entities that are universally comparable, whereas the international reality shows intricate heterogeneity in organizational structures. Universities increasingly present themselves as networked organizations rather than monolithic entities; faculties, as the core of the university, are related to research institutes, academies of science and other universities through the exploitation of research centres. Usually, such systems are organized by national or regional government regulations, which allows for comparison between universities in one such system. The comparison of units within different systems, however, is fundamentally problematic. Within the multitudinous complex relations, we focus on the affiliations of universities with hospital trusts in the exploitation of hospitals. Since medical research in particular is quantitatively very productive and strongly represented in the Web of Science, these affiliations are consequential in university rankings and therefore further investigation is urgent.

Focus on university hospitals systems

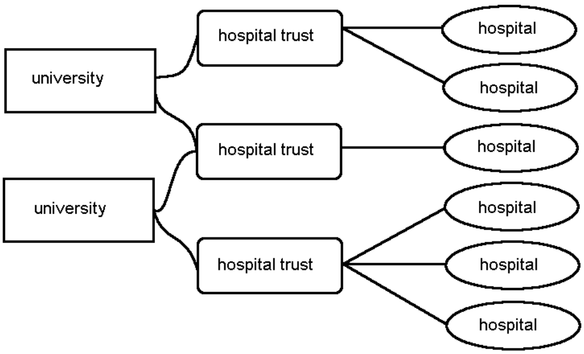
University hospitals are traditionally broadly defined. Their common denominator is that they combine research, education and care by combined employment of staff (Peart, 1970; Davies et al., 2010: 1091; Lewkonia, 2002: 289). In some systems, universities with a school or faculty of medicine have a symbiotic relationship with one academic medical centre. University and hospital then have integrated research and teaching departments, but can be independent legally and organizationally. This exclusive relation between a university and a hospital can be formalised in regulations or law, as is the case in, among others, the Netherlands and most states in Germany [Figure 1](WHO, 2010; Siewert & Merk, 2009).[ii]

Figure 1. Symbiotic relations between university and hospital.



In other systems, such as the English or Norwegian, academic hospitals are not so directly linked to universities, but organized in trusts, regulated by regional or national governance (WHO, 2006; 2011). One such trust can contain many hospitals, having non-uniform affiliations to universities. Moreover, multiple hospital trusts can be linked to one university; and many universities to the same hospital group [Figure 2].

Figure 2. Network of universities, trusts and hospitals.



Research from these networks is published under various labels: that of the trust or of the individual hospital, and with or without mentioning 'university', affiliated however closely. Increasingly and in various countries, multiple universities and many hospital groups integrate into larger complex networks, extending possible variation in practice (Darzi, 2008; WHO, 2006). In the United States, examples of both the symbiotic and the networked systems exist.

The complex and increasing variance of university hospital systems must be taken into account when preparing data for the compilation of university rankings. A case study with English university hospitals will further illustrate this.

Case study: English university hospitals

The relations between seven universities and hospitals in England, diagrammatically represented above [Figure 2], have been explicated below [Figure 3]. Although all the hospitals listed are closely related to universities via trusts, these relations are not necessarily reflected in their names. The table is exemplary for the variety of associations possible: relations between universities and hospital trusts are usually direct, but can be intermediated by a jointly governed Academic Health Science Centre (King's College) or medical school (Exeter/Plymouth); multiple universities can be related to multiple hospital trusts (Sheffield); and trusts can govern one, multiple, or zero hospitals (Oxford).

Figure 3. Complex affiliations between English universities and hospitals.

University	Medical School	Academic Health Science Centre	Affiliated Hospital Trust	Hospital
King's College, University of London		King's Health Partners	Guys & St Thomas' NHS Foundation Trust	Guy's Hospital St Thomas' Hospital
			King's College Hospital NHS Foundation Trust	King's College Hospital
			South London & Maudsley NHS Foundation Trust	Bethlem Royal Hospital Lambeth Hospital Lewisham University Hospital, Ladywell Unit Maudsley Hospital
St George's, University of London			St George's Healthcare NHS Trust	Queen Mary's Hospital St George's Hospital St John's Therapy Centre
			South West London & St George's Mental Health NHS Trust	Springfield University Hospital
University of Oxford			Oxford Health NHS Foundation Trust	
			Oxford University Hospitals NHS Trust	Churchill Hospital Horton General Hospital Nuffield Orthopaedic Centre John Radcliffe Hospital
University of Exeter	Peninsula Medical and Dental School		Plymouth Hospitals NHS Trust	Derriford Hospital Scott Hospital, Child Development Centre
			Royal Cornwall Hospitals NHS Trust	Royal Cornwall Hospital West Cornwall Hospital St Michael's Hospital
University of Plymouth			Royal Devon and Exeter Hospitals NHS Foundation Trust	Royal Devon and Exeter Wonford Hospital Heavitree Hospital
			South Devon NHS Healthcare	Torbay Hospital
Sheffield Hallam University			Sheffield Children's Hospital NHS Foundation Trust	Sheffield Children's Hospital
Sheffield Hallam University			Sheffield Teaching hospitals NHS Foundation Trust	Charles Clifford Hospitals Jessop Wing Northern General Hospital Royal Hallamshire Hospital Weston Park Hospital
University of Sheffield				

Figure 4 shows the outcomes of two different approaches for identification of these universities and their related medical institutions within the WoS. The difference between the two approaches indicates the bandwidth in which a university can be delimitated. The strict approach includes author affiliations that mention the name of the university or any of its constituent schools

explicitly, such as the Institute of Psychiatry within King's College. The relaxed approach furthermore includes relations to hospitals or trusts, for example Guy's Hospital, or Guys & St Thomas NHS Foundation Trust.

Figure 4. Number of WoS-covered publications 2005–2009 according to two different approaches for seven universities in England.

University	Strict approach	Relaxed approach	Difference
King's College, University of London	11,999	17,479	46%
St George's, University of London	1,017	3,267	221%
University of Oxford	24,466	29,488	21%
University of Exeter	3,932	5,708	45%
University of Plymouth	2,487	4,105	65%
Sheffield Hallam University	982	3,466	253%
University of Sheffield	11,372	13,002	14%

The two approaches can be considered as a lower and upper boundary for the bandwidth of delimitating these universities. This is substantial for all seven universities but varies significantly. St George's and Sheffield Hallam's output would more than triple in the relaxed, maximally inclusive approach; both would then have surpassed the threshold number of publications to enter in the Leiden Ranking 2011/2012; as would the University of Exeter. These figures clearly demonstrate that the handling of affiliated organizations importantly influences the compilation of university rankings.

Discussion

As the case study shows, the optional inclusion of university hospitals can significantly influence university rankings, due to the share of medical research in the total research output. In the compilation of a ranking, clear delimitations are therefore elementary to allow comparison of maximally similar types of 'research conglomerates'. Author affiliations are not self-explanatory in publications, nor is international practice homogeneous. Therefore publication data must be edited extensively before it can be used in a ranking, but clear-cut standards for this are lacking. As the methodology of editing can be crucially important, ranking organizations should be transparent about it, and about the limitation of their products (Waltman et al., 2012; Hardeman, 2013: 1187).

Since the start of the Leiden Ranking, we have been well aware of the challenges presented above, but the possible consequences have not been studied in detail yet (Van Raan, 2005: 137). Although the Leiden Ranking has always dealt with the problem of universities and related hospitals, the current approach is not fully satisfactory, because of limited transparency (cf. Waltman et al., 2012: 2424–2429). Therefore, the aim of our current research is to study alternative methodologies, and their consequences, to apply to the Leiden Ranking.

References

Barrett, D.J. (2008). The evolving organizational structure of academic health centres: The case of the University of Florida. *Academic Medicine*, 83, 804–808.

Birtwistle, T. (2003). What is a "university"? (The English Patient). *Education & The Law,* 15, 227–236.

Carnegie Foundation (2013). Methodology. Retrieved 30 May 2013, from: http://classifications.carnegiefoundation.org/methodology/basic.php.

Darzi, K.B.E. Lord of Denham (2008). High Quality Care for All: NHS Next Stage Review final report, Ministry of Health.

Davies, S.M., Tawfik-Shukor, A., De Jonge, B. (2010). Structure, governance, and organizational dynamics of university medical centres in the Netherlands. *Academic Medicine*, 85, 1091–1097.

Dill, D. D. (2006 September). Convergence and diversity: The role and influence of university rankings. In: *Keynote Address presented at the Consortium of Higher Education Researchers (CHER) 19th Annual Research Conference* (Vol. 9).

Dill, D. D., & Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher Education*, 49(4), 495–533.

Hardeman, S. (2013). Organization level research in scientometrics: a plea for an explicit pragmatic approach. *Scientometrics*, 94(3), 1175–1194.

Lewkonia, R. (2002). The functional relationships of medical schools and health services. *Medical Education*, 36, 289–295.

Van Raan, A. F. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1), 133–143.

Rochford, F. (2006). Is there any clear idea of a university? *Journal of Higher Education Policy and Management*, 28, 147–158.

Peart, W.S. (1970). A University Hospital, *Proceedings of the Royal Society for Medicine*, 63.

Siewert, J. R., & Merk, S. (2009). The relationship between German university hospitals and medical schools. Cooperation model versus integration model. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, 52(8), 807.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E., Tijssen, R. J., van Eck, N. J., ... & Wouters, P. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, 63(12), 2419–2432.

WHO (2006), Jan Roth Johnson, *Health Systems in Transition*: Norway, 8.1.

WHO (2010), W. Schäfer et al., *Health Systems in Transition*: The Netherlands, 12.1.

WHO (2011), Seán Boyle, *Health Systems in Transition*: England, 13.1.

---

i Rochford (2006) emphasizes the importance of a sovereign power by which exercise universities are established by means of a legislative act, in any case in the UK and Australia (p. 149).

ii We are aware of notable complex university-hospital systems in Germany, such as Charité in Berlin, or the University Clinic of the Ruhr University Bochum (UK RUB). However, these are exceptions: in most *Bundesländer*, the law prescribes exclusive relations between one university and one hospital.

Measuring opening of national R&D programmes: what indicators for what purposes?[1]

Emilia Primeri e.primeri@ceris.cnr.it / CERIS CNR, Via dei Taurini, 19 00185 Rome (Italy)

Emanuela Reale e.reale@ceris.cnr.it / CERIS CNR, Via dei Taurini, 19 00185 Rome (Italy)

Benedetto Lepori benedetto.lepori@usi.ch / Università della Svizzera italiana, Faculty of Economics, Centre for Organizational Research (Switzerland)

Philippe Larédo philippe.laredo@enpc.fr / ARMINES-LATTS, Laboratoire Territoires Techniques et Sociétés, University of Paris Est (France)

Maria Nedeva mnedeva@dom01.mbs.ac.uk / Manchester Institute of Innovation Research (MioIR), University of Manchester (United Kingdom)

Duncan Thomas duncan.thomas@mbs.ac.uk / Manchester Institute of Innovation Research (MioIR), University of Manchester (United Kingdom)

Abstract

Opening of national research programmes has gained importance as mean for increasing international collaborations and for improving quality and efficiency of scientific research both at the national, European and international level. The concept of opening refers to the fact that actors not belonging to the national research space can participate in research funding programmes. This complex and multidimensional phenomenon can be operationalized through different measures: the participation of foreign partners in domestic research activities with or without funding, the portability of grants when moving abroad, agreements for international collaborations (with or without complementary funding). This underlines the importance for descriptors and indicators, which could provide evidences of different patterns of opening and contrasting perspectives on policy motivations and goals behind opening decisions. The paper aims at presenting the descriptors and indicators used for exploring opening patterns and logics which characterize main project funding instruments in three countries (CH, FR, IT) on the base of data collected in the within of the JOREP project trying to identify national research policy perspectives behind the decision of opening large national research programmes. Evidence emerging for the three countries surveyed are presented and discussed.

1 For this work the authors acknowledge support from the European Commission through the JOREP study (Contract No. RTD/DirC/C3/2010/SI2.561034). We acknowledge the contribution and the information provided by the European Commission that has kindly agreed to the dissemination of the results of the study. The authors would also like to thank the national experts involved in the JOREP data collection. The views expressed in this article are those of the authors and do not necessarily reflect the opinion or position of the European Commission.

Background

Opening of national research programmes is considered a key dimension in the establishment of a European market of knowledge, together with the creation of joint programmes (Pérez, de Dominicis and Guy, 2010; Optimat, 2005). "Opening" refers to the fact that actors not belonging to the national research space can participate in research funding programmes and it represents a complex and multidimensional phenomenon (Pérez, De Dominicis and Guy, 2010).

The question is why and for serving what purposes do countries adopt different strategies and measures of opening. For instance opening can represent a strategic choice towards the increasing push for internationalization and Europeanization of research. To that respect countries can behave differently, acting with formal opening or by putting in place effective measures of opening. On the other hand, opening of national research programmes could serve national policy goals of improving research capabilities and assets through improving collaborations with foreign researchers and research organizations. Thus opening represents a policy priority especially as it serves filling national scientific gaps or when R&D capabilities are weak. This is the case mostly for investigator driven research where foreign collaborations are considered important. Thirdly opening can be fostered by national scientific community, which aims at maintaining a leading edge position in its field or at improving it, through international collaborations, at the national level or rather at the EU level or in the international arena (EC, 2007).

In so far, a multidimensional approach to the analysis of opening of largest national research programmes, which represent main project funding instruments, could help detecting different ways for achieving national goals and objectives besides channelling national funding abroad (Lepori et al, 2007). So opening is an interesting and specific dimension of project funding scheme which is worth to be investigated more in depth. Moreover, different patterns of opening also highlight the existence of institutional arrangements that orient differently national research policies and funding decisions although facing the same external pressures (as in the case of increasing internationalization push) (Lepori et al, 2007; Nedeva 2013; Svanfeldt, 2009).

Nevertheless, despite the fact that opening represents a more and more diffused and complex phenomenon, available studies are quite few. Main indicators used to assess the opening of national research systems provide often insights about researchers' mobility or levels of spending for research collaborations with foreign organizations rather than providing evidences of different patterns of opening and contrasting perspectives on policy motivations and goals behind opening decisions.

The paper presents descriptors and indicators useful to comparatively position the opening patterns and logics characterizing the main project funding instruments in three countries, trying to link evidences to national research policies, motivations and goals. Thus the focus of this study is on the national research policy perspectives, which can be identified behind the decision of opening large national research programmes in the largest context of project funding instru-

ments that can be detected at the national level rather than on level of funding transferred abroad. The paper does not analyse the following and not even the benefits which might stem from opening up of national research programmes to foreign collaborations (e.g. the scientific outputs and outcomes stemming from the opening up of programmes) this being related to an analysis of impact of foreign collaborations rather than to the analysis of different features of opening and policy drivers behind them. The research questions driving the analysis are: what patterns of opening of national research programmes can be identified? How relevant is opening in the frame of national project funding instruments? What logics emerge behind opening decisions? Do evidence support the discussion that national specificities and needs drive different directions towards opening?

In the definition of opening adopted in this study looking at funding flows abroad, which emerge to be generally quite limited, does not represent the main issue, rather different dimensions of opening are considered, which can also help operationalizing the wide notion of opening: the participation of foreign partners in domestic research activities with or without funding, the portability of grants when moving abroad, agreements for international collaborations (with or without complementary funding).

To that end descriptors and indicators are built to characterize the opening of national programmes, which would cover the different dimensions of opening (and their combination) as the portability of funding, the type of beneficiary sectors the programmes are open to, the subject domains covered, the share of projects with foreign partners.

Methodology

The evidences come from a pilot survey of large national research programmes in three countries, Switzerland, France and Italy, realized in the context of the JOREP (Analysis of investments in joint and open research programmes) project[i].

Programmes were selected on the base of their financial volume and importance in the national funding landscape. As a whole 52 programmes were analysed. The following caveat about the sample used in the analysis have to be considered: for Switzerland and Italy the main instruments for project funding at the national level are covered, for France only programmes of the Agence Nationale de la Recherche (ANR) were included.

The table below summarizes main information of the programmes surveyed.

Table 1. Programmes for the pilot analysis on opening

	France	Italy	Switzerland
Programmes included	Thematic programmes by the Agence Nationale de la Recherche (39 thematic programmes by field).	FAR (Fund for the promotion of Research) FIRB (Basic Research Investment Fund) FIRB (Futuro nella Ricerca) FIT (Fund for Technological Innovation) Programme for the diffusion of scientific culture (Law 6/ 2000) Funding of "Programme agreements". PNRA-National Programme of Research in Antarctica National Space Plan	Swiss National Science Foundation: Research projects. National research programmes. National competence centres in research. Sinergia. Swiss Innovation Agency: cooperation projects with the industry.
Involved agencies	Agence Nationale de la Recherche (ANR)	Ministry of Research (MIUR); Ministry of Economy. Italian Space Agency	Swiss National Science Foundation; Swiss Innovation Agency.
Total budget 2009	392 mio. Euros	137 mio. Euros	426 mio. euros
Share of national project funding	N.A.	N.A. but the largest instruments are covered.	76% (OECD NESTI)

Problems of availability of data in the three countries due to differences in the national funding systems were considered and to that end only descriptors including available information in the three countries were exploited. The analysis also exploited programmes documents, mostly publicly available, and, in some cases, information retrieved through interviews to programmes' officers.

Indicators

Besides a unique and very broad definition of opening, this notion in the work is operationalized taking into account several dimensions captured exploiting different types of descriptors which also allow distinguishing between a formal and an actual level of opening.

A first set of descriptors is used to characterize the programmes included in the analysis and patterns of opening in the countries considered. These include descriptors as the countries the programme is open to, the year in which the programme was opened to foreign participants (year of opening), the subject domains covered and the beneficiary sectors foreign collaborations are open to.

A second set of descriptors is aimed at characterising different dimensions of opening. For instance, the criteria for eligibility of participation of researchers and research organizations, the role of partners in the projects, the modes of participation, the time of opening and the possibility to move funding abroad (following researchers) are considered.

The former refers to the nationality of researchers entitled to participate to domestic research programmes (national or foreign) or the location of research organizations (located nationally or abroad). To that respect we can distinguish between programmes open to collaborations of foreign researchers performing research in their own countries or when hosted by a national institution, this being the case of programmes which support mobility of researchers and aimed at attracting talented scientists from abroad. As for the modes of participation, this refers to the way participation is enacted by foreign researchers (i.e. playing or not a formal role in the project-e.g. coordinator or partner- and with or without receiving funding for activities performed). To that respect the following categories of opening can be distinguished: a) international cooperation which includes foreign partners participation to project activities through regular exchanges of information and results but without a formal role in the project, b) participation with a formal role and specific duties in the project-co-applicant or subcontractor- and formal responsibilities but no research funding (expenses for meetings and travel only), c) participation of foreign partners abroad with formal role also eligible to receive research funding. Then the time of opening refers to the moment of the life cycle when opening occurs (application, funding decisions, later opening through involvement of foreign partners). Finally the possibility to move project funding abroad following researchers' move, the portability of grant, is considered.

A third set of descriptors mostly refers to factors that could be considered as enablers of opening, namely the visibility of information concerning foreign participations, the availability of information in English and the possibility to submit proposals in a foreign language.

Finally, the level of opening of countries surveyed is analyzed through two indicators, namely:

- the share of projects with at least a foreign partner (thus considering only collaborations in which foreign partners have an official role);
- the share of projects with foreign participants receiving funding compared to the whole project funding volume.

The table below summarizes the descriptors and indicators used in the work with indications of measures used.

Table 2. Descriptors and indicators on opening

Descriptor/Indicator	Details
Countries opened	- ERA countries - Worldwide - Opening conditional towards some countries only
Year of opening	- The year programmes open to foreign collaborations
Beneficiaries and research topics	- Beneficiary sectors - Subject domains covered/open
Time of opening	- Application and funding stages - Later opening
Eligibility of partners	- Nationality of researchers - Territoriality of research organizations
Modes of participation	- Participation without formal role in the project - Participation with formal role (foreign partner, foreign coordinator) - Foreign partner not receiving research funding - Foreign partner receiving research funding - Availability of funding for international cooperation (travels, exchanges)
Portability of grant	- Possibility of funding to follow researcher (if moving abroad)
Language barriers and visibility conditions	- Availability of information in foreign languages - Calls text in English - Proposals submission in English
Level of opening of national research programmes	- Share of projects with partners abroad - Share of projects with foreign participants receiving funding as % of the whole project funding volume

The descriptors and indicators indicated above allow the characterization of patterns of opening in the countries considered and also distinguishing between a formal and an effective level of opening. This supports a more differentiated analysis and careful consideration of different facets and patterns of opening, which are required in order to understand this complex and multidimensional phenomenon.

Results

Here the main results for the considered countries can be summarized.

Italy has a limited level of openness of national research programmes, with no distinction according to programme types and scientific priorities addressed. Factors which should enable foreign participations appear to be generally weak (visibility of information and use of foreign language). Most of the programmes are open at least in principle, as in the case of the FIRB programme, but lacking effective measures to support it (e.g. portability of grants, official status of foreign organizations in the programmes). Thus most of the programmes mainly envisage the availability of funding for international collaborations, focusing for instance on the support to mobility of researchers or bilateral collaborations.

France displays a diversified attitude towards opening. Although main research programmes show an effective level of opening (e.g. foreign research organizations are entitled of an official status also as coordinator), and measures are adopted to improve foreign participation (e.g. possibility to submit proposals in English), the effectiveness of opening is burden by the lack of appropriate facilitating means (e.g. the availability of calls in English) and funding measures (e.g. portability of grants and funding abroad). The organization of the ANR by scientific domain allows the agency to develop, although to a limited extent in the frame of national regulations, its own strategy to improve international collaborations consistently with the needs and priorities of the scientific community.Thus opening seems more a strategic way for answering to pushes for increasing international collaborations rather than for serving national interests. For instance, programmes considered strategic for national research priorities (e.g. in the field of energy) mainly remain national, and collaborations with countries considered to be relevant partners are managed through different instruments for international collaborations (e.g. bilateral agreements).

Differently, in the Swiss case internationalization policies are mostly reflected in the opening features of national research programmes, with few exceptions when domestic industrial interests are concerned. Switzerland, in fact, shows an effective level of opening so that, except for programmes which account industries among the main beneficiary sectors and support applied research, all programmes are open at least in principle and adequate instruments to support foreign collaborations are envisaged (e.g. official status of foreign organizations, high visibility of programmes information through the widespread use of English, different grant portability schemes). Nevertheless opening seems mostly to be enacted when cross borders cooperation (e.g. in the case of Lead Agency agreements) and reciprocity in collaboration agreements are almost the case. Moreover higher degree of opening can be detected for investigator driven research programmes, which could benefit the most from wider collaborations, instead of applied research programmes which are mostly context and national related and mainly serve national economic interests.

The indicators and descriptors tested in the pilot proved a substantial value to position the opening of national programmes and enabled to point out different dimensions and levels of opening. The pilot highlights as the countries behave differently when promoting international collaborations, supporting the idea that opening up of national R&D programmes represents highly diversified choices within the frame of project funding instruments.

Some general considerations can be drawn. Firstly, opening features are quite heterogeneous and display clear country patterns (e.g. orientation towards basic research and the public sector in Switzerland, towards industrial research in Italy or the clear organisation by research topics in France). Moreover, where national funding systems are more flexible and different decisions centres exist (Government, funding agencies) highly diversified strategies of opening can emerge (e.g. Switzerland, France).

Secondly, different levels of opening seem to be related to the types of programmes considered, science-oriented programmes being more open than programmes driven by economic interests

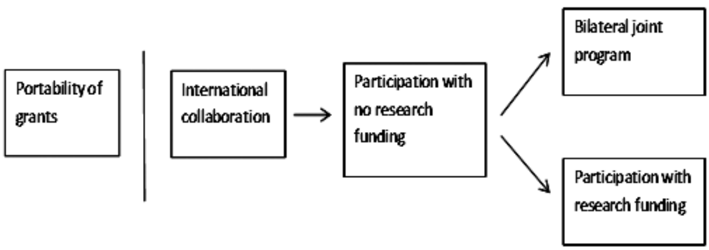
whose benefits are mostly national. For the former visibility and availability of information tend to be good and well detailed, whereas for the latter instruments to facilitate foreign researchers involvement are often weak.

Finally, data show as generally for all countries the possibility to fund research abroad is very limited, the only case being Switzerland which also shows, by the way, a tendency towards conditional opening (i.e. countries collaborating on the base of the Lead Agency Agreement as Germany, Austria, and Luxembourg). A different issue is represented by the possibility of grants to follow researchers and principal investigators, which does not seem to be related to the openness of programmes rather to national funding systems characteristics and programmes' goals. As a consequence, when there is a need of providing research funding to partners abroad, bilateral joint programmes seem to remain the preferred option.

Summing up opening emerges as a widespread phenomenon in the frame of national project funding, highly diversified and consistent with national funding system characteristics.

Instead of letting emerge isomorphic pushes, opening decisions assumes a variety of patterns, which mirror different set of policy motivations and goals behind them (Figure 1).

Figure 1. Dimensions and levels of opening of national programmes



Moreover the levels of opening emerge to be highly selective: international collaboration is encouraged and supported in many national programmes whereas research funding to partners abroad is possible only under specific circumstances which are beneficial to the national research system (like acquiring specific competences) or pursue foreign policy goals (like development aid or cooperation with other countries). This implies also that the amount of funding flows abroad does not represent a reliable measure of opening.

Conclusions

The focus of the study was on the national research policy perspectives behind the decision of opening large national research programmes to foreign collaborations in the largest context of project funding instruments. Although the sample analyzed still needs to be improved including a larger set of countries and programmes to draw general conclusions, some interesting insights emerge.

Compared to a restrictive definition of open programmes, focused on funding flows abroad, which would lead to the conclusion that the phenomenon is relatively marginal and limited to specific cases (e.g. bilateral agreements or opening agreement towards specific countries based on reciprocity criteria) a broader definition of opening, which include different dimensions, displays much more interesting results and supports a discussion on policy perspectives of opening as a specific feature in the frame of national research funding instruments.

The results show that the opening of national programmes is more widespread than expected and that some levels of opening increasingly characterize the large national research programmes, which constitute the bulk of national research funding. There is also anedoctal evidence that opening is in most cases recent and rapidly developing.

Concerning the types of programmes, as expected, science-oriented programmes tend to be more open than programmes driven by national economic needs and interests. Thus a common feature in the observed trend of opening of national R&D programmes is the "additional" value of non-resident researchers and research organizations involvement.

Moreover, the study shows as national states generally are not willing to transfer funding abroad and opening of national research programmes mainly takes place modifying or softening national rules in order to allow foreign partners involvement, although with varying degrees of engagement and different roles, with no or conditional funding commitment by national authorities.

Thus it can be argued that the phenomenon of opening represents a highly relevant evolution in the making of the European Research Area and that the internationalisation of research is a driving force of opening, fostering researchers' cross-border cooperation.

Also, opening allows improving knowledge on strategic issues which cannot be fully addressed by relying on national resources only, especially when challenging issues are to be addressed, with no or conditional financial commitment by national authorities (differently from joint programmes).

In so far opening can represent a strategic choice for internationalise national research preferable to undertaking joint activities especially when national supremacy in strategic domains is at stake or when increasing budget cuts and constraints might limit national involvement in internationally funded research initiatives and programmes.

References

European Commission (2012). *International Cooperation in Science, Technology and Innovation: Strategies for a Changing World*, Report of the Expert Group established to support the further development of an EU international STI cooperation strategy, eds. Sylvia Schwaag Serger and Svend Remoe.

European Commission (2011). *Innovation Union Competitiveness Report*, Brussels: European Commission.

European Commission (2008). *Optimising research programmes and priorities*, Report of Amanatidou. Luxembourg, 2008. EUR 23324.

European Commission (2007). The European Research Area: new perspectives. Green paper, COM (2007), 161, Brussels See http://ec.europa.eu/research/era/pdf/era- greenpaper_en.pdf.

Lepori B.& Durkel, A. (2011). *Evaluation of the Impact of Swiss Bilateral Research Programmes*, Lugano.

Lepori, B., Dinges, M., Reale,E., Slipersaeter,S., Theves,J. and Van den Besselaar,P. (2007a). Comparing the Evolution of National Research Policies: What Patterns of Change? *Science and Public Policy*, 34, 6, 372–88.

Lepori, B., van den Besselaar,P., Dinges,M., van der Meulen,B., Potì,B., Reale,E., Slipersaeter,S. and Theves,J. (2007b). Indicators for Comparative Analysis of Public Project Funding: Concepts, Implementation and Evaluation, *Research Evaluation*, 16, 4, 243–255.

Nedeva, M. (2013). Between the global and the national: Organising European science. *Research Policy*, 42, 220–230.

Optimat Ltd and VDI/VDE/IT (2005). *Examining the Design of National Research Programmes*, Study for EC DG Research, December 2005.

Pérez, E. S., de Dominicis, L. and Guy, K. (2010). *Developing the European Research Area: Opening-up of National R&D Programmes and Joint R&D Policy Initiatives*, JRC Scientific and Technical Research series, Luxembourg: Publications Office of the European Union.

Svanfeldt, C. (2009). A European Research Area built by Member States?, in Henri Delanghe, Ugur Muldur, Luc Soete (eds), *European science and technology policy: towards integration or fragmentation*?: Edward Elgar.

i JOREP is a contract funded by the European Commission [Contract. No. RTD/DirC/C3/2010/SI2.561034) under the Seventh Framework Programme. The project covers eleven countries which are (in alphabetical order): Czech Republic, Denmark, France, Germany, Italy, Netherlands, Norway, Poland, Switzerland, Spain and the United Kingdom.

Journal Name As Proxy for research quality – how the Cult of Journal Impact Factors is Affecting Research

Bernd Pulverer bernd.pulverer@embo.org / Scientific Publications, EMBO, Meyerhofstrasse 1, Heidelberg, 69117 (Germany)

Michele Garfinkel michele.garfinkel@embo.org / Science Policy Programme, EMBO, Meyerhofstrasse 1, Heidelberg, 69117 (Germany)

The peer reviewed research paper remains the main conduit for the exchange of research discoveries. With the diversification and growth of global research output, publishing in selected journals and citation metrics are increasingly employed as surrogates for quality in research assessment. Consequently, the pressure to publish in just a handful of journals has increased dramatically, putting the peer review system and the whole research enterprise under strain. At least as a point for discussion, there could be a relationship between the desire of some to publish in an extremely limited number of journals and the rising incidences both of research misconduct and, in an attempt to be first to a journal and with a lack of training about the consequences, of sloppiness or incompetence in data presentation. All of this degrades the scientific literature as a whole, disrupts the progress of science, and damages relationships between researchers, funders, and the public.

The Journal Impact Factor (IF) metric was originally developed to aid librarians in title selection. Despite the presence of a number of related metrics, in the last two decades the implementation of the Journal IF has gained a dominant role in research assessment at every level, from postdoctoral fellowships and tenure appointments, to research grants and institutional evaluations. The 2-year Journal IF thus is being used in ways that it was not designed for, and that are not serving science optimally.

Our thesis is that journal hierarchies naturally evolve and that they are meaningful, and that they can serve as useful navigational aid to validated, important research. However, these hierarchies, accentuated by the Journal IF metric, must not replace the need to assess the actual scientific output of an individual at face value.

Recently, policy concerns and practical frustrations about the misuse of the Journal Impact Factor were manifested in a document now called the San Francisco Declaration on Research Assessment (DORA, 2013). The principle drafters of the document were from the publications and research administration communities; signatories now include organizations and individuals from a wide range of backgrounds and specialties. Although concerns about the uses of the JIF and other individual impact factors have been expressed for many years, this is perhaps the first example of a stance on their use made publicly across a wide range of disciplines and involving

representatives of all the stakeholders – including journals that themselves benefit from the use of the JIF (see, for example, Alberts, 2013).

Integrating observations and ideas from the scientific publishing and science policy communities and from reactions to the San Francisco Declaration, we will discuss the inherent limitations of this metric and damaging applications of the Journal IF to research assessment in the life sciences. We note that both the scientific and science publishing enterprises become distorted by, for example, institutional requests or requirements that researchers publish only in journals with a minimum but arbitrary Journal Impact Factors.

We will consider the manifestations of gaming of the Journal IF by researchers, journals and institutions, and of a culture of excessive rewards based on this single metric, in particular in emerging research cultures in Europe and Asia. At the same time, in some institutions the use of the JIF is appreciated by those who are concerned about rewards made to individuals based on their status in a department or personal relationships; in this type of case the use of the Journal Impact Factor may be demonstrably better than nothing. In both cases, a related concern then is how metrics-issuing companies may sway JIFs, influencing them in large and small ways by deciding what types of articles count or do not count toward Journal IFs. The need for transparency in these calculations is readily apparent but there seems to be little movement by metrics-issuing groups to discuss how to address this need.

We will reflect on the implications of research assessment that does rely on a single citation metric. A straightforward way to get away from a single metric is to use several metrics. However, the tendency will be to turn such a collection of metrics into a single number, recapitulating the current problems in this assessment system. We will describe the outlines of a system that can holistically take into account a range of measures, and will discuss alternative metrics that focus at the article level as well as improvements to the existing IF metric that would add meaning to this number. We will argue, however, that no single current metric can be used in isolation to assess research performance in the complex and vast arena of biomedical research. We will also argue that research assessment needs to be resourced to allow scientifically founded evaluation of research publications as well as research output other than research papers.

References

Alberts B. 2013. Editorial: Impact Factor Distortions, Science 340: 787.

DORA: The San Francisco Declaration on Research Assessment. (2013) http://am.ascb.org/dora/

Collaboration Benefits and International Visibility of Two-Country-Collaborations[1]

Mathias Riechert riechert@forschungsinfo.de / iFQ – Institute for Research Information and Quality Assurance, Schützenstr. 6a, D-10177 Berlin (Germany)

Daniel Sirtes sirtes@forschungsinfo.de / iFQ – Institute for Research Information and Quality Assurance, Schützenstr. 6a, D-10177 Berlin (Germany)

Haiko Lietz haiko.lietz@gesis.org / GESIS – Leibniz Institute for the Social Sciences, Sachsenhausen 6–8, D-50667 Köln (Germany)

Abstract

We introduce two novel ways to capture the impact benefits two countries receive from collaborations. For both indicators we compare the value of a specific collaboration with the value of average collaborations for each of the countries. As we restrict our analysis to only two-country collaborations and calculate the values for each scientific field individually, many of the problems introduced by former attempts of collaboration indicators dissipate. Additional to a mean based indicator (*Citation Benefit*) we introduce an *International Citation Share* indicator that measures the share of international citations on an item basis. By aggregating and correlating these indicators we show that two different factors of collaboration return, highly cited publications and a general domestic/international bias, *i.e.* the tendency of a publication to be cited more in the originating country, can be measured exclusively by these two indicators. These approaches open up the field to a new kind of deep analysis of scientific collaborations.

Introduction

Many studies imply that collaboration increases the amount of received citations (Chinchilla-Rodríguez, Vargas-Quesada, Hassan-Montero, González-Molina, & Moya-Anegóna, 2010; Glänzel, 2001, 2002; Hsu & Huang, 2011; Katz & Hicks, 1997; Lewinson & Cunningham, 1991; Narin, Stevens, & Whitlow, 1991; Persson, Glänzel, & Danell, 2004). In order to analyze the collaboration link's successfulness, the Relative Citation Eminence (RCE) has been introduced (Glänzel & Schubert, 2001). RCE is computed by dividing the mean observed citation rate of publications co-authored by countries X and Y by the geometric mean of the mean observed citation rates of X and Y. However, the RCE is a symmetric value and does not show which country benefits more from the collaboration or not at all. It was shown that countries differ in their scientific collaboration work depending on their size, citation culture, productivity and geographical proximity (Ding, Foo, & Chowdhury, 1998; Glänzel, Schubert, & Czerwon, 1999; Glänzel, 2001; Katz &

1 Thanks to Jörg Neufeld (iFQ) for useful discussions and Almuth Lietz (iFQ) for help with the figures.

Hicks, 1997; Luukkonen, Tijssen, Persson, & Sivertsen, 1993; Pao, 1981; Singh, 2005; Zhao & Guan, 2011). An asymmetric citation behavior is therefore to be expected and focus of the present paper. To measure aspects of this inequality, several indicators have been proposed recently. The *Citation Rate Increment from the Collaborator* (CRIC) and the *Domestic Citation Rate Comparison when Collaborating* (DCRCC) measure if a collaboration between countries O and P yields more citations from O to P and O to O, respectively. The *Domestic Impact Rate Increment when Collaborating* (DIRIC) measures the increment in average domestic citations relative to the non-collaboration case (Lancho-Barrantes, Guerrero-Bote, & De Moya-Anegón, 2012). The *International Collaboration Gain in Normalized Citation* (ICGNC) is the difference between the field normalized citation rate of collaborative and non-collaborative publications (Guerrero Bote, Olmeda-Gómez, & De Moya-Anegón, 2013).

We contend that there are several methodological problems with these indicators. CRIC, DCRCC, and DIRIC lack field normalization which skews the results toward the highly cited disciplines. Third collaborating countries are not controlled for. The focus, in the case of CRIC and DCRCC, on raw citation scores is very sensitive to extreme values. OCGNC is field normalized but its construction as a difference can be challenged. Finally, all indicators compare collaboration with non-collaboration. This may introduce a bias between international collaborating authors vs. non-collaborating ones.

We introduce methods where specific collaboration is compared with general collaboration. In the indicators to be described, the benefit for one country to collaborate with another is compared to the average[i] of its collaborations. We also avoid some of the methodological problems of the indicators described above. All our indicators are field normalized. Our set of publications includes only those with two collaborating countries. Additionally, in order to measure international visibility we introduce a citation share indicator which is not sensitive to extreme values. The comparison between the values of our indicators helps clarify the latent factors determining high collaboration returns and high international visibility.

Data and Methods

The publications for the analysis are drawn from the Competence Centre for Bibliometrics for the German Science System's bibliometric database based on Web of Science (WoS), published between 2005 and 2009 in journal articles or reviews in all fields. The focus of this study are publications which were collaboratively authored by exactly two countries, as multilateral collaboration requires a differentiated approach (Glänzel & De Lange 1997). 843,666 distinct publications from WoS fulfilled the above criteria. This selection does not exclude the authors with more than one affiliation as was shown problematic by Katz and Martin (1997). However, for our study this only becomes virulent if one author is affiliated with two countries and no other author has a different country affiliation. Our analysis shows that only 0.7 percent of authors have this kind of double country-affiliation, which makes this problem rather negligible.

A sliding citation window of three years was used. All citations were counted without self-citations to prevent self-citing effects from skewing the results (especially on international visibility). To reduce the effect of extreme values, only papers were selected that were in a set of at least 20 papers in a country-field and a country-country combination. The Publications below the threshold of 20 were discarded for both countries, but only for the respective field (as it can be present in different fields simultaneously). From the initial sample, 807,535 (95.7%) publications from 116 distinct countries and 222 distinct fields (using the subject classification scheme of WoS) were consequently used as the basis for the calculations.

The following five indicators are proposed:

Starting point for the *Collaboration Benefit (CB)* indicator construction are the citations per paper, which is associated with a field *f* and was penned by collaborators from countries *o* and *p*. This is the Mean Observed Citation Rate $MOCR_{opf}$. The $MOCR_{opf}$ is divided by the Mean Observed Citation Rate of country *o*'s co-authored papers in that field f. This allows for comparisons between different fields. The indicator for collaboration citation benefit of a country with another country in a field is therefore proposed as follows:

$$CB_{opf} = \frac{MOCR_{opf}}{MOCR_{of}^{co}}$$

$MOCR_{opf}$	Mean Observed Citation Rate of country *o*'s papers co-authored with country *p* in field *f*.
$MOCR_{of}^{co}$	Mean Observed Citation Rate of country *o*'s co-authored papers in field *f*.

To illustrate the construction of the indicator, the collaboration between the USA (*o*) and France (*p*) in the field of thermodynamics (*f*) is used: The USA and France have collaborated in 57 papers and received 301 citations (without self-citations) for them. Consequently, the $MOCR_{opf}$ is 5.28 for both countries in this field. This value is to be compared with the average citations the respective country receives, while collaborating with any one other country, in that field: While the USA has published 1,185 collaborative papers receiving 7,359 citations in the field thermodynamics ($MOCR_{USA}^{co}$ = 5.03), France published 479 such papers receiving 2,407 citations ($MOCR_{FRA}^{co}$ = 6.24). Therefore the Collaboration Benefit for the USA in collaboration with France in the field Thermodynamics is $CB_{USA,FRA,f}$ = 5.28/5.03 = 0.94. while the corresponding Collaboration Benefit is $CB_{FRA,USA,f}$ = 5.28/6.24 = 0.89. Both countries have fewer citations per paper in collaboration compared to the citations both countries generally receive in collaborated papers in the field of thermodynamics.

International Collaboration Benefit (CB_INT)

The second indicator we propose is similarly designed, but targets the international citations of the collaborated:

$$CB\_INT_{opf} = \frac{MOCR\_INT_{opf}}{MOCR\_INT_{of}^{co}}$$

$MOCR\_INT_{opf}$ International Mean Observed Citation Rate of object *o*'s papers co-authored with object *p* in field *f*.

$MOCR\_INT_{of}^{co}$ International Mean Observed Citation Rate of object *o*'s co-authored papers in field *f*.

For illustration the same example is used: The Collaboration Benefit for the USA in collaboration with France in the field Thermodynamics is $CB\_INT_{USA,FRA,f} = 3.63/4.48 = 0.85$ while the corresponding International Collaboration Benefit is $CB\_INT_{FRA,USA,f} = 4.70/4.38 = 1.05$. The USA has fewer international citations per paper in collaboration compared to the international citations normally received from collaborated papers in the field of thermodynamics. In contrast, France has more international citations compared to all international citations from collaborations. In terms of international citations France benefits more from the collaborations.

Domestic Collaboration Benefit (CB_DOM): The third proposed indicator is designed analogically to the International Collaboration Benefit to measure the Domestic Citation Benefit CB_DOM_{opf}.

$$CB\_DOM_{opf} = \frac{MOCR\_DOM_{opf}}{MOCR\_DOM_{of}^{co}}$$

$MOCR\_DOM_{opf}$ Domestic Mean Observed Citation Rate of country *o*'s papers co-authored with country *p* in field *f*

$MOCR\_DOM_{of}^{co}$ Domestic Mean Observed Citation Rate of country *o*'s co-authored papers in field *f*

International Citation Share (CS_INT):The fourth indicator we propose measures the Mean International Citation Share (MICS) of two collaborating countries in a field *f*. The main reason to use the share of international and domestic citations received in collaboration is that this is not dependent on highly cited papers as the share of international vs. domestic citations is weighed for each publication equally, disregarding how many citations an individual paper has yielded. In contrast to the *CB*, the *CS* therefore measures the typical international visibility of the collaboration. The *Item's International Citation Share* is computed by dividing the international citations per paper by all citations. The *Mean International Citation Share* is the average of the IICS in the field. The *International Citation Share* is consequently computed by dividing the *Mean International Citation Share* of the countries co-authored papers in the field *f* by the *Mean International Citation Share* of all publications of the country *o* in field *f*:

Item's International Citation Share (per paper *i*):

$$IICS_i = \frac{non - domestic\ citations_i}{all\ citations_i}$$

International Citation Share:

$$CS\_INT_{opf} = \frac{MICS_{opf}}{MICS_{of}^{co}}$$

$MICS_{opf}$	Mean International Citation Share of country *o*'s papers co-authored with country *p* in field *f*
$MICS_{of}^{co}$	Mean International Citation Share of country *o*'s co-authored papers in field *f*

Example: The USA and France did collaborate in 51 papers. This number differs from the one above as only cited papers are counted for citations shares. In the field Thermodynamics these papers received 301 total citations. For each paper the IICS is computed and subsequently averaged over all papers resulting in a $MICS_{USA,FRA,f} = 0.68$ and a $MICS_{FRA,USA,f} = 0.90$. The average value for the USA in the field is $MICS_{USA,f}^{co} = 0.68$ and for France $MICS_{FRA,f}^{co} = 0.87$. The resulting *CS_INT* values are $CS\_INT_{FRA,USA,f} = 1.00$ and $CS\_INT_{FRA,USA,f} = 1.03$.

Domestic Citation Share (CS_DOM)

The Domestic Citation Share indicator is constructed analogously to *CS_INT* only with the Item's *Domestic* Citation Share as basis.

Aggregation: The proposed indicators are calculated for each collaboration relationship in each field. For each collaboration pair, the number of publications *n* is used for a weighted average, *i.e.* a field-normalized aggregation. Therefore, each indicator can subsequently be aggregated to the field-independent country-country level. Finally, again a weighted aggregation on the collaborating country p is possible in order to get indicators on the country level. It is possible to aggregate these indicators to the country-field and field level as well. As the intention of our study is to analyze the collaboration for countries and not for fields we focus on this aggregation path in the following.

Results

On the country-country-field level, 181,110 different combinations with the five described indicators were computed in 116 countries and 222 fields. Table 1 shows the descriptive statistics for the indicators on the country-country-field level. It is evident that, as to be expected, the CS indicators have less dispersion than the CB indicators.

Table 1. Indicator Descriptive Statistics

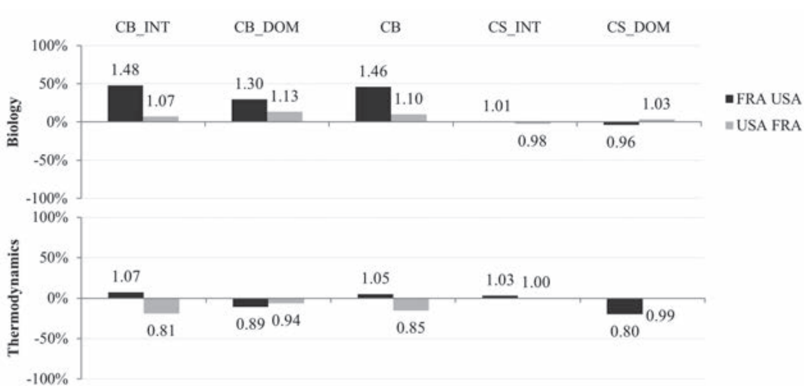
	N	*CB_INT*	*CB_DOM*	*CB*	*CS_INT*	*CS_DOM*
Min:	1	0	0	0	0	0
1st quartile:	1	0.3407	0	0.3503	0.975	0
Median:	3	0.707	0.4059	0.7112	1.028	0.571
Mean:	14.39	0.9006	0.8868	0.8975	1.005	0.98
3rd quartile:	10	1.1409	1.1288	1.1381	1.075	1.218
Max:	2346	106.8925	78.1844	103.3742	2.773	48

Table 2. Indicator Values in Different Fields:

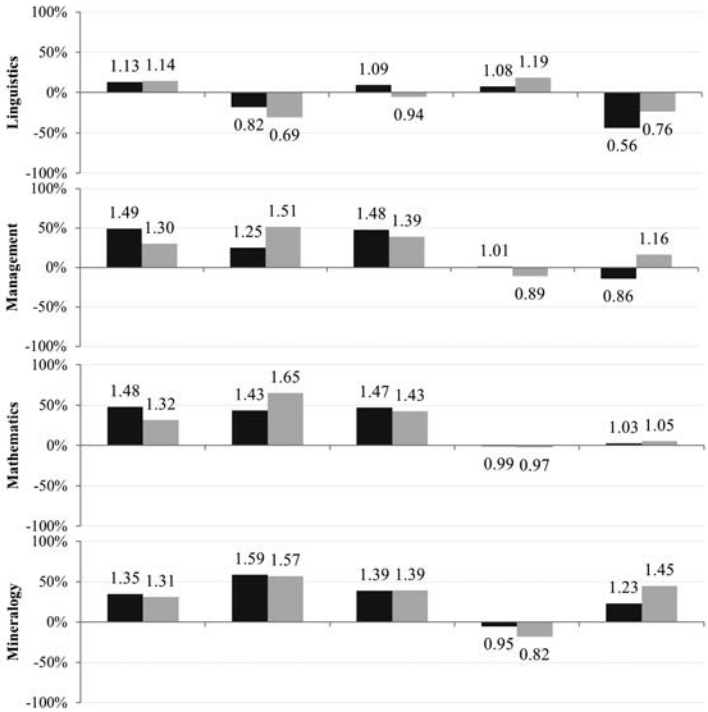
Country O	Country P	Field	*CB_INT*	*CB_DOM*	*CB*	*CS_INT*	*CS_DOM*
FRA	USA	Biology	1.48	1.30	1.46	1.01	0.96
USA	FRA	Biology	1.07	1.13	1.10	0.98	1.03
FRA	USA	Thermodynamics	1.07	0.89	1.05	1.03	0.80
USA	FRA	Thermodynamics	0.81	0.94	0.85	1.00	0.99
FRA	USA	Linguistics	1.13	0.82	1.09	1.08	0.56
USA	FRA	Linguistics	1.14	0.69	0.94	1.19	0.76
FRA	USA	Management	1.49	1.25	1.48	1.01	0.86
USA	FRA	Management	1.30	1.51	1.39	0.89	1.16
FRA	USA	Mathematics	1.48	1.43	1.47	0.99	1.03
USA	FRA	Mathematics	1.32	1.65	1.43	0.97	1.05
FRA	USA	Mineralogy	1.35	1.59	1.39	0.95	1.23
USA	FRA	Mineralogy	1.31	1.57	1.39	0.82	1.45

The results shown in Table 2 are on the already introduced example of the collaboration between the USA and France. Both countries collaborated in 205 disciplines. For reasons of space we will show only a limited subset of fields. The indicator values are depicted as a graph in Figure 1.

Figure 1. Indicator Values for the USA and France



… Continuation Figure 1



CB: France benefits in all six fields compared to its overall collaborations when collaborating with the USA, albeit the benefit varies considerably between the fields (1.48 in management vs. 1.05 in thermodynamics). The USA benefits less from the collaboration with France and has lower citation rates in the fields of thermodynamics and linguistics compared to their usual citations in the fields. The benefit in terms of international citation (*CB_INT*) is close to the *CB* in the fields analyzed, although there are some variations for the USA. The domestic citations are observably different for the fields. The collaboration in the field mathematics is, for example, characterized by high benefits for both countries 1.43 and 1.65 as well as in mineralogy while collaboration in linguistics and thermodynamics seems to attract fewer domestic citations.

CS: The results for the citation share indicator differ substantially from the CB results. In coherence with Table 1, the *CS_INT* and *CS_DOM* have less variance compared to the CB indicators. The *CS_INT* is only above 1 in the field of linguistics for the USA; therefore only in this field the USA receives a higher International Citation Share compared to their overall collaborations. Mineralogy is an example for how both indicator types can differ in the same field. Although both countries benefit from the collaboration in terms of international, domestic and overall citations, the *International Citation Share* is clearly lower compared to their overall collaborations. This conspicuous difference between the indicators is the subject of a further correlation analysis, which is conducted in order to obtain an interpretation of the indicators in the following.

Table 3 shows the Pearson correlation coefficient for the proposed indicators computed for all 181,110 combinations. The results show that none of the indicators is correlating with the number of publications (*N*). The highest correlation (0.991) is found between the *CB_INT* and *CB* indicator. There are considerably high correlations between *CS_INT* and *CS_DOM* (– 0.722) and between *CB_DOM* and *CS_DOM* (0.615). Additionally there are minor correlations between *CB_INT* and *CB_DOM* (0.331), *CB_DOM* and *CB* (0.414) and *CB_DOM* and *CS_INT* (– 0.417). All other indicator combination correlations are below 0.1.

Table 3. Indicator Correlation

	CB_INT	*CB_DOM*	*CB*	*CS_INT*	*CS_DOM*
N	0.026	0.018	0.027	-0.008	0.003
CB_INT		0.331	0.991	0.106	-0.071
CB_DOM			0.414	-0.417	0.615
CB				0.028	-0.016
CS					-0.722

Discussion

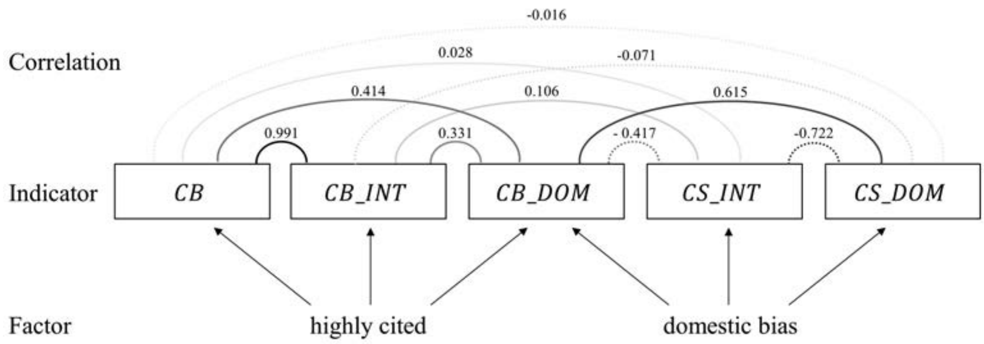
The results of the analyzed collaboration between the USA and France have shown a multifaceted collaboration. The high variation between the values in different fields reinforces the argument that collaboration on the basis of citations should be measured field-specific to prevent skewed results. An aggregation of citation-based collaboration indicators should therefore be based on the weighted field values.

While both countries observably do benefit in the fields of mineralogy, mathematics and management, France is benefiting substantially more in biology in terms of citations. The collaboration in the fields of linguistics and thermodynamics feature fewer citation benefits for both countries. In terms of domestic citation benefits, both countries have even fewer citations than their usual collaboration papers receive. An important difference is visible when comparing the citation benefits to the citation shares. As has been shown exemplarily, the share of citation per item differs substantially from the citations per paper. The collaborations of USA and France in the fields of mathematics and management show how the citations per paper can increase while the citation share is stable or even declining. We infer that the indicators measure different factors of collaboration, and have therefore to be analyzed in more detail.

Figure 2 depicts the correlations of the indicators. From the correlations between the citation benefit and the citation share indicators there are following points that can be learned. *CB* and *CB_INT* correlate very strongly. The non-domestic publication share, *i.e.* the share of all publications where the country at question is not involved is almost always the majority and in over 96%

of country-field combinations it amounts to over 90%. Therefore it is not surprising that both indicators will coincide.

Figure 2. Indicator Factor Dependencies



The citation benefit indicators are most influenced by very successful publications, as the mean is calculated over the set of papers in the combination. Therefore, a few extremely highly cited publications will skew the *CB* indicator to a very positive value. In contrast, the citation share indicator is not affected by these 'lucky few.' The share of international vs. domestic citations is weighed for each publication equally, disregarding how many citations an individual paper has yielded. Therefore the citation share indicator gives us a better picture of the *typical* distribution of citations rather than just reiterating that there are a couple of highly cited publications in the mix. The *International Citation Share* indicator being very lowly correlated with the *International Citation Benefit* indicator while quite highly negatively correlated with the domestic citation benefit indicator is a strong indication that the two international indicators truly measure two different latent dimensions.

The prior probability to be cited domestically is very low and therefore the *International Citation Benefit* is hardly affected by domestic citations. On the other hand the *Domestic Citation Benefit* is influenced by two factors: very highly cited papers will also have quite a few domestic ones, and for lowly cited papers it will show whether these few citations are over-proportionally domestic, this is what we call a *domestic bias*. The international citation share does only measure the second aspect as it does not measure highly cited papers more highly than lowly cited and therefore it really picks up on the domestic vs. international orientation of publications. This second, overlapping factor of Domestic Citation Benefit and International Citation Share is how the relatively high negative correlation between the two can be explained.

In conclusion, we suggest using only two of these indicators *CB* and *CS_INT* (or *CS_DOM*) in order to capture the two main factors that influence collaboration impact indicators, highly cited

publications and the general domestic bias. In this short paper we have only dipped into the vast possibilities of analyzing these indicators on different aggregation levels and as the basis of an in-depth *directed* network analysis (The manuscript of this analysis is in preparation).

References

Albarrán, P., Crespo, J. A., Ortuño, I., & Ruiz-Castillo, J. (2011). The skewness of science in 219 sub-fields and a number of aggregates. *Scientometrics*, 88(2), 385–397.

Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Hassan-Montero, Y., González-Molina, A., & Moya-Anegóna, F. (2010). New approach to the visualization of international scientific collaboration. *Information Visualization*, 9(4), 277–287.

Ding, Y., Foo, S., & Chowdhury, G. (1998). A bibliometric analysis of collaboration in the field of information retrieval. *International Information and Library Review*, 30, 367–376.

Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, 51(1), 69–115.

Glänzel, W. (2002). Coauthorship patterns and trends in the sciences (1980–1998): A bibliometric study with implications for database indexing and search strategies. Katholieke Universiteit Leuven.

Glänzel, W. & De Lange, C. (1997). Modelling and measuring multilateral co-authorship in international scientific collaboration. Part II. A comparative study on the extent and change of international scientific collaboration links. *Scientometrics*, 40(3), 605–626.

Glänzel, W., & Schubert, A. (2001). Double effort = Double impact? A critical view at international co-authorship in chemistry. *Scientometrics*, 50(2), 199–214.

Glänzel, W., Schubert, A., & Czerwon, H. J. (1999). A bibliometric analysis of international scientific cooperation of the European Union (1985–1995). *Scientometrics*, 45(2), 185–202.

Guerrero Bote, V. P., Olmeda-Gómez, C., & De Moya-Anegón, F. (2013). Quantifying the benefits of international scientific collaboration. *Journal of the American Society for Information Science and Technology*, 64(2), 392–404.

Hsu, J., & Huang, D. (2011). Correlation between impact and collaboration. *Scientometrics*, 86(2), 317–324.

Katz, J. S., & Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics*, 40(3), 541–554.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research policy*, 26(1), 1–18.

Katz, J. S. (2012). Scale-Independent Measures: Theory and Practice. *Proceedings of the 17th International Conference on Science and Technology Indicators*,1–19.

Lancho-Barrantes, B. S., Guerrero-Bote, V. P., & De Moya-Anegón, F. (2012). Citation increments between collaborating countries. *Scientometrics*, 1–15.

Lewinson, G., & Cunningham, P. (1991). Bibliometric studies for the evaluation of trans-domestic research. *Scientometrics*, 21(2), 223–244.

Luukkonen, T., Tijssen, R. J. W., Persson, O., & Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics*, 28(1), 15–36.

Narin, F., Stevens, K., & Whitlow, E. S. (1991). Scientific co-operation in Europe and the citation of multinationally authored papers. *Scientometrics*, 21(3), 313–323.

Pao, M. L. (1981). Co-authorship as communication measure. *Library Research*, 2(4), 327–338.

Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60(3), 421–432.

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management science*, 51(5), 756–770.

Zhao, Q., & Guan, J. (2011). International collaboration of three 'giants' with the G7 countries in emerging nano-biopharmaceuticals. *Scientometrics*, 87(1), 159–170.

i Citation averages may happen to be power-law distributed (Albarrán, Crespo, Ortuño, & Ruiz-Castillo, 2011). Averages only give a valid characterization as long as the Central Limit Theorem holds (Katz, 2012). For the calculation and interpretation of the presented indicators one must be aware of a possible bias.

Erratum:

There has been a deplorable mix-up in the numeric examples:

p. 304:

To illustrate the construction of the indicator, the collaboration between the USA (*o*) and France (*p*) in the field of thermodynamics (*f*) is used: The USA and France have collaborated in 57 papers and received 301 citations (without self-citations) for them. Consequently, the $MOCR_{opf}$ is 5.28 for both countries in this field. This value is to be compared with the average citations the respective country receives, while collaborating with any one other country, in that field: While the USA has published 1,185 collaborative papers receiving 7,359 citations in the field thermodynamics ($MOCR_{USA}^{co} = 6.21$), France published 479 such papers receiving 2,407 citations ($MOCR_{FRA}^{co} = 5.03$). Therefore the Collaboration Benefit for the USA in collaboration with France in the field Thermodynamics is $CB_{USA,FRA,f} = 5.28/6.21 = 0.85$ while the corresponding Collaboration Benefit for France is $CB_{FRA,USA,f} = 5.28/5.03 = 1.05$. Thus, for France a collaboration with the USA is more beneficial than average, while the USA have fewer citations per paper in collaboration with France than in all collaborated papers in the field of thermodynamics.

p. 305

For illustration the same example is used: The Collaboration Benefit for the USA in collaboration with France in the field Thermodynamics is $CB\_INT_{USA,FRA,f} = 3.63/4.48 = 0.81$ while the corresponding International Collaboration Benefit is $CB\_INT_{FRA,USA,f} = 4.70/4.38 = 1.07$.

A bottom-up approach to assess the interdisciplinarity of journals from a multidisciplinary corpus of bibliographical records[1]

Ivana Roche ivana.roche@inist.fr / CNRS, Institut de l'Information Scientifique et Technique, UPS 76, 2 allée du Parc de Brabois, Vandoeuvre-lès-Nancy, F-54519 Cedex (France)

Dominique Besagni dominique.besagni@inist.fr / CNRS, Institut de l'Information Scientifique et Technique, UPS 76, 2 allée du Parc de Brabois, Vandoeuvre-lès-Nancy, F-54519 Cedex (France)

Claire François claire.francois@inist.fr / CNRS, Institut de l'Information Scientifique et Technique, UPS 76, 2 allée du Parc de Brabois, Vandoeuvre-lès-Nancy, F-54519 Cedex (France)

Marianne Hörlesberger marianne.hoerlesberger@ait.ac.at / AIT, Austrian Institute of Technology GmbH, Donau-City-Strasse 1, 1220 Vienna (Austria)

Edgar Schiebel edgar.schiebel@ait.ac.at / AIT, Austrian Institute of Technology GmbH, Donau-City-Strasse 1, 1220 Vienna (Austria)

Abstract

This work is investigating the possibility of assessing the interdisciplinarity of scientific and technological journals without using any taxonomy or classification scheme as those usually adopted in bibliographical databases. For that, we start from a large corpus of bibliographic records from which we extract terms, either keywords already present or obtained by text mining techniques. With the help of a clustering method, that corpus is split into clusters defining a given number of scientific fields. Those fields and the keywords indexing each document are the basis of the calculation of an interdisciplinarity score applying the diffusion model approach. Then, we calculate an interdisciplinarity indicator for each journal by combining the scores obtained by its articles.

Introduction

Interdisciplinarity and its corollary, specificity, are useful characteristics to locate metaphorically a journal in the scientific and technological (S&T) literature landscape. Previous works were done on the issue of subject classification and the creation of coherent journal sets. Usually these approaches are based on data available from Thomson Reuters' Journal Citation Reports (JCR) where the citations got by each publication are aggregated at the journal level, i.e. the work by Leydersdorff and Rafols (2011) that makes use of measures of interdisciplinarity like network

1 This work was partially inspired by DBF (Development and Verification of a Bibliometric Model for the Identification of Frontier Research), a Coordination and Support Action of the IDEAS specific programme of the European Research Council (ERC). The authors wish to acknowledge this contribution.

indicators or unevenness indicators. In a recent study Thijs et al. (2013) introduced a very interesting approach building a network among journals based on bibliographic coupling. Hybrid approaches based on citation-based and lexical similarities are also known (Janssens et al., 2008) as well as approaches combining several indicators of journal specificity based on textual coherence and research communities (Boyack and Klavans, 2011).

In our work, an exclusively content-based approach is developed to determine from a large multidisciplinary corpus of bibliographical records an indicator measuring the interdisciplinarity of each record and, from that, the interdisciplinarity of the journals where these documents were published.

Our approach is directly inspired on the methodology developed in the context of the DBF project (Hörlesberger, 2013), the goal of which was to infer attributes of 'frontier research' in peer-reviewed research proposals under the scheme of the European Research Council (ERC). To this end, indicators across scientific disciplines and in accord with the strategic definition of frontier research by the ERC are elaborated, exploiting textual proposal information and other scientometric data of grant applicants. In particular, an indicator was devised to characterized any project that "*… pursues questions irrespective of established disciplinary boundaries, involves multi-, inter- or trans-disciplinary research that brings together researchers from different disciplinary backgrounds, with different theoretical and conceptual approaches, techniques, methodologies and instrumentation, perhaps even different goals and motivations*" (EC, 2005) and that we defined as interdisciplinarity.

That indicator is built upon the basic assumption and previously successfully tested concept (Schiebel et al., 2010) that the frequency of occurrence and distribution of discipline specific keywords in scientific documents can be used to classify and characterize disciplines. The concept is consistent with the practice of bibliometric clustering, where the contents of each cluster (e.g., words and articles, or cited references and articles) are ranked by some index (e.g. TF-IDF) of specificity to the cluster.

In the next section the developed methodology is presented followed by some preliminary results.

Methodology

Our methodology is based on a diffusion model approach (Schiebel et al., 2010; Roche et al., 2010), developed in the context of a previous project aiming at detecting emerging technologies, that evaluates the status of each term in a considered discipline by measuring its so-called degree of diffusion.

The diffusion model is founded on the assumption that new findings in a research field are published in journals, conference proceedings, books etc. That S&T literature is collected in bibliographical databases where the content of each document is represented with a set of key-

words. Keywords that describe the innovative results occur in the first stage in an unusual manner. In the second stage the research intensifies and established keywords are used. In later stages, the results cross the disciplinary barrier by diffusing to other research fields where they follow a similar evolution cycle. Consequently, the diffusion status is obtained by the calculation for each keyword of a diffusion degree that can be either "unusual", "established" or "cross-section".

Two pragmatic approaches are successively employed to realise this categorisation. Firstly, the so-called Home Technology terms (H-T terms) are defined. We assumed keywords which are specific for a field occurred with a higher probability in that field rather than in others. The probability is defined by the frequency of one term in a field divided by the number of articles in this field, namely the relative term frequency (rtf_{Field}). For a term, we calculate its rtf_{Field} in each field and the field with the highest probability is declared to be its Home Technology. So after this assignment we obtain for each field the list of its H-T terms. Therefore the complete terminology associated to a field consists of the union of its H-T term list and the set of terms imported from the other fields.

Secondly, we use the Gini index (or also Gini coefficient), a measure of statistical dispersion developed by the Italian statistician Corrado Gini (Gini, 1921) at the beginning of the 20th century. The Gini index (*GINI*) is a measure of the inequality of a distribution and it varies from *0* to *1*, a value of *0* expressing total equality and a value of *1* maximal inequality. It is commonly used as a measure of inequality of the income or wealth of the countries. It is, in this study, employed as a measure of the dispersion of a term in a scientific domain. A Gini index equal to *0* means a completely uniform distribution and indicates that the term occurs in all the considered fields of the domain. Conversely, a Gini index of *1* tells us that the term is very specifically limited to the only field where it appears.

If we consider a set of *n* Home Technologies, the Gini index of a term can be calculated by the Brown formula:

$$GINI = 1 - \sum_{k=0}^{n-1} (X_{k+1} - X_k)(Y_{k+1} + Y_k)$$

where *X* is the cumulative share of Home Technologies, and *Y* the cumulative share of occurrences of the considered term.

In the present study, we are not concerned with detecting innovative technologies but with characterizing scientific fields by analysing their related terminologies. So, hereafter we will not speak anymore of Home Technology but of Home Field (H-F). Moreover, we do not consider the categorization of the keywords according to their diffusion degree, but only the determination for each keyword of its rtf_{Field} and its Gini index. These values allow for each keyword either to assign it to a H-F or to discard it if it is not discriminant enough.

Determining the set of H-F can be done either with the help of a pre-defined taxonomy, or with a content analysis approach starting with a term extraction, followed by a validation step operated

on the extracted keywords indexing each document and finishing by a clustering splitting the corpus into a given number of H-F.

In our case, we apply a non-hierarchical clustering algorithm, the axial K-means method, coming from the neuronal formalism of Kohonen's self-organizing maps, followed by a principal component analysis in order to represent the obtained clusters on a 2-D map (Lelu and François 1992). This step is realized by employing an in-house software tool, Stanalyst (Polanco et al. 2001), dedicated to the scientific and technical information analysis.

The axial K-means is a variant of the well-known K-means clustering algorithm: it derives half-axes, or "axoïds" maximizing a global inter-axes inertia criterion, instead of deriving cluster centroïds maximizing the inter-class inertia. One can sort the cluster's describers and documents along one of these half-axes as well as project the other terms and documents onto it. These projections on any given axis represent the weight of the describers and the document in the corresponding cluster. In this way, one can derive a fuzzy interpretation of the resulting axes, though the method is a strict clustering technique. This method is fast and can handle very large amounts of data. It is formally related to neural models with unsupervised winner-take-all learning. With that clustering algorithm a document may belong to one or more clusters. We consider that the cluster where the document has the higher weight is its H-F.

At this stage, as each publication is allocated to an H-F, it is possible for each one to calculate its HFT and AFT, respectively, the share of its H-F terms and the share of its "abroad terms", i.e. terms assigned to the other H-F. The higher its AFT value, the more interdisciplinary the publication. However this value does not account for the diversity of origins of these imported terms: do they come from a unique H-F or from several? Indeed, for two publications with an equal value of AFT, the one with the greater number of different origins of abroad terms should receive a higher value.

Finally, for each journal represented in the corpus with a statistically significant number of articles, we combine the AFT values calculated for all its articles.

Results

The data set is extracted from the PASCAL database that is specifically adapted to the purpose of our approach. It provides broad multidisciplinary coverage of scientific publications and contains nowadays about 20 million bibliographic records from the analysis of the scientific and technical international literature published predominantly in journals and conference proceedings. On the other hand, the PASCAL records benefit from an indexing by both keywords and thematic categories of a classification scheme assigned to each individual publication, either manually by scientific experts or automatically based on a content analysis. It is this terminology, formed by the indexing keywords that we can also refer to as "terms", that we employ in our analysis, after an assessment step done by a scientific expert.

The query operated in this work aims to represent the whole spectrum of disciplines in the PASCAL database by following the magnitude of their representation. The obtained corpus, that comprises 105,254 bibliographic records, is a set of randomly chosen weekly updates of the PASCAL database.

Although each PASCAL record has at least one classification code, we did not exploit this information to define our H-F because the fine-grained classification scheme produces a huge number of disciplines, way too large for our purpose. As indicated previously, to determine our set of H-F without the help of that in-house taxonomy, we decided to unfold the "content analysis + indexing validation + clustering" sequence described in the Methodology section.

The examination of the clusters and their content bring to a final validation of the cluster list corresponding to the list of H-F. That is at that time, for instance, that some clusters the content of which is very close could be merged.

At this stage, we have the list of H-F, the list of documents assigned to each H-F and the set of keywords representing the content of each document. So, all the conditions have been met to apply the diffusion model approach to calculate an interdisciplinarity score for each document. Then, the set of values got by all the documents from the same journal are combined to produce a journal interdisciplinarity indicator. For obvious statistical reasons, only the journals with a significant number of articles in the studied corpus are taken into consideration. Finally, we put the obtained results into perspective.

References

Boyack K.W. and Klavans R. (2011). Multiple dimensions of journal specificity: Why journals can't be assigned to disciplines, Proceedings of the 13th International Conference of the International Society for Scientometrics & Informetrics, Durban, 04–07 July, 123–133

EC – European Commission (2005). Frontier research: The European Challenge. High Level Expert Group Report, EUR 21619

Gini C. (1921). Measurement of inequality of incomes, *Economic Journal* 31, 124–126

Hörlesberger M., Roche I., Besagni D., Scherngell T., François C., Cuxac P., Schiebel E., Zitt M. & Holste D. (2013). A concept for inferring 'frontier research' in grant proposals, submitted to Scientometrics

Janssens F., Glänzel W. & De Moor B. (2008). A hybrid mappping of information science, *Scientometrics*, 75, 3, 607–631

Lelu A. & François C. (1992). Hypertext paradigm in the field of information retrieval: A neural approach, 4th ACM Conference on Hypertext, Milano

Leydesdorff L. and Rafols I. (2011). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations, *Journal of Informetrics* 5, 87–100, doi:10.1016/j.joi.2010.09.002

Polanco X., François C., Royauté J., Besagni D. & Roche I. (2001). STANALYST: An integrated environment for clustering and mapping analysis on science and technology, 8th International Conference on Scientometrics and Informetrics, July 16–20, 2001, Sydney, Australia, Proceedings Vol. 2, 871–873.

Roche I., Besagni D., François C., Hörlesberger M. & Schiebel E. (2010). Identification and characterisation of technological topics in the field of Molecular Biology, *Scientometrics*, 82, 663–676.

Schiebel E., Hörlesberger M., Roche I., François C. & Besagni, D. (2010). An advanced diffusion model to identify emergent research issues: the case of optoelectronic devices, *Scientometrics*, doi: 10.1007/s11192-009-0137-4

Thijs B., Zhang L. & Glänzel W. (2013). Bibiographic coupling and hierarchical clustering for the validation and improvement of subject-classification schemes, 14th International Conference on Scientometrics and Informetrics, July 15–19, 2013, Vienna.

Publication productivity expressed by age, gender and academic position – an analysis on Norwegian scientists

Kristoffer Rørstad Kristoffer.rorstad@nifu.no / NIFU – Nordic Institute for Studies in Innovation, Research and Education, Wergelandsveien 7, NO-0167 Oslo (Norway)

Dag W. Aksnes Dag.w.aksnes@nifu.no / NIFU – Nordic Institute for Studies in Innovation, Research and Education, Wergelandsveien 7, NO-0167 Oslo (Norway)

Abstract

This study investigates how scientific performance in terms of publication productivity is influenced by the gender, age and academic position of the researchers. Previous studies have shown that these factors are important variables when analysing scientific productivity at the individual level. The study is based on almost 12,400 Norwegian university researchers, and about 35,800 observations. We have carried out regression analyses of each major academic field. The results show that publication productivity is a function of academic position, gender and age in the natural sciences, engineering and medicine. However, for the social sciences and humanities, the productivity does not fit into the same regression model, due to different publication pattern and larger individual differences.

Introduction

Can scientific publication output be expressed by age, gender and academic position? And if so, to what extent do these variables have any effect on scientific performance in terms of publication productivity? These questions have been raised several times and have been recurring issues in the bibliometric literature. Previous research has shown that all these variables have an effect on scientific output, but the results have not been entirely consistent. The relationship between age and productivity has been found to be curvilinear in several studies. The average production of publications increases with age and reaches a peak at some point during the career and then declines (see for instance Aksnes, Rørstad, Piro and Sivertsen (2011); Kyvik, 1990; Cole, 1979; Barjak, 2006; Gonzalez-Brambila & Veloso, 2007). Studies on gender and productivity have shown that female scientists tends to publish fewer publications than their male colleagues do (e.g. Cole and Zuckerman, 1984; Kyvik and Teigen, 1996; Long, 1992; Xie and Shauman, 1998). Finally, scientific productivity has been found to increase within the hierarchy of academic positions where professors are the most prolific personnel (see e.g. Kyvik, 1991, Aksnes, Rørstad, Piro and Sivertsen, 2011).

In this large-scale study, we will investigate if age, gender and scientific position have the same effect on productivity across fields. In previous studies where we investigated these effects, we looked at the average sum of article equivalents for each scientist in the period from 2005 to 2008. In this study, we will look at the average sum of article equivalents per scientist per year. As a result, we reduce the effect of the time a researcher is employed in an academic position.

Methods and data

For our study on publication output, we applied the Norwegian publication database "Current Research Information System in Norway (Cristin)". This database has a complete coverage of all peer-reviewed scientific and scholarly publication output, including books, articles, reviews and conference series. Our analysis covers a 7-year period from 2005 to 2011 at the four major universities in Norway (Oslo, Bergen, Trondheim and Tromsø) which account for about 75 per cent of the total publications in the higher education sector in Norway.

To provide information of individual characteristics of the researchers; field of science, position, age and gender, the bibliographic database was coupled to NIFU's Research Personnel Register. This database contains individual data on all researchers in the higher education sector and Institute Sector in Norway every second year from 1977 to 2007, and every year from 2008. Every scientists publications were coupled with his or hers age and position at the time of publications. The publications were assigned to five major fields; the humanities, social sciences, natural sciences, engineering and medicine.

Table 1. Number of persons and observations by field of science.

Field of sciences	Number of persons	Number of observations
Humanities	1 340	3 691
Social Sciences	1 979	5 821
Natural sciences	3 151	9 558
Engineering and technology	1 912	4 757
Medicine	4 021	11 974
Total	**12 403**	**35 798**

The coupling of these two databases resulted in a dataset we analysed, containing about 12,400 persons in almost 35,800 observations (i.e. publication per years). Non-publishing personnel have not been included in the analysis. Publication output is measured as article equivalents per person per year. In this calculation, co-authored publications are fractionalised among the authors and monographs are weighted as equal to 5 articles (in journals or books) in order to make the research efforts behind publications comparable. The weighting of books is based on

Kyvik's summary of such weighting procedures from other studies, which shows that most studies equate 4–6 articles to one full monograph (Kyvik, 1991).

For our analysis, we included four main academic positions: professors, associated professors, postdocs and PhD-students. In addition, physicians/medical doctors where included for medicine. To simplify, the researchers were divided into five-year age categories.

Table 2. Numbers of observations by field of science, position and gender

Field of science/Position	Women	Men	Total
Humanities	1 434	2 257	3 691
Professor	512	1 228	1 740
Associate professor	459	577	1 036
Post doc	200	176	376
PhD-student	263	276	539
Natural sciences	2 066	7 492	9 558
Professor	428	3 501	3 929
Associate professor	323	1 067	1 390
Post doc	505	1 244	1 749
PhD-student	810	1 680	2 490
Medicine	4 645	7 329	11 974
Professor	719	2 463	3 182
Associate professor	511	446	957
Post doc	743	640	1 383
PhD-student	1 926	1 171	3 097
Physician/medical doctor	746	2 609	3 355
Social sciences	2 240	3 581	5 821
Professor	764	2 006	2 770
Associate professor	609	812	1 421
Post doc	299	226	525
PhD-student	568	537	1 105
Engineering and technology	849	3 905	4 754
Professor	118	1 378	1 496
Associate professor	64	329	393
Post doc	163	532	695
PhD-student	504	1 666	2 170
Total	**11 234**	**24 564**	**35 798**

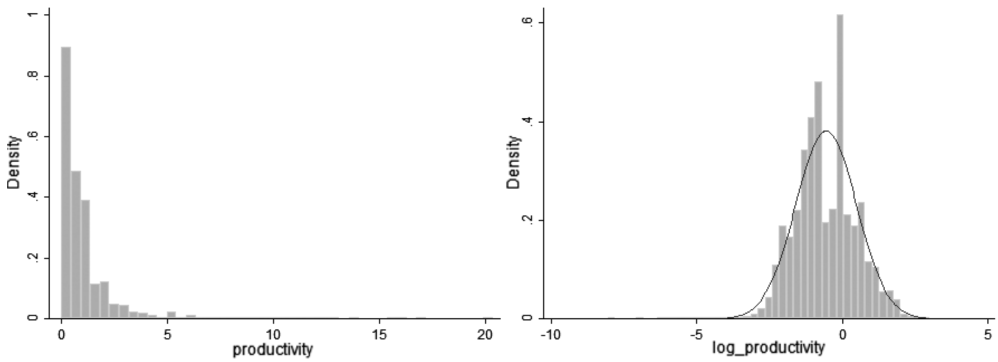
In order to be able to make regression models on our data, we recoded our variables. First, positions were encoded to numerical categorical variables (professors = 1, associated professors = 2, postdocs = 3, PhD-students = 4 and physicians/medical personnel = 6). Secondly, dummy variables were made for each of the categorical position variables. Genders were recoded (men=0 and women=1). Age groups were categorized into 20–24 years=1, 25–29=2, 30–34=3 and so forth.

Our assumption that productivity (average number of article equivalent per person per year is a function of gender, age and scientific position can be written as:

$$Y(\text{productivity}) = b_0 + b_1{*}\text{gender} + b_2{*}\text{age} + b_3{*}\text{position} + \text{residuals} \quad (1)$$

The productivity of publications varies significantly among the individuals, and is not normally distributed. A histogram of our productivity indicator shows a positive skew in Figure 1. A common way to eliminate skewness is to transform the data to naturally logarithmic scale.

Figure 1. Distribution of numbers of average article equivalents per person per year and natural logarithmic numbers. N=35,798



We analysed the productivity by fields of sciences, academic position, age groups and gender. Regression analysis was performed on each major field separately in order to investigate whether one regression model is valid for all fields. Since we previously have found that age and productivity often has a curvilinear relationship, we include the power of age in our regression model. Our new regression model is then:

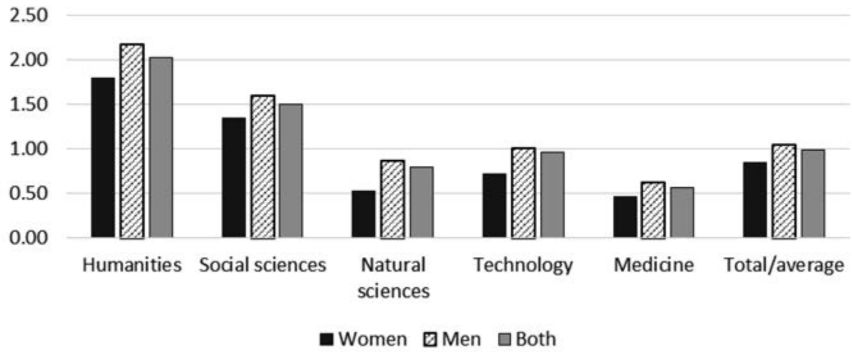
$$\text{Log}(Y) = b_0 + b_1{*}\text{age} + b_2{*}\text{age}^2 + b_3{*}\text{gender} + b_4{*}\text{position} + \text{residuals} \quad (2)$$

Results and discussion

The distribution of the population by fields in terms of numbers of researches and observations are given in Table 1. Medicine is the largest field and accounts for about one third of the population, in terms of both persons (4,021) and observations (11,974). With about 1,340 persons and about 3,700 observations, humanities is the smallest field, but with sufficient large numbers to analyse.

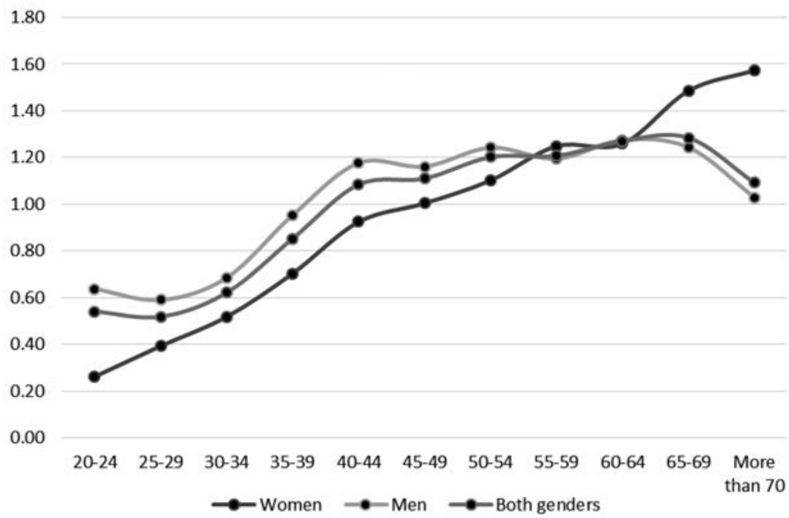
Figure 2 shows the average numbers of article equivalents per person per year by gender and fields. This figure shows that there are major inequalities in the publication patterns between the fields. The same result was also found by Prio, Aksnes and Rørstad (2013), and supports our decision that separate regression analysis of each field is justified.

Figure 2. Article equivalents (mean) by fields and gender (N=35,798)



The productivity for the whole population by gender and age groups shows an interesting picture. Productivity among women keep increasing by age almost in a straight line (the lower darker line in Figure 3), while the productivity among men has a curvilinear shape. Overall men are more productive than women up to the age of 55–59 years.

Figure 3. Article equivalents (mean) by gender and age groups for all fields (N=35,798).



Analysis of natural sciences

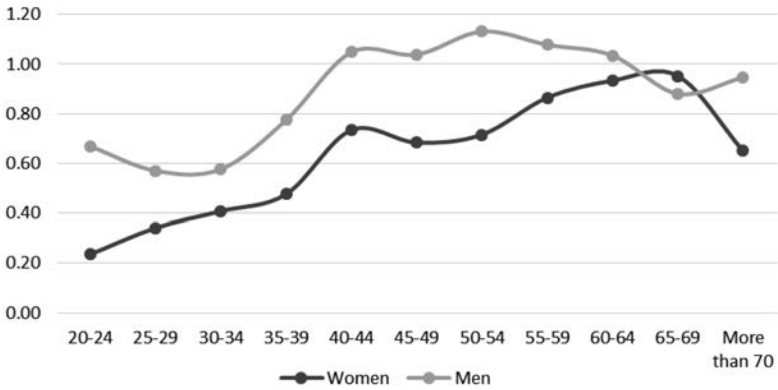
Table 3 shows the average number of article equivalents per person and year by academic position and gender. Overall, we can observe a decrease in productivity from professor towards PhD-students in the natural sciences. Moreover, for all positions female scientists tend to be less productive than their male colleagues. On average, men have 0.34 more article equivalents annually than women. The overall gender difference is larger than for the separate position categories due to a larger share of female researchers in the lower productive categories (PhD-students and Postdocs).

Table 3. Article equivalents (mean) by academic position and gender in the natural sciences (N=9,558)

Academic position	Women	Men	Both genders	Diff, Men-Women
Professors	0.91	1.14	1.11	0.23
Associate professors	0.61	0.86	0.80	0.25
Postdocs	0.44	0.65	0.59	0.21
PhD-students	0.35	0.48	0.43	0.13
Total/average	**0.53**	**0.87**	**0.79**	**0.34**

Divided by age groups, curvilinear relationships between publication output and age can be seen with several peaks. In addition, men are more productive than women are in all but one age-groups.

Figure 4. Article equivalents (mean) by gender and age groups in natural sciences. (N=9,558)



In order to test whether these variables can be used to predict the productivity level, we carried out an ordinary least squares regression analysis. The results from our regression models show that both gender and academic position are statistically significant variables. However, neither the beta coefficients for age nor age-squared are significant in our regression model. As seen in Figure 4, the productivity does not have a linear or a curvilinear relationship as assumed. Moreover, our regression analysis shows that PhD-students are less productive than their more experienced colleagues are.

Table 4. Regression analysis of natural sciences

Productivity	Beta Coefficients	Std. Err.	t-value	P>t	[95% Conf.	Interval]
Age	-0.025	0.025	-0.97	0.332	-0.074	0.025
Age squared	-0.002	0.002	-1.11	0.266	-0.006	0.002
Gender	-0.253*	0.024	-10.42	0	-0.301	-0.206
Professors	0.997*	0.044	22.42	0	0.910	1.084
Associated professors	0.664*	0.042	15.81	0	0.582	0.746
Postdocs	0.227*	0.032	7.11	0	0.164	0.289
PhD-student	*0.000*	*(reference)*				
Constant term	-0.984*	0.058	-16.84	0	-1.099	-0.869

R-squared = 0.1345, F-value = 247.33, N = 9,558, * significant at 0.05 level

Analysis of engineering and technology

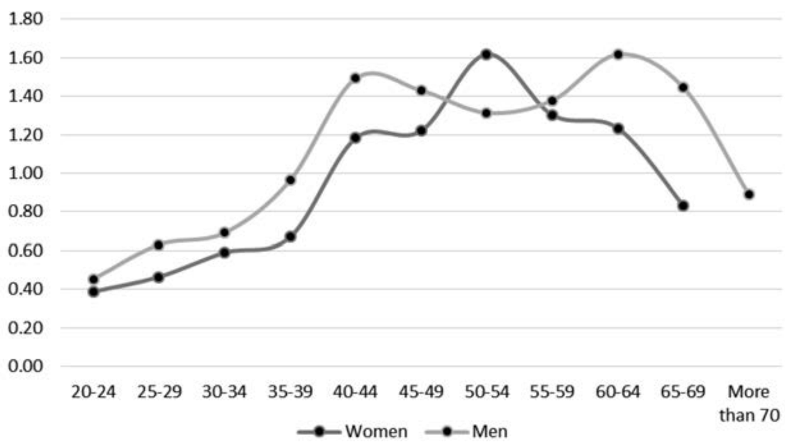
Table 5 shows that researchers within engineering and technology are slightly more productive (with 0.96 article equivalents per person per year) than their colleagues within natural sciences (with 0.79) (see also fig 2). Not surprisingly, professors are the most prolific personnel, followed by associate professors, postdocs and PhD-student. In this field, women account for less than 20 per cent of the observations and they publish 0.30 less article equivalents on average than men. The gender differences in productivity range from 0.08 for professors to 0.20 for postdocs.

Table 5. Article equivalents (mean) by academic position and gender in technology (N=4,754)

Academic position	Women	Men	Both genders	Diff. Men-Women
Professors	1.47	1.55	1.54	0.08
Associate professors	1.07	1.09	1.09	0.02
Postdocs	0.62	0.82	0.78	0.20
PhD-students	0.52	0.61	0.59	0.09
Total/average	**0.71**	**1.01**	**0.96**	**0.30**

If we look at productivity by age, there are different patterns for men and women (cf. Figure 5). For both genders, productivity increases by age up to about 40–44 years, then declines for men, while women remain at the same productivity rate for their next five year. While women reach a productivity peak around their mid-fifties, their male colleagues have a slight decrease, and then further increase their productivity rate to a peak in their sixties. This productivity pattern indicates that age and age squared should be included in a regression analysis for predicting productivity in engineering.

Figure 5. Article equivalents (mean) by age groups in technology (N=4,754)



Our regression analysis of engineering gives a model where all variables are significant. The size and sign of the beta coefficients provide information on the extent productivity is correlated to each of the independent variables. Both age and age squared are significant, but slightly less important than gender. However, academic position is the most important variable to express the productivity rate. Explained variance in this model is about 15 per cent, and slightly more powerful than we found for natural sciences (13.5 per cent).

Table 6. Regression analysis of technology

Productivity	**Beta coefficients**	**Std. Err.**	**t-value**	**P>t**	**[95% Conf.**	**Interval]**
Age	0.12*	0.03	3.8	0	0.06	0.18
Age squared	-0.01*	0.00	-4.58	0	-0.02	-0.01
Gender	-0.17*	0.03	-5.18	0	-0.23	-0.10
Professors	0.31*	0.05	6.12	0	0.21	0.41
Associated professors	*0.00*	*(reference)*				
Postdocs	-0.32*	0.06	-5.73	0	-0.43	-0.21
PhD-student	-0.43*	0.06	-7.72	0	-0.54	-0.32
Constant term	-0.55*	0.10	-5.34	0	-0.75	-0.35

R-squared = 0.1509, F-value = 140,56, N = 4,754, * significant at 0.05 level

Analysis of medicine

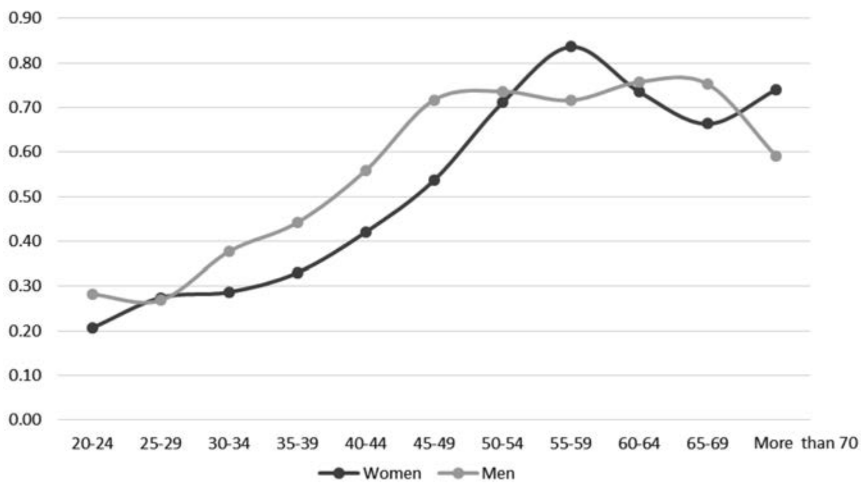
Table 7 shows the productivity by academic position and gender within the field of medicine. Overall men are more productive than their female colleagues are, but female professors and associated professors are slightly more productive than their male colleagues. On average, a male researcher within medicine publishes 0.63 articles per year, while a female researcher produces 0.47 article, which gives a difference of 0.16 articles per year.

Table 7. Article equivalents (mean) by academic position and gender in medicine (N = 11,974)

Academic positions	**Women**	**Men**	**Both genders**	**Diff. Men-women**
Professors	0.97	0.95	0.96	-0.01
Associate professors	0.71	0.66	0.69	-0.05
Postdocs	0.40	0.47	0.43	0.07
PhD-students	0.28	0.31	0.29	0.04
Medical doctors/physicians	0.38	0.50	0.47	0.12
Total/average	**0.47**	**0.63**	**0.57**	**0.16**

The scientific productivity as a function of age is shown in Figure 6. The productivity pattern is similar to the one fond for engineering. Women (the darker grey line) are less productive than their male colleagues within the same age group, up to their mid-fifties. Then they become equally productive, and in their late fifties a productivity peak can be seen, where female researchers are more productive than their male colleagues are.

Figure 6. Article equivalents (mean) by age groups, medicine (N=11,974)



The results of the regression analysis are shown in table 8. About 19 per cent (R-squared=0.1895) of the variance is explained in our regression model, and is thus a little more powerful than the models for the previous studied fields. All variables are also shown to be significant. Academic position is more important than age, which again is more important than gender in terms of productivity rate.

Table 8. Regression analysis of medicine

Variables	Beta coefficients	Std. Err.	t-value	P>t	[95% Conf.	Interval]
Age	0.153*	0.018	8.31	0	0.117	0.190
Age squared	-0.011*	0.001	-7.72	0	-0.014	-0.008
Gender	-0.083*	0.017	-4.9	0	-0.117	-0.050
Professors	0.306*	0.032	9.54	0	0.243	0.369
Associated professors	*0.000*	*(reference)*				
Postdocs	-0.351*	0.037	-9.56	0	-0.422	-0.279
Medical doctors/physicians	-0.405*	0.031	-13.02	0	-0.466	-0.344
PhD-student	-0.589*	0.034	-17.24	0	-0.656	-0.522
Constant term	-1.231*	0.066	-18.68	0	-1.360	-1.102

R-squared=0.1895, F-value=39.61, N=11,974. *Significant at 0.05 level

Analysis of social sciences

As shown in Figure 1, researchers in social sciences are more productive than researchers in the three previous studied fields in terms of article equivalents. The main reason for this is a different publication pattern in the social sciences compared to the previously studied fields. In the social sciences, one publication has in general fewer authors than a publication in medicine natural sciences and technology In addition; researchers in the social sciences publish more monographs.).

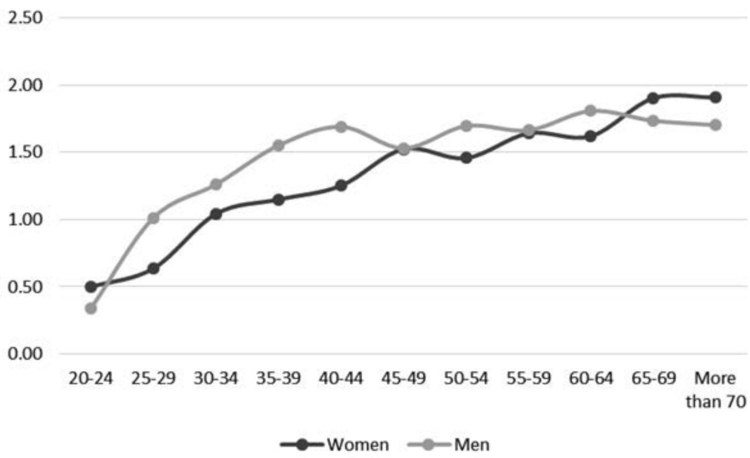
On average, a professor publishes about 1.77 article equivalents per year. In contrast to the other major fields, postdoc fellows are more productive than the associate professors, 1.53 and 1.44 article equivalents, respectively. The PhD-students have on average 0.96 article equivalents per year. For all academic positions, men are a slightly more productive than their female colleagues, on average 0.25 publications per person per year.

Table 9. Article equivalents (mean) by academic position and gender in social sciences (N=5,821)

Academic position	Women	Men	Both genders	Diff. Men-women
Professors	1.72	1.79	1.77	0.07
Associate professors	1.38	1.49	1.44	0.10
Postdocs	1.45	1.65	1.53	0.20
PhD-students	0.81	1.11	0.96	0.30
Total/average	**1.35**	**1.60**	**1.51**	**0.25**

Productivity versus age for researches in the social sciences is shown in Figure 7. Overall, the productivity rate shows a less steep increase by age compared to the hard sciences, but male researchers are more productive than their female colleagues are at the same age, except for the end of their career.

Figure 7. Article equivalents (mean) by age groups, social sciences (N=5,821)



The results of a linear regression model are presented in Table 10. As the explained variance is only 6.5 per cent, a linear regression model with three variables is not sufficient to explain the publication productivity in the social sciences. Age is not a significant variable and gender is just outside our 95 per cent confidential interval. Associated professors and PhD-students are significantly less productive than professors.

Table 10. Regression analysis of social sciences

Variables	Beta coefficients	Std. Err.	t-value	P>t	[95% Conf.	Interval]
Age	0.017	0.010	1.65	0.099	-0.003	0.036
Age squared	0.000	0.000	-1.31	0.192	0.000	0.000
Gender	-0.079	0.024	-3.22	0.001	-0.126	-0.031
Professors	*0.000*	*(reference)*				
Associated professors	-0.167*	0.031	-5.31	0	-0.228	-0.105
Postdocs	-0.115	0.050	-2.28	0.023	-0.213	-0.016
PhD-student	-0.492*	0.049	-9.95	0	-0.589	-0.395
Constant term	-0.303	0.258	-1.17	0.241	-0.809	0.204

R-squared=0.065, F-value=67.87, N=5,821. *Significant at 0.05 level

Analysis of humanities

Scholars in the humanities are overall more productive than the research personnel in the other major fields (see Figure 2). On average, a researcher within this field produces 2.02 article equivalents per year, which is twice as much as the average for the whole population of all fields. The main reason is a difference publication pattern, which has similarities to the one found for the social sciences, albeit with an even larger proportion of monographs.

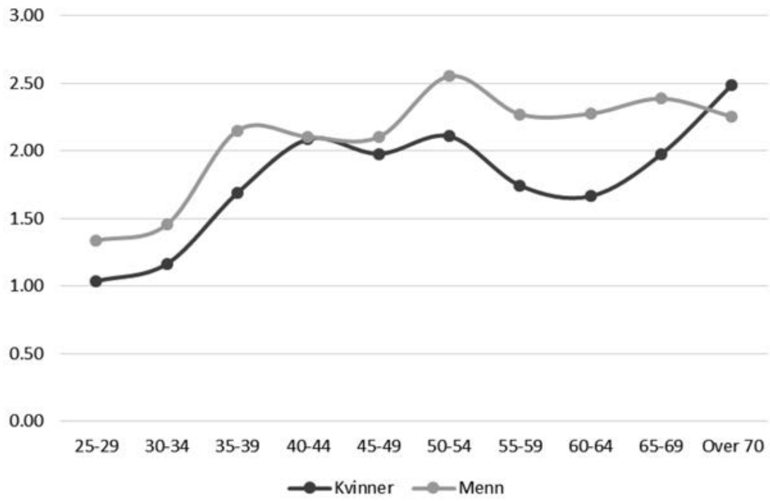
Male professors publish on average 2.39 article equivalents per year, and are more productive than their female colleagues who have an average output of 2.08 article equivalents per year. Male PhD-students are also a slightly more productive than the females, but for associated professors and postdocs, the gender difference is only minor.

Table 11. Article equivalents (mean) by academic position and gender in humanities (N=3,691)

Academic position	Women	Men	Both genders	Diff. Men-Women
Professors	2.08	2.52	2.39	0.44
Associate profes-sors	1.83	1.92	1.88	0.09
Postdocs	1.85	1.94	1.89	0.08
PhD-students	1.12	1.31	1.22	0.19
Total/average	**1.79**	**2.17**	**2.02**	**0.38**

The productivity pattern as a function of age and gender, show that male researchers are more productive than their female colleagues for all age groups except at their early forties. Overall, the productivity increases by age, but the relationship is not linear, and the curve has several peaks.

Figure 8 Article equivalents (mean) by age groups in the humanities (N=3,691)



The regression model for the humanities has less explanatory power than for the other fields. Nevertheless, male researchers are significant more productive than female. Academic position is also a significant variable and professors are significantly more productive than the personnel in the other positions are. Nor age or age squared are significant variables in our model.

Table 12. Regression analysis of the humanities (N=3,691)

Variables	Beta coefficients	Std. Err.	t-value	P>t	[95% Conf.	Interval]
Age	-0.021	0.038	-0.55	0.585	-0.096	0.054
Age squared	0.000	0.003	0.11	0.911	-0.005	0.006
Gender	-0.093*	0.026	-3.61	0	-0.144	-0.043
Professors	*0.000*	*(reference)*				
Associated professors	-0.222*	0.032	-6.87	0	-0.286	-0.159
Postdocs	-0.236*	0.053	-4.49	0	-0.339	-0.133
PhD-student	-0.599*	0.059	-10.15	0	-0.714	-0.483
Constant term	0.726*	0.136	5.32	0	0.458	0.993

R-squared=0.060, F-value=39.28, N=3,691. *Significant at 0.05 level

Discussion and conclusions

Our study has shown that productivity in terms of article equivalents can be expressed as a function of gender, age, age squared and academic position for natural sciences, engineering and

technology and medicine, and all variables are significant in our regression model. However, the regression analysis of the social sciences and humanities resulted in poorer models, and age is not a significant variable. Academic position has shown to be the most important variable for all fields, followed by gender.

If we look at professors only, the differences between men and women in productivity are rather small. The largest difference is found in humanities with 0.44 article equivalents, followed by natural science with 0.23. For engineering and the social sciences, the difference is less than 0.10, and in medicine 0.01 in favour female professors.

In this study of researchers at four Norwegian universities, we did not include institution as a variable. As the selected universities do not differ in terms of working conditions and time available to spend on research (e.g. for professors approx. half of their time), we did not expect any significant differences between the institutions in terms of productivity. However, in a second regression analysis, we did include institution as a variable to control for organizational differences in terms of publication output. This second analysis showed no or minor differences between the major fields.

Scientific productivity is strongly skewed at the level of individuals, this holds for all fields, positions, age groups. In the fields analysed, we have found that 6 to 19 per cent of the variance can be explained by our regression model. Thus, most of the variance is due to other factors. This is an interesting observation and may function as basis for further research on the topic.

References

Aksnes, D. W., Rørstad, K., Piro, F., & Sivertsen, G. (2011). Are Female Researchers Less Cited? A large-scale study of Norwegian scientists. *Journal of the American Society for Information Science and Technology (JASIST)*, 62(4), 628–636.

Barjak, F. (2006). Research productivity in the internet era. *Scientometrics*, 68(3), 343–360.

Cole, J.R., & Zuckerman, H. (1984). The productivity puzzle: Persistence and change in patterns of publication of men and women scientists. In P. Maehr & M.W. Steinkamp (Eds), Advances in motivation and achievement (Vol. 2pp 217–258). Greenwich, CT: JAI Press.

Cole, S. (1979). Age and Scientific Performance. *The American Journal of Sociology*, 84(4), 958–977.

Gonzalez-Brambila, C., & Veloso, F. M. (2007). The determinants of research output and impact: A study of Mexican researchers. [Article]. *Research Policy*, 36(7), 1035–1051.

Kyvik, S. (1990). Age and scientific productivity. Differences between fields of learning. *Higher Education*, 19, 37–55.

Kyvik, S. (1991). *Productivity in academia. Scientific publishing at Norwegian universities*. Oslo: Universitetsforlaget.

Kyvik, S. (1993). Academic staff and scientific production. *Higher Education*, 5, 191–202.

Kyvik, S., & Teigen, M. (1996). Child Care, Research Collaboration, and Gender Differences in Scientific Productivity. *Science, Technology, & Human Values*, 21(1), 54–71.

Long, J.S. (1992). Measures of sex differences in scientific productivity. Social Forces, 71(1), 159–178.

Piro, F, Aksnes, D. W., Rørstad, K. (2013). A macro Analysis of *Productivity differences across fields: Challenges in the Measurement of Scientific Publishing – Journal of the American Society for Information Science and Technology (JASIST)*, 64(2), 307–320.

Xie, Y. & Shauman, K.A. (1998). Sex differences in research productivity: New evidence about an old puzzle. American Sociological Review, 63(6), 847–870.

The Best Five – Do Authors' Assessments of their own Publications Correspond to Metric Based Rankings?[1]

Marion Schmidt schmidt@forschungsinfo.de / iFQ Institute for Research Information and Quality Assurance / Schützenstr. 6a / Berlin, 10117 (Germany)

Jörg Neufeld neufeld@forschungsinfo.de / iFQ Institute for Research Information and Quality Assurance / Schützenstr. 6a / Berlin, 10117 (Germany)

Introduction

For some years now, the on-going discussion on the validity of bibliometric indicators in research evaluation and especially their potential effects on scientists – who might try to publish as much as possible – have evolved into a paradigm which can be described by the phrase "Quality not quantity"[i]. Funding agencies like the German Research Foundation (DFG), the European Research Council (ERC), the American National Science Foundation (NSF) as well as the National Institute of Health (NIH) have adopted policies which demand scientists to cite or highlight only a limited number of their most important publications in funding proposals. Reviewers of funding decisions are to be focussed on those selected papers as well as on the proposal itself. The respective selection of papers is supposed to be based on the scientists' subjective ratings which are not expected to correspond completely with bibliometric impact measures. Applicants therefore have the possibility to highlight publications that are especially innovative – which may not be adequately reflected by bibliometric indicators. However, as far as researchers adhere to metrics – or believe the funding agency or the reviewers being adhered to metrics, they will simply base their selection on just these metrics. On the other hand, subjective researcher ratings and metrics may conform with each other although the respective researchers' selection is independent from metrics.

On a wider level, the funding agencies' policy change aims to generally counteract the importance of metrics in evaluation of science, allocation of funding and appointments.

How do selected papers actually fare when evaluated with bibliometric indicators? Subjective ratings of their own papers by researchers have been analysed quite rarely although they are apparently getting more and more important to funding decisions. Based on a sample of the publication corpora of highly cited scientists, Aksnes (2006) asked them how they assessed the scientific contribution of their papers, and to categorize their papers as empirical, theoretical, methodical, or review. He observes overall correlations of 0.56 and 0.52 between raw and field-

1 This work was supported by the German Federal Ministry of Education and Research (BMBF)

normalized citation counts, respectively, and authors' perception of their papers' scientific contribution. For the empirical, theoretical and methodical papers, there is no significant difference to the overall value, while for the reviews, the correlation is weaker.

Porter et al. (1988) gather two cohorts of Sloan Chemistry Fellows and ask them to augment publication data and nominate their three best papers. In a basic probability estimate, they calculate a ratio of observed to expected overlap as 34% to 11% between best and most cited papers in the 1974 cohort. They use content analysis in order to categorize papers as theoretical, empirical and methodological papers with the result that most cited papers are methodological to a larger extent, whereas best ones are more often theoretical and empirical.

A number of studies have compared bibliometric indicators with peer ratings:

Rinia et al. (1998) report significant correlations between peer ratings with citation-based indicators, notably citations per publications and the field normalized citation rate, in physics research groups. Van Raan (2006) compares the h-index and the field normalized citation rate with peer judgments. Both indicators discriminate very well between excellent or good chemistry research groups on the one side and less good on the other side. Similar results are reported by Moed (2005).

Bornmann and Leydesdorff (2013) correlate peers' ratings of papers based on the online service *Faculty of 1000* with citation-based indicators, of which the ratio of highly-cited papers correlates best. Franceschet & Costantini (2010) report significant correlations between peer ratings and raw citations, and, to a lesser degree, journal impact factors in a study comprising several fields.

Some studies are concerned with scientists' general perception of citations:

Aksnes and Rip (2009), in a follow-up study of Aksnes (2006), conclude that, on the individual paper level, citations are not a reliable indicator as quite a few scientists perceive own papers as under- or overcited. An older survey of academics' views of citations and evaluative bibliometrics is presented by Collins (1991). Most respondents accepted quantitative indicators as part of research assessment, but favoured productivity indicators over citation indicators. Hargens and Schuman (1990) analyse the relationship between biochemists' and sociologists' use of citation data and the field-normalized medium citation score of their own publication corpus. The number of citations a researcher receives is only weakly correlated with his usage of citation indexes in case of biochemists, but the correlation is significantly stronger in case of sociologists.

This research in progress paper evolved from a service project which was conducted by the iFQ recently and which was focussed on calculating bibliometric indicators for a benchmarking process of three German universities in chemistry and physics. As part of a publication validation process, scientists' subjective ratings of own papers have been requested and are compared with several state-of-the-art bibliometric indicators. In this paper, we analyse to which extent the best

papers according to subjective assessment are concordant with the best papers identified by metrics; and furthermore, which indicator shows best concordance.

Data and Methods

Data and bibliometric indicators

Publications of scientists belonging to chemistry and physics institutes of three universities have been searched in *Web of Science* [ii] by way of institutional addresses and names of currently employed scientists. After that, these scientists have been contacted and invited to validate the publication data from 2005 and 2010 searched for them, and if necessary to delete or add publications or to upload own publication lists. They also have been asked to mark five publications they considered being their best in the period from 2005 to 2010, independently of institutional affiliations. They were asked in exact wording: "In order to answer the question whether bibliometric indicators correspond to your own assessment we ask you to flag up to five publications you consider your best."[iii]

In comparison to Asknes (2006) and Porter et al. (1988), who both preselected their respective samples based on citation numbers and funding applications, our sample includes scientists in different career steps and irrespectively of their specific publication performance.

For each of the 182 scientists who have marked their best publications, all citable items[iv] between 2005 and 2009 have been selected. A threshold of 10 publications per person has been applied resulting in a sample of 109 persons. For all publications the following indicators have been calculated:

– citations per paper (within 3-year citation window), (CPP),
– field-normalized citation rate (within 3-year citation window), (FNCR),
– journal-normalized citation rate (within 3-year citation window), (JNCR),
– Journal Impact Factor, (JIF)[v].

Due to the fact that relevant portions of a year's publications are entered into WoS during the first months of the following year, we have omitted the publication year 2010 and will include these publications in an updated calculation in summer. The same holds for the possibility of considering the field-normalized ratio of highly-cited papers whose distribution among the analysed papers between 2005 and 2009 is too sparse.

Comparing researcher's and metrics based rankings

(1) Cohen's Kappa

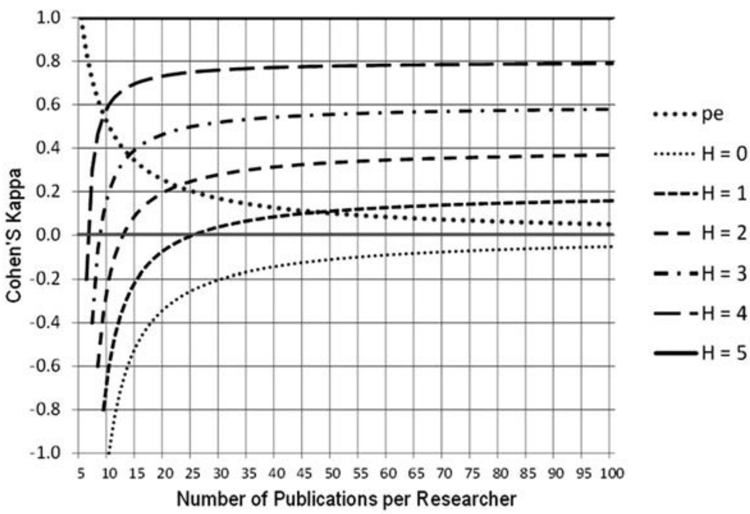
Depending on the total number of publications a researcher has published, there is a certain probability that researcher assessment and metric based ranking coincide by chance. We therefore chose Cohen's Kappa (*K*) as a measure of accordance because it takes into account this randomly expected concordance[vi].

$$K = \frac{p_0 - p_e}{1 - p_e}$$

p_0 = observed relative agreement
p_e = concordance expected by chance

Kappa varies between -1 and +1 whereby 0 indicates the accordance of empirical concordance and the concordance expected by chance. Figure 1 illustrates how Kappa relates to the number of publications and the number of concordant judgments.

Figure 1. Values of Kappa and expected concordance (p_e) depending on the number of publications per researcher and the number of hits (H) – provided that exactly five publications have been marked as best.



Obviously with a growing number of publications the values of Kappa converge to limiting values which are equal to the respective shares of concordant judgments.

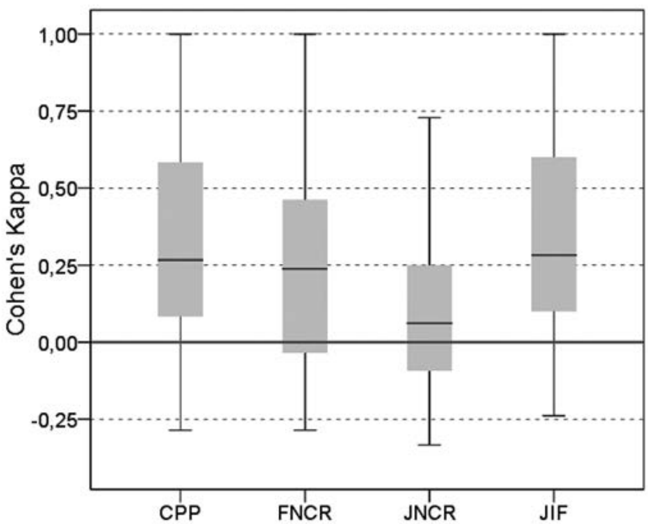
(2) Median relative rank of the best five

Kappa ignores the ordinal information, which is given by the metric based rankings. Each publication which is marked as “best five” by a scientist has its specific position in the metric based ranking of all publications of this scientist. For example, Kappa treats a best five publication which is ranked sixth regarding CPP as discordant (no hit) even if the number of publications is quite high. By contrast the *median relative rank* of the best five reflects the relative position of the best five in the metric based rankings. Medians beneath 0.5 indicate scientists who tend to evaluate their publications contrary to the metric based rankings and medians above 0.5 point to a similar/equal assessment.

Results

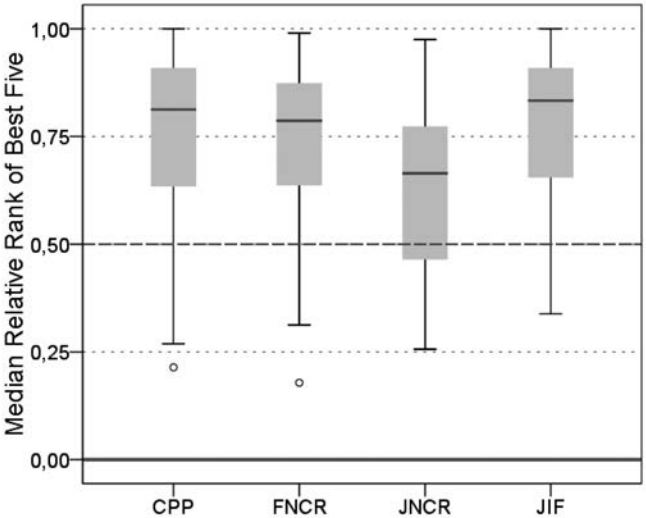
For each researcher we calculated four values of Kappa, one for each indicator (CPP, FNCR, JNCR, JIF). Figure 2 shows boxplots for each distribution of Kappa. Medians of CPP, FNCR and JIF are located near to .25. Based on usual evaluation scales for Kappa this indicates rather weak fit of scientists' judgments and metrics based rankings. However, values are quite dispersed; roughly 25 per cent of the scientists show values of .60 or higher regarding CPP and JIF. The accordance of scientists' judgments and source normalized impact (JNCR) is even smaller and barely better than would be expected by chance.

Figure 2. Boxplots of Cohen's Kappa measuring accordance between scientists' ratings (best five) and indicator based ranking.



The boxplots of the median relative rank in Figure 3 show the same pattern for all indicators. Again the correlation between source normalization (JNCR) and scientists' judgments is weaker than with the other indicators. Apart from the JNCR, the median relative rank of the best five publications is located in the upper third of the distributions in 75 per cent of the cases. Roughly 50 per cent are located in the top quintile (> 0.8).

Figure 3. Boxplots of median relative rank of the best five publications in metric based rankings.



Discussion

As the results obtained so far demonstrate, all in all researcher assessments and indicator based rankings are not perfectly consistent. However, values are dispersed, and regarding CPP, JIF and FNCR some cases show a perfect match. It still remains an open question whether these scientists actually base their perception of their own papers on the citation scores or JIF – which they can easily access via the databases' online interfaces – , or whether the concordance is merely a product of the underlying quality of the papers that leads both to citations and a good subjective rating (which in turn could also lead to submitting the paper to a journal with a higher impact factor, which in turn may lead to higher visibility and more citations). In the course of our project we would like to further investigate this issue by analysing the scientists' specific patterns. The next planned step is to survey them about the specific criteria for perceiving and defining publications as best five and to relate subjective assessments with respective author positions and estimations of their own contribution to a paper.

Regarding the funding agencies' policies the question arises if the desired effect of giving scientists the opportunity to behave blind to metrics and highlight unconventional research is actually realistic or if scientists' view of the world tends to be already shaped by bibliometrics. Funding agencies which are interested in a selection of best five publications independent of metrics should explicitly state this in their guide lines. Possibly they could even state criteria (like innovativeness). However, when, as in the case of the ERC (European Commission 2012: 20), applicants are invited to quote citation counts of the named best five, there is a risk that the selection of the best five is based just on these counts.

References

Asknes, D. W. (2006). Citation Rates and Perceptions of Scientific Contribution. *Journal of the American Society for Information Science and Technology*. 57, 169–185.

Asknes, D.W. & Rip, A. (2009). Researchers' perceptions of citations. *Research Policy*, 38, 895–905.

Bornmann, L. (2011). Scientific Peer Review. *Annual Review of Information Science and Technology*, 45, 199–245.

Bornmann, L. & Leydesdorff, L. (2013). The validation of (advanced) bibliometric indicators through peer assessments: A comparative study using data from InCites and F1000. *Journal of Informetrics*, 7, 286–291.

Cicchetti, D. V. & Feinstein, A. R. (1990). High Agreement but Low Kappa: II. Resolving the Paradoxes. *Journal of Clinical Epidemiology*, 43, 551–558.

Collins, P. M. D. (1991): *Quantitative assessment of departmental research. A survey of academics' views*. SEPSU Policy Study No. 5. Science and Engineering Policy Studies Unit of the Royal Society and the Fellowship of Engineering, London.

DFG (2010). *Quality not Quantity – DFG Adopts Rules to Counter the Flood of Publications in Research*. Press Release. No. 7, 23. February 2010. Retrieved March 11, 2013 from http://www.dfg.de/en/service/press/press_releases/2010/pressemitteilung_nr_07/index.html.

European Commission (2012). *ERC Work Programme 2013*, European Commission, C(2012) 4562 of 09 July 2012.

Feinstein, A. R. & Cicchetti, D. V. (1990). High Agreement but Low Kappa: I. The Problems of Two Paradoxes. *Journal of Clinical Epidemiology*, 43, 543–549.

Franceschet, M. & Costantini, A. (2010). The first Italian research assessment exercise: A bibliometric perspective. *Journal of Informetrics*, 5, 275–291.

Hargens, L. L. & Schuman, H. (1990). Citation Counts and Social Comparisons: Scientists' Use and Evaluation of Citation Index Data. *Social Science Research*, 19, 205–221

Moed, H. F. (2010). New Developments in Electronic Publishing and Bibliometrics, ESSS Summer School, Berlin, 16 June 2010. Retrieved March 11, 2013 from http://www.scientometrics-school.eu/images/esss1_Moed.pdf

Moed, H. F. (2005). *Citation Analysis in Research Evaluation*. Information Science and Knowledge Management, 9. Dordrecht: Springer

Porter, A. L., Chubin, D. E. & Jin, X.-Y. (1988). Citations and Scientific Progress: Comparing bibliometric measures with scientist judgements. *Scientometrics* 13, 103–124.

Rinia, E. J., van Leuween, T. N., van Vuren, H. G. & van Raan, A. F. J. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria – Evaluation of condensed matter physics in the Netherlands. *Research Policy*, 27, 95–107.

Van Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67, 491–502.

i http://www.dfg.de/en/service/press/press_releases/2010/pressemitteilung_nr_07/index.html

ii In a bibliometric database based on WoS raw data.

iii We abstained from a more elaborate survey because information letters and validation tool should have not been too complex.

iv Articles, Letters, Reviews published in journals.

v This indicator has been chosen because we assumed that scientists are rather attentive to the impact factors of the journals they publish in. The journal impact factor of a given journal and year is calculated by us in accordance with Thomson Reuters as ratio of the number of citations to all documents of two preceding years and the number of all articles and reviews of that journal in these two years (see Moed, 2010).

vi Note that in every case both researchers and rankings have the same number of positive (normally five) evaluated publications. Therefore the issues concerning the Kappa coefficient discussed by Feinstein & Cicchetti (1990) and Ciccetti & Feinstein (1990) – e.g. symmetry/ asymmetry of the marginal totals – are basically not relevant in this context.

Marks as indicators for analysing the service orientation of universities

Ulrich Schmoch schmoch@uni-speyer.de / German University of Administrative Sciences, Freiherr-vom-Stein-Str. 2, Speyer. 67346 (Germany)

Abstract

Science policy increasingly requires that universities engage in transfer activities with a focus on technology transfer. With the considerable weight of the service sector the service transfer will get more attention. An appropriate indicator for analysing service innovations are service marks. This type of investigation reveals a growing activity of German universities in services with a focus on education, in particular further education. The mark analysis shows a relevant competition in the market of higher education with a high number of foreign universities. In other fields like financial and business consultancy or medical services the German universities show a certain activity as well, first of all university professors privately.

Introduction

In the last 20 years, European universities have to cope with the enormous growth of student numbers. In addition to their traditional missions education and research, the universities are expected to be more active in the "third mission" transfer. This explicit orientation on transfer gets relevance also in education, as it can introduce "practical" elements. Professors with a relevant involvement in transfer will introduce more practice-oriented topics in their teaching. The analysis of transfer is a major topic of scholarly discussion under the label of "university-industry cooperation". An important tool of analysis is patent indicators, but patents only refer to technology. Now more than half of the employees in Germany work in the service sector (EFI 2012: 148 ff.), in other countries even more. Therefore the universities should also be active in the transfer to the service sector, a topic which is dealt with by only a few publications, e.g., by Chesbrough & Spohrer (2006). A potential tool for analysing this aspect may be marks, as marks cover goods as well as services, in particular knowledge-intensive services. Schmoch (2003) and Mendonca et al. (2004) already showed that marks are an appropriate innovation indicator in the service sector as complementary to patents. The following investigation examines the service activities of universities by means of mark indicators.

Methodology

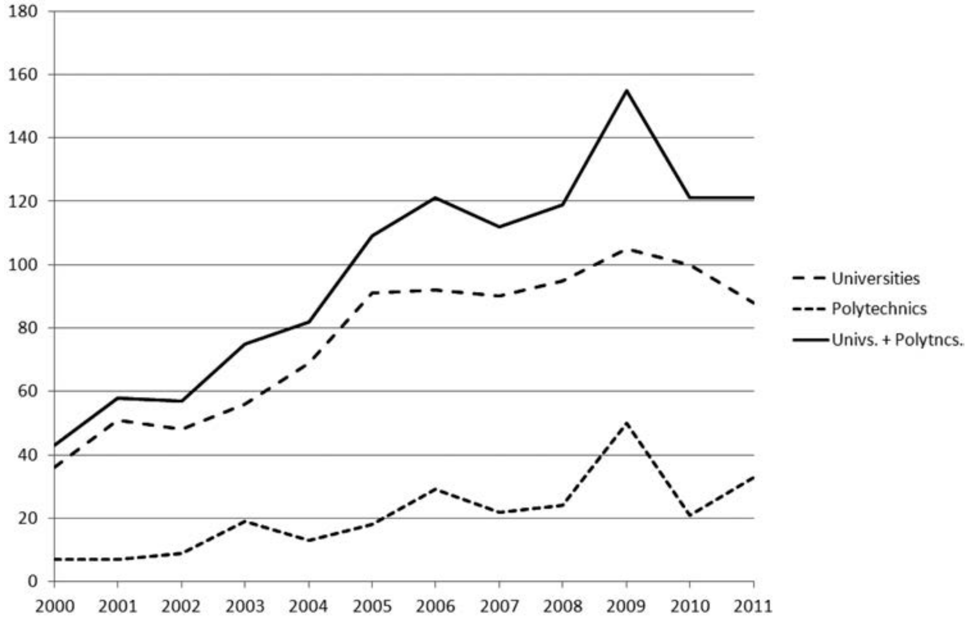
The analysis is based on German marks as provided by the DPMAregister of the German Patent and Trademark Office, thus on domestic marks of German universities. We have to be aware of the different character of patents and marks. Patents refer to technical inventions with the option of marketing referring products. Marks aim at the protection of names referring to products and services, and by a mark other firms can be prevented from using this name for similar products or services. Marks can be divided into trademarks referring to products and service marks. Marks protect products and services already offered in the market, thus the economic stage is one step further than in the case of patents. In the case of marks, the non-use in the market is a legal reason for cancellation.

Results

As universities are no producing enterprises, it can be expected that the number of marks is much lower than in the case of patents, as in the latter case a production can be realised by a license to a manufacturing enterprise. Thus, the universities must not manufacture the products on their own terms. In the case of marks, the universities have to provide the services themselves. As universities are no commercial service enterprises, their number of marks will be rather low.

Looking at mark applications of German universities, the first observation is that they exclusively own marks referring to services, products only appear in addition to services. A second observation is that the level of mark applications by universities is quite low with about 0.6 % of all German service marks compared to a share of about 5 % of all German patent applications. But since 2000, the number of mark applications by German universities and polytechnical schools (applied universities) is steadily increasing from about 40 in 2000 to 120 in 2011 (Figure 1). The data for 2011 are complete, as for mark applications, the data are published about half a year after the application, thus much earlier than for patent applications (18 months). The majority of these institutional applications refer to universities, a minor one to polytechnical schools. This finding meets the expectations, as one the one hand, polytechnical schools are more applied, so that they are close to transfer, but on the other hand, they have less resources to engage in activities outside teaching to a larger extent. Furthermore, the number of students in Universities is destinctly higher by a factor of 1.9 than those in polytechnical schools.

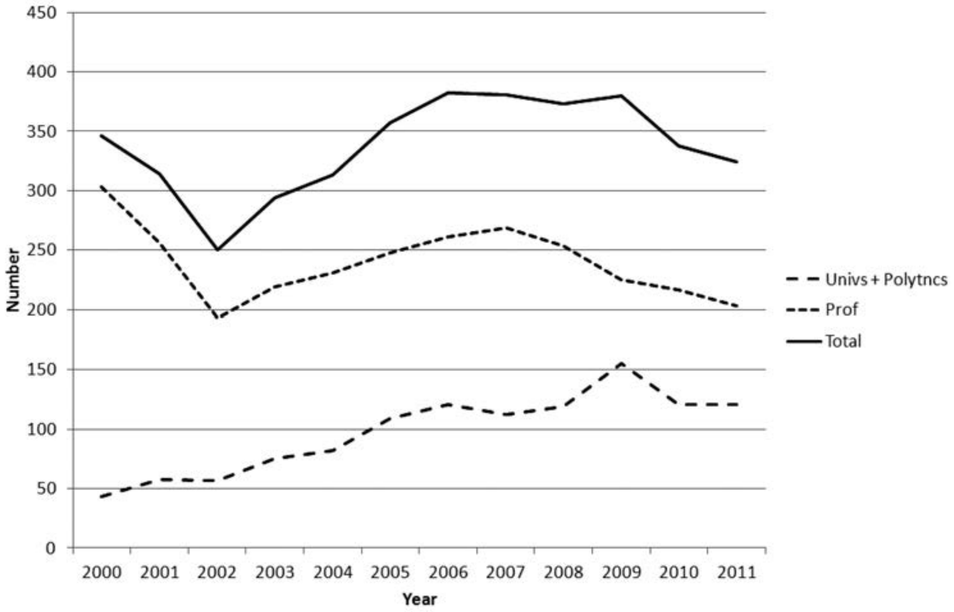
Figure 1. Number of German or European mark applications of German universities and polytechnical schools by application year



Source: DPMAregister, searches and compilation by the author

A substantial number of service marks is not applied institutionally by the universities or the polytechnical schools, but by the professors individually. In many cases, the professors have established their own enterprises and offer their services in the market. As to professors, the database does not offer the possibility to distinguish professors from universities and those from polytechnical schools. Comparing the number of institutional and individual applications of German origin, the number of applications by professors proves to be even higher than the institutional ones. Whereas the applications by universities and polytechnical schools increased since 2000, the number of applications by professors decreased, but is still higher than the institutional one.

Figure 2. Number of German or European mark applications of German universities and polytechnical schools and of German professors by application year



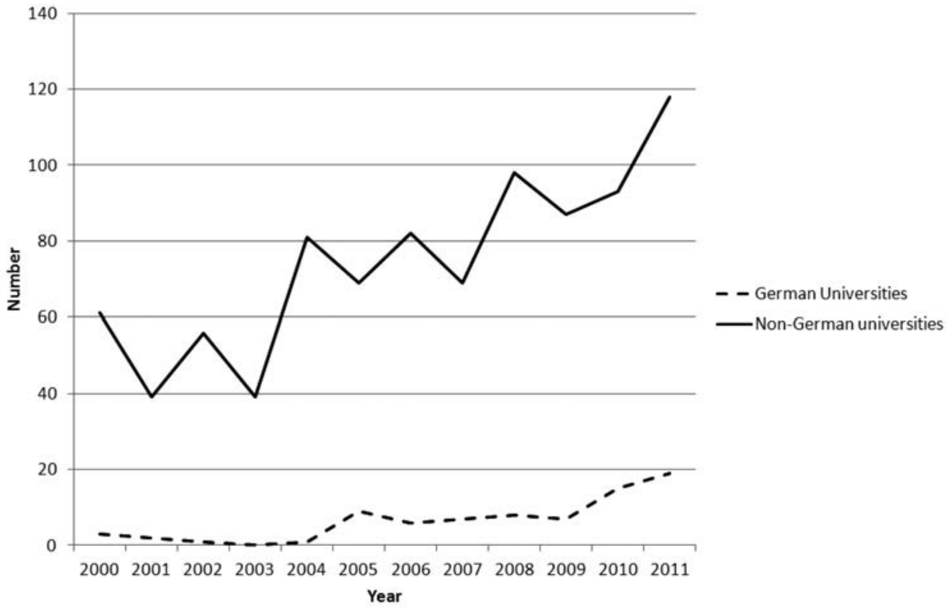
Source: DPMAregister, searches and compilation by the author

Most marks in the higher education sector are applied as German marks at the German Patent and Trademark Office (DPMA) with validity in Germany. A small part is applied at the Office for the Harmonization of the Internal Market (OHIM) with validity for the whole European Union. these European marks primarily belong to universities and refer to education, the class 41 of the so called Nice Classification for marks. In the case of education, the marks often refer to courses in addition to the standard ones such as further education with relevant tuition fees, thus with a definite economic aspect. Meanwhile a substantial competition in the market of higher education can be observed (Hahn 2003), reflected in a growing number of European marks of universities with different national origin. Therein, the German universities increasingly engage as well (Figure 3).

By disaggregating by mark classes, the majority of service marks of German universities and polytechnical schools is applied in the classes for education and computer services (Figure 4). On third position, but already at a much lower level are applications in business services. There is also a relevant number of marks in communication (telecommunication), medical and personal (in general legal) services, but education is dominant and only in computer services a substantial transfer to other fields appears. The core area education of universities and polytechnical schools, the higher education institutions (HE institutions), is also reflected in service marks. As to the

appropriate interpretation of data, it has to be taken into account that in most cases, the marks refer to several classes, e. g., for 2009 to 2011, 2.5 classes per application on average., and in most cases either education or computer services are one of theses classes.

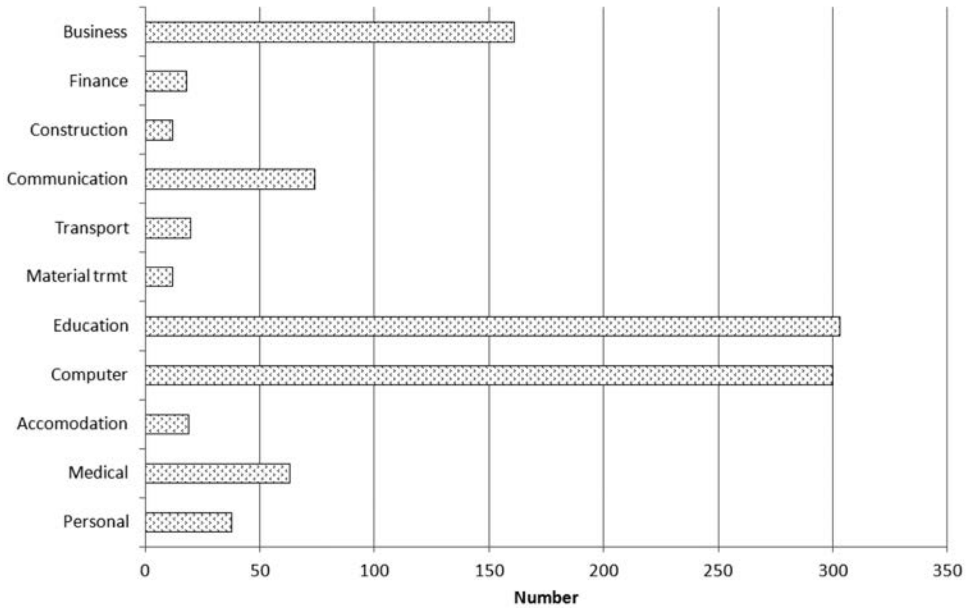
Figure 3. Number of European marks of universities with different national origin



Source: DPMAregister, searches and compilation by the author

The size of the classes differs considerably. Therefore specialisation indexes are often calculated where the absolute numbers in one class are normalized by the referring totals, in this case all marks with German origin. For HE institutions, also in terms of specialisation a clear focus on education and computer services can be asserted (Figure5).[i] The computer services can be interpreted as transfer to industry, but still with a close link to technology. For the analysis, the broader period of 2009 to 2011 was selected for achieving sufficiently high numbers of items for statistically relevant results.

Figure 4. Number of German or European mark applications of German universities and polytechnical schools by service classes, 2009–2011



Source: DPMAregister, searches and compilation by the author

Furthermore it is revealing to look at the specialisation pattern of marks of professors. Although the specialisation of HE institutions versus professors is similar in education, the professors have a stronger focus on financial and medical services (Figure 5). Obviously they have a more distinct activity in sectors of economic enterprises such as financial and business services, and this may be described as transfer of (commercial) services. In addition, they have a clear focus on medical services.

Mark classes cover quite broad areas. Only 45 classes for goods and services are available (Nice Classification), thereof 10 service classes. This is a very crude classification compared to the about 75,000 sub-groups of the International Patent Classification (IPC). The patent offices provide fines classifications for marks as orientation where specific goods or services should be referred to. However, these finer codes o not appear in official mark applications. However, in many mark applications, the specific goods and services are indicated in addition to the broad official definitions of the classes. Therefore, it is possible to get better insights in the contents of mark applications by looking into samples of mark documents with reference to specific classes.

With reference to marks of HE institutions, we examined the documents referring to relevant classes more in depth.

In the case of class 35, "Advertising; business management; business administration; office functions", the marks of HE institutions comprise business consulting, but also the building up and updating of databases, market research, opinion research, expert opinions, thus contract research for enterprises. In the case of professors, business consulting, e. g., expert opinions, are more prominent.

The class 36, "Insurance; financial affairs; monetary affairs; real estate affairs", is primarily important for private marks of professors. They offer financial consulting or consulting for insurances. This can be considered as real knowledge transfer to the economy, thus transfer in services. However, the activity of In the case of class 41, "Education; providing of training; entertainment; sporting and cultural activities", most marks refer to the offer of all types of higher education, in particular special offers for education in parallel to a job, but also the organisation of conferences or cultural events as well as the publication of scientific or educational documents appears in the specifications of the marks.

Class 42, "Scientific and technological services and research and design relating thereto; industrial analysis and research services; design and development of computer hardware and software. comprises much more than computer-related services, but also scientific and research services. It is obvious that HE institutions are quite prominent in this area. However, scientific and research services may refer to all fields of science and research, not necessarily to the service sector.

The Class 44, "Medical services; veterinary services; hygienic and beauty care for human beings or animals; agriculture, horticulture and forestry services", primarily refers to the private activities of professors. But in many cases, there is a co-classification with education (class 41) and the further education of physicians is explicitly mentioned. The latter activities may be addressed as service transfer form HE institutionns.

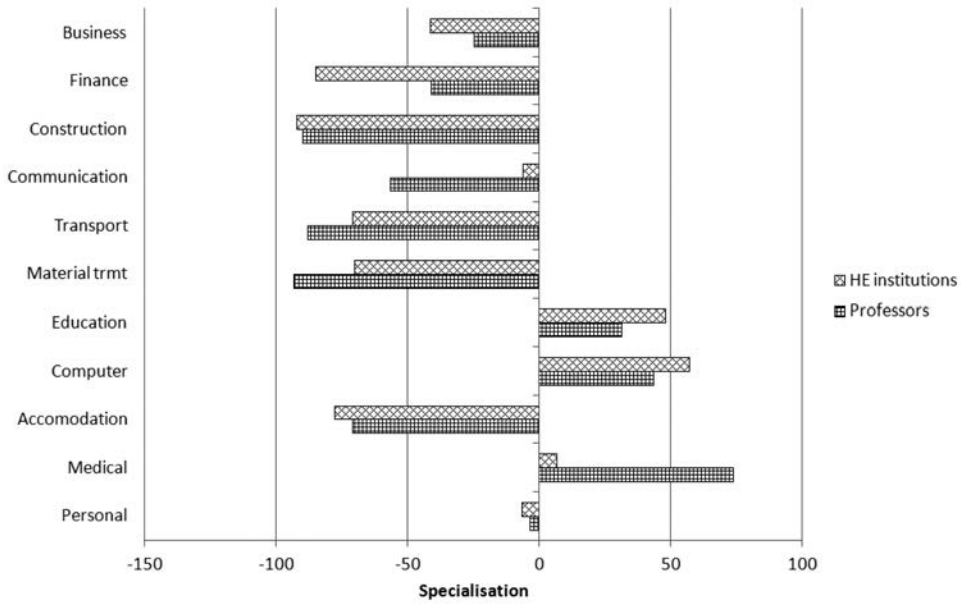
Conclusions

Service marks prove to be a useful indicator for analysing the activities of HE institutions in the service sector. Although the number of mark applications by HE institutions is rather low, major pattern of service activities get visible. First, the activities in services increase reflected by a growing number of service marks since 2000. A clear dynamics in EU marks in education points to a strong international competition in the offer of HE courses, in particular post-graduate courses for people already working in industry or the service sector in the context of lifelong learning. Second, the number of marks of professors is even higher than marks of HE institutions themselves. The focus of these individual marks is on education as well, but professors exhibit more activities in economic fields such as financial or business services and in medical services. Thus the transfer in services to economic applications is primarily linked to the activities of individual professors. However, the decreasing number of individual marks in parallel to a growing number of institutional ones might be interpreted as a shift from individual to institutional marks, and it will be interesting to track whether this trend will be associated to a stronger activity

of HE institutions in fields with relevance for transfer to economy. At least in the present situation, transfer in services is obviously no major topic in the agenda of HE institutions.

The results of this paper refer to Germany. It will be interesting to compare the German structures to those of other countries such as Great Britain, France, Italy or the United States. Methodically, it will be important to analyse domestic marks, as International Marks (IR marks) or European marks primarily reflect service exports. Only domestic marks show the structures of services of HE institutions for domestic clients. The international marks, however, are useful for the analysis of the international competition in the HE education market.

Figure 5. Specialisation of mark applications by German HE institutions and professors, 2009–2011.



References

Chesbrough H. & Spohrer, J. (2006): A research manifesto for services science, *Communications of the ACM*, Vol 49, No. 7, 35–40.

Expertenkommission Forschung und Innovation (EFI) (2012). Report 2012. Research, Innovation, and Technological Performance in Germany, Berlin: EFI.

Karola Hahn (2003). Die Globalisierung des Hochschulsektors und das ‚General Agreement on Trade in Services' (GATS). *Die hochschule*, 1, 48–73.

Sandro Mendonça, S., Santos Rereira, T. & Mira Godinho, M. (2004). Trademarks as an indicator of innovation and industrial change. *Research Policy*, 33(9), 1385–1404.

Schmoch, U. (2003). Service marks as novel innovation indicator, *Research Evaluation*, 12(2), 149–156.

i The specialization index covers a range from -100 to +100. Positive values indicate specializations above average, negative ones activities below average. Values above 30 refer to distinctly positive orientations, values below -30 to distinctly negative ones.

Is post-doc funding stimulating research performance? Evidence from Danish research policy!

Jesper W. Schneider jws@cfa.au.dk / Danish Centre for Studies in Research & Research Policy, Department of Political Science & Government, Aarhus University, Bartholins Allé 7, Aarhus, DK-8000 C (Denmark)

Thed N. van Leeuwen leeuwen@cwts.leidenuniv.nl / Centre for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62a, Leiden, 2333 AL (the Netherlands)

Abstract

We present main results from the bibliometric part of a recent evaluation of two different post-doctoral funding instruments used in Denmark. We scrutinize the results for robustness, sensitivity and importance, and eventually come out questioning the official conclusions inferred from these results. We specifically examine whether there is a long-term citation performance difference between groups of researchers funded by the two instruments. Through an elaborate matching process, potential differences in performance are also compared to a control group of researchers that has not received postdoctoral funding, but otherwise are comparable to the postdoc-groups. Hence, we are also able to indicate the effectiveness of being postdoc-funded, given its benefits, when it comes to citation performance. The results show that all three groups perform well above the database average impact, however, we conclude that there is no difference in citation performance between the two postdoc-groups. There is, however, a difference between the postdoc groups and the control group, but we argue that this difference is trivial and certainly not robust. Our conclusion, contradicts the official conclusion given in the evaluation, where the Research Council emphasizes the success of their funding programs and neglects to mention the rather high performance of the basically tenure-tracked control group. We speculate that reservations against tenure-track positions may lay behind this conclusion? We demonstrate and stress that indicators should come with robustness and sensitivity analyses, as well as some sort of yardstick, in order to make decisions concerning the importance of findings.

Introduction

Postdoctoral research fellowships are most often designed to facilitate the transition of a research career for young researchers. The Council for Independent Research in Denmark (DFF) has for more than a decade supported a large number of junior researchers through postdoctoral (postdoc) funding programs (DFF, 2012). Basically two instruments have been used. An instrument where individual researchers apply for and receive grants to fund their own research (for short "individual postdocs"), and an instrument where junior researchers are funded within larger

research projects, where typically more senior researchers are the principle investigators and the names of the intended postdocs can either be specified in the application, or alternatively, only the amount expected to be used to fund a postdoc position is given (for short "embedded postdocs"). In the latter case, specific junior researchers are selected following successful funding of the research project and their *curriculum vita* therefore plays no role in the funding review process.

Evidence for the effectiveness and benefits of such postdoc-programs remains vague. Attention is most often directed towards questions concerning application and selection, e.g., do the "best" young researchers apply and are the "best" selected for funding, where "best" is typically defined and measured in terms of publication performance (which consists of output and impact analyses). (e.g., Bornmann & Daniel, 2006; Bornmann, Wallon & Ledin, 2008; Böhmer, Hornsbostel & Meuser, 2008; Hornbostel et al., 2009; Bornman, Leydesdorff & van den Besselar, 2010; Neufeld & von Ins, 2011; Neufeld & Hornbostel, 2012). Discriminating between young researchers, often coming straight from a PhD-program, at the time of application based on past publication performance can be challenging due to the possible short active publishing career at that point. Further, if resources are allocated to those young researchers already on their way to a productive career, the marginal impact of the intervention may turn out to be small. In a regression-discontinuity designed case study of National Institutes of Health's (NIH) F32 individual postdoc fellowships, Jacob and Lefgren (2011) find that the F32 program does appear to increase the amount of health science research, as well as the number of individuals engaged in a biomedical research career (i.e., not leaving academia). Jacob and Lefgren (2011) argue that among benefits for those receiving funding, are a higher probability of getting funded by NIH in the future; more time to research and less teaching obligations; easier to establish networks with high performing peers; and/or a general increase of the postdoc's visibility in the domain. Again visibility is defined and measured in terms of publication performance, i.e., publication activity and citation impact.

As mentioned above, in Denmark two funding instruments have been in place for more than a decade, i.e., individual and embedded postdocs. During this period, the DFF has experienced a rising pressure on postdoc applications due to large intake of PhD-students into the Danish research system from 2004 and onwards. A general concern among the Board of Directors in DFF is whether funds are used to best effect. DFF assumes that postdoc-funding in general is beneficial to the individual researcher and his or her future career, however, at the same time DFF suspects that the funding instrument, where individuals are funded directly, is superior compared to the embedded funding instrument, and if so, money is perhaps best spent by exclusive support to the individual postdoc instrument? The rationale is that the selection process in the individual postdoc instrument, *ceteris paribus*, ensures that the "best" applications and applicants are funded, and perhaps more importantly, it is also assumed that being funded by this instrument provides the best future career opportunities, such as visibility, impact, networking and mobility, compared to the postdocs funded indirectly through larger research projects. No evidence is provided for this presumption. Hence, the DFF commissioned an evaluation of the two postdoc funding instruments (DFF, 2012).

This study presents the results of the bibliometric part of this evaluation. Besides the bibliometric analysis of publication performance, the evaluation also constitutes a survey and two register-based analyses of career paths, mobility, internationalization, subsequent funding successes and leadership opportunities.

The aim of this study is to examine and compare the longer-term productivity and impact of researchers funded by these two instruments. Additionally, in order to measure the potential effect of being funded compared to not being funded, a carefully constructed control group is also set up and their productivity and impact is likewise examined and compared to the two postdoc-funding instruments. The research design therefore enables us to answer the research question regarding long-term performance differences between the two instruments, as well as the question regarding effectiveness of being postdoc-funded at all. Notice, while the research design is considered strong due to its strict matching process, we do not claim any causal effects for funding interventions (e.g., Freedman, 2005)

The paper is organized as follows, data and methods, especially the matching process, are presented in the next section; the subsequent section presents some of the most important results; and we end with a brief discussion of the approach, the results and some consequences.

Method and data

DFF comprises five field specific research councils (FSS: health sciences, FNU: natural sciences, FTP: technical sciences, FSS: social sciences, and FKK: arts and humanities). The evaluation examines postdocs funded between 2001 and 2009; in that period 62% of the funded postdocs went directly to individual applicants and the remaining 32% went indirectly to positions embedded in larger research projects. The bibliometric evaluation comprises all research councils, but only the evaluation of FSS, FNU and FTP comprise citation analyses. In this paper we only focus on postdoc-funding from the latter three research councils comprising citation analyses.

The comparisons between the two postdoc instruments and between the instruments and the control group are carried out combined for all three research councils on the aggregated level (e.g., sub-fields aggregated to the overall council). In order to be able to compare the bibliometric performance-effect of the two postdoc instruments, in between them and between them and a comparable group of non-funded researchers, a stringent matching process is set up to deal effectively with covariates. Notice, we have access to all postdoc grants from 2001 to 2009, including grant information, names and demographic data both for individual and embedded postdocs. We also have access to a general population of researchers in Denmark for the same period.

Matching procedures

Matching methods can be considered a stronger approach than simple statistical controls when designing an observational study in the sense of selecting the most appropriate data for reliable estimation of effects when it is impossible to have full control (e.g., randomization) (Rubin, 1997; Rosenbaum, 2002). Notice, most observational studies, especially ones with strict matching procedures, cannot fulfil the required assumptions needed for frequentist inferential statistics (e.g., Gill, 2010), this is also the case in the present study. However, we do not consider this a problem at all (see Schneider, 2013).

The control group was created from the general population of researchers in an exact match (1:1) with postdocs from both groups, where exact matching criteria included research council, active subject field and year of PhD-graduation. Further matching criteria include age, where a 2-year variation in age was allowed between the matched individuals. Finally, in order to produce roughly equal-sized treatment and control groups we randomly selected eligible matched individuals. Initially around 730 researchers were included (i.e., postdocs and control group).

Bibliometric data, measures and analyses

After the matching procedure, an attempt was made to construct publication portfolios for all 730 researchers by searching the Web of Science (WoS) database on name variants, affiliations and the specified publication period. Eventually, some 60 researchers were discarded as no portfolio could be established, either due to failure of identification or dubious identification. Con–sequently, 670 researchers were eventually selected for the overall analysis, i.e., control group (202), embedded postdoc (228) and individual postdoc (240). The 670 researchers is the basis for the publication analyses in the overall evaluation. In this paper we only focus on the results of the citation analyses and only 632 of the 670 researchers went into the citation analyses. The reduction in the number of authors is primarily caused by the shorter publication window and for a few because they had only published proceedings papers. On the other hand, the reduction in the number of researchers seem to be just about evenly distributed among the examined groups, i.e., control group (195), embedded postdocs (207) and individual postdocs (230), the corresponding drop in publications also do not seem to alter the publication profiles of three groups.

While the number of researchers in each group turned out to be marginally different, the homogeneity between groups on matching parameters and demographic variable remained, so we considered the groups representative and suitable for comparison of publication activity and impact. Notice, the publication portfolios for some of these researchers are most likely incomplete, including false positives and missing publications. Unfortunately, we were not able to contact researchers in order to validate the publication lists. We assume these 'errors' to be equally distributed across all three groups, furthermore since we only produce results on the group level, this type of error on individual level cancels out. We should stress that what we examine is publication performance of the three groups within the constraints of Web of Science.

Bibliometric analyses require robust data. A pilot study indicated that the publication history of most postdocs were meagre around the time they received their grants, many acquiring their grants within one or two years after finishing their PhD. As a consequence, a pre-test of bibliometric performance between the funded postdocs, as well as potential applicants is not feasible in the present case; indeed we conjecture that this is the case in many circumstances. Instead, we establish more robust publication portfolios for each individual researcher and measure the longer-term impact accordingly. Notice, this implies that potential publications not related to the postdoc-funding will be included in the portfolio and as a result analyzed.

Several publication windows for the portfolios were tested, but the general homogeneity between groups, including granting years between 2001 and 2011, meant that relative differences between groups were stable regardless of the length of the publication window. We therefore report results with publication windows spanning from the granting year (estimated year after PhD graduation for the control group) to 2009, i.e., 2009 is last suitable year for citation analysis. We apply a citation window of three years, the publication year plus two consecutive years. A relative short window like this enables us to include publications from 2009. The following publication types are included in the citation analyses: Articles, articles; proceedings paper[i] , letters and reviews. We apply CWTS' version of WoS, where letters are weighted as one fourth of paper. All citation analyses are carried out with field normalizations excluding self-citations. We present productivity measures, mean (MNCS) and percentile ($PP_{top\,x\%}$) citation indicators and self-citation rates (Waltman et al., 2012). The MNCS is calculated for each researchers portfolio of publications and an overall weighted MNCS is calculated for the entire group (i.e., number of publications is used for weighting). Focus is on differences in weighted MNCSs between the three groups and results are scrutinized with bootstrapped 95% confidence intervals (Efron, 1987; Colliander & Ahlgren, 2011) and effect sizes (Cohen, 1988). In this short paper we are not able to go into details of all the analyses carried out, hence the next section will only present some of the main results directly relating to the research questions asked in this paper, i.e., whether there is a substantial difference in the long-term performance between the two postdoc-funding instruments, as well as a substantial difference in long-term performance between being postdoc-funded or not.

Results

The overall publication analyses, including 670 researchers and a publication window from the granting year (estimated year after PhD graduation for the control group) to 2011, showed that individual postdocs' average and median publication activity is 17.5 and 13; embedded postdocs' is 13.4 and 9; and finally the average and median activity for the control group is 10.2 and 8. The average publication period for all groups is 6 years and follows a Gaussian distribution. We see that the publication activity distributions for the two postdoc-groups are considerably more skewed compared to the control group. The publication activity result seems to support the presumptions of the Board of Directors at DFF, however, unfortunately publication activity often plays a too important role in peer review and policy processes at the expense of impact

analyses. A reasonable bibliometric evaluation of the two funding instruments should be based on a comparison of the impact of the groups' publications.

As stated in the methods section, the citation analysis is restricted to publications published from the year of granting (estimated year after PhD for the control group) to 2009. Table 1 below shows that the differences in publication activity between the three groups in absolute numbers are intact in the citation analysis, but more important is the marginal differences in the average and median impact (MNCS) of these publications between the groups.

Table 1. Relative indicators for the three groups.

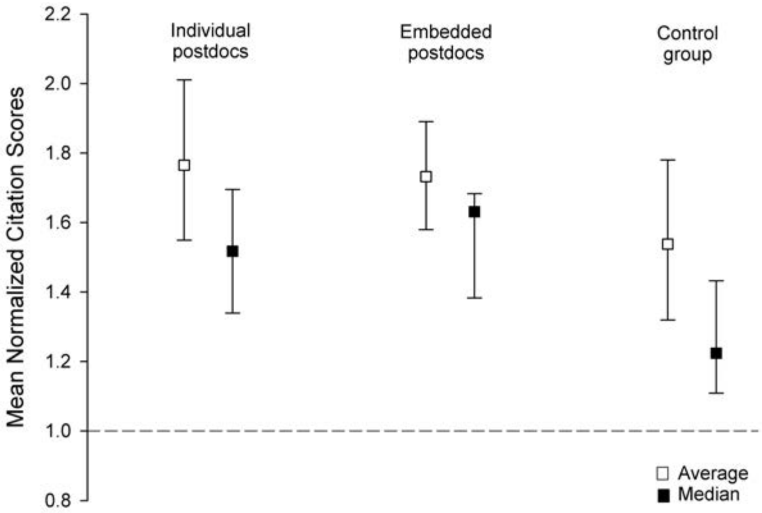
	Control group	Embedded postdocs	Individual postdocs
Total publications	1512.75	1947.75	2735.25
Self-citation rate	29%	29%	31%
Average MNCS*	1.54	1.73	1.76
Median MNCS**	1.22	1.63	1.52
$PP_{top10\%}$***	1.46	1.94	1.80
$PP_{top5\%}$***	1.61	2.13	2.19
$PP_{top1\%}$***	1.66	2.51	3.07

*Average weighted MNCS for all researchers in the group.
**Median weighted MNCS for all researchers in the group.
***Ratio of observed and expected publications among the top *x*% highly cited.

The performance of all three groups is well beyond the "database average" of one, but the publications coming from the two postdoc-instruments seem to have a larger impact on average, some 19 and 22 percentage points higher than the control group (average MNCS), whereas the performance of the two postdoc-instruments seems to be equal. In fact, if we turn to the median MNCS, we see that the embedded postdocs perform slightly better than the individual postdocs. The same pattern is visible with the percentile indicators. All three groups have more publications than expected among the most highly cited in the database, while the two postdoc-groups have larger odds-ratios of highly cited publications compared to the control group. But we also see that at the top 10% level, the embedded postdoc group have a marginally higher ratio of highly cited publications compared to the individual group; however this is reversed at the top 1% where the individual postdocs have a higher ratio. Notice the actual difference in ratios at the top 1% should not be overestimated as data is not robust at this level.

The impression we get from these indicators is that there is no difference between the two postdoc-groups, and a difference between the postdoc-groups and the control group. One should ask whether these results are robust and important. Consider Figure 1 below. Here we see a plot of the mean and median weighted MNCS indicators.

Figure 1. Differences in relative citation impact.



What is also presented in this plot is bootstrapped 95% confidence intervals (CI) surrounding the indicators. It gives no meaning to produce standard CIs as we do not have a probability sample. Instead, we can estimate the robustness of the data a hand by simulating a frequentist experimental situation, in this case with 1000 iterations (Efron, 1987). Several characteristics emerge from the CIs. All groups have rather large CIs for the weighted average MNCS and they all overlap with each other. To a certain extent this can be ascribed to the skewness of the underlying distributions and the strong dependence of highly cited papers in the calculation of averages. Nevertheless, the embedded postdoc-group seems to have the most robust weighted averages. The CIs for the median MNCS reveal asymmetric confidence limits. We notice that the embedded postdoc-group's median MNCS is slightly higher than the individual postdoc, testifying to a more stable underlying distribution in the inter quartile range. Notice again that all CIs overlap with each other. The bootstraped CIs reveal fragile underlying distributions and blurs the immediate impression that there exists a noteworthy difference between the postdoc-groups and the control group.

Further evidence for indistinctive results are provided by scrutinizing the differences by calculating standardized effect sizes for averages and medians presented in Table 2 below.

Table 2. Relative indicators for the three groups.

	Individual vs. embedded postdoc groups	**Embedded postdoc group vs. control group**
Hedge's g (unbiased) converted to r effect size	.01	.09
Effect size for medians	0.51	.57

We only examine the difference between individual and embedded postdocs, as well as the difference between embedded postdocs and the control group; a further comparison between individual postdocs and the control group is superfluous as the performance of the postdoc-groups is in essence the same. Effect sizes basically quantify how far we are from a null finding. We convert Hedge's *g* for differences in means to *r* for interpretive reasons, and if we compare the results of the two comparisons to Cohen's benchmarks (Cohen, 1988), we see that the differences can be characterized as "trivial". Since, the underlying distributions are skewed, we also calculate an effect size based on medians as suggest by Grissom and Kim (2012). This statistic estimates the probability that a score randomly drawn from population *a* will be greater than a score randomly drawn from population *b*. As we can see from Table 2, the probability for the two postdoc-groups is basically fifty-fifty; and with .57, the probability is only slightly in favour for embedded group compared to the control group. Consequently, the results give evidence to the claim that there is no difference in long-term citation performance between the two postdoc-funding instruments, but also inconclusive evidence for a substantial difference in long-term citation performance between being postdoc-funded or not.

Notice, it could be argued that the individual postdoc group outperforms the two other groups, and likewise that the two postdoc groups outperforms the control group, due to the considerable difference between the groups when it comes to the total number of fractionalized and field normalized citations:, 4814 (individual postdocs), 3369.6 (individual postdocs) and 2329.6 (control groups). Such an argument weights citations by publications thus emphasising the general higher average publication activity among postdocs[ii]. The argument is certainly valid, especially at the individual level when comparing relatively small *n* publication profiles, however we think the argument is more complex at the aggregate level where publication numbers are considerably larger, individual profiles much more skewed within the group, and the publication expectations between groups are unclear. In order to consider such an argument in more detail we need to do more specialized analyses that investigate differences in the distributions of author profiles between the three groups. We also need to consider that postdoc-funding is expected to yield higher publication output. We will do this in a future study.

Discussion

The bibliometric part of the overall evaluation played a dominant role in the official conclusion given by the Board of Directors at DFF. They conclude that the research councils are able to select very competent researchers for postdoc fellowships and that their performance is considerably better than the researchers in the control group, both in relation to publication activity and citation impact (DFF; 2012, p. 3). In other words, the two different postdoc-funding instruments function equally well, contrary to what was presumed, "and certainly much better that traditional tenure-track positions"! The present analysis demonstrates that this conclusion is somewhat overstated and perhaps politically motivated? As indicated above, publication activity is a poor parameter for performance evaluation in this context. Postdocs generally have more time for

research compared to tenure-track positions such as assistant professors. Consequently, it is expected that their activity is higher. When it comes to impact, the present analysis questions the robustness of the findings. First, what seems to be a "significant" difference between the performance of the postdoc-groups compared to the control group, turns out to be a "trivial" difference if we use Cohen's benchmark, or close to a fifty-fifty chance in probabilities. The general conclusion should be that all groups perform above the database average and that the postdoc-groups appear to perform slightly above the control group, but these differences are non-robust and most likely unimportant. Like Kreiman and Maunsell (2011), and others before them, we assert that bibliometric indicators, especially at micro- and meso levels should come with robustness and sensitivity analyses; but we also argue that some sort of yardstick should be provided, in order to make decisions concerning the importance of findings (e.g., Schneider, 2012). We think that this study provides some illustrations of how the robustness of indicators can be scrutinized. Finally, why should the conclusion by DFF be politically motivated? The current trend in Denmark is to terminate traditional tenure-track programs and only give support to postdoc-fellowships. No official reasons are given, but the fact that postdocs are financed by external funding is probably one of them.

References

Bornmann, L. & Daniel, H.-D. (2006). Selecting scientific excellence through committee peer review: a citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. *Scientometrics*, 68(3), 427–440.

Bornmann, L., Wallon, G., & Ledin, A. (2008). Does the committee peer review select the best applicants for funding? An investigation of the selection process for two European Molecular Biology Organization programs. *PLOS One*, 3(10): e3480.

Bornmann, L., Leydesdorff, L., & van den Besselaar, P. (2010). A meta-evaluation of scientific research proposals: Different ways of comparing rejected to awarded applications. *Journal of Informetrics*, 4(3), 211–220.

Böhmer, S., Hornbostel, S., & Meuser, M. (2008). Postdocs in Deutschland: Evaluation des Emmny Noether-Programs. *IFQ-Working Paper No. 3*, Bonn. http://www.forschungsinfo.de/Publikationen/Download/working_paper_3_2008.pdf [accessed 13.03–2013]

Colliander, C., & Ahlgren, P. (2011). The effects and their stability of field normalization baseline on relative performance with respect to citation impact: A case study of 20 natural science departments. *Journal of Informetrics*, 5(1), 101–113.

Efron, B. (1987). Better bootstrap con?dence intervals (with Discussion). *Journal of the American Statistical Association*, 82, 171–200.

Evaluering af postdocfinansiering i Det Frie Forskningsråd [*Evaluation of postdoctoral funding within the Free Research Council*], rapport udgivet af Styrelsen for Forskning og Innovation (2012): http://fivu.dk/publikationer/2012/evaluering-af-postdocfinansiering-i-det-frie-forskningsrad.

Freedman, D.A. (2005). *Linear statistical models for causation: A critical review*. In: Everitt, B. & Howell, D., eds., *Encyclopedia of Statistcis in Behavioral Science*, New York: Wiley, 1061–1073.

Gill, J. (2010). Critical differences in Bayesian and non-Bayesian inference. In: Statistics in the social sciences : current methodological developments, Kolenikov, S. ed., Wiley: Hoboken, N.J., 135–158.

Hornbostel, S., Böhmer, S., Klingsporn, B., Neufeld, J., & von Ins, M. (2009). Funding of young scientist and scientific excellence. *Scientometrics*, 79(1), 171–190.

Jacob, B. A., & Lefgren, L. (2011). The impact of NIH postdoctoral training grants on scientific productivity. *Research Policy*, 40(6), 864–874.

Kreiman, G., & Maunsell, J.H.R. (2011). Nine criteria for a measure of scientific output. *Frontiers in Computational Neuroscience*, 5(48): doi: 10.3389/fncom.2011.00048.

Neufeld, J. & von Ins, M. (2011). Informed peer review and uninformed bibliometrics? *Research Evaluation*, 20(1), 31–46.

Neufeld, J. & Hornbostel, S. (2012). Funding programs for young scientists – do the 'best' apply?. *Research Evaluation*, 21(4), 270–279.

Rosenbaum, P. (2002). Observational studies. 2nd edition. Springer-Verlag: New York.

Rubin, D. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, 127, 757–763.

Schneider, J.W. (2012). Testing University Rankings Statistically: Why this perhaps is not such a good idea after all. Some reflections on statistical power, effect Size, random sampling and imaginary populations. In: Archambault, É., Gingras, Y., & Larivière , V., eds., *Proceedings of 17th International Conference on Science and Technology Indicators*, 719–732.

Schneider, J.W. (2013). Caveats for using statistical significance tests in research assessments. *Journal of Informetrics*, 7(1), 50–62.

Waltman, L, Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., van Eck, N. J., an Leeuwen, T.N., van Raan, A. F. J., Visser, M. S., Wouters, P. (2012) The Leiden ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, 63(12): 2419–2432.

i Due to a change in the classification of articles in WoS, after the inclusion of the proceedings citation index, some regular articles now carry a double document type classification, i.e., articles; proceedings papers.

ii We thank one of our reviewers for bringing up this point.

Internationalization of Large Firms R&D activities: towards a shift in trend?

Antoine Schoen — a.schoen@esiee.fr / Université Paris-Est, ESIEE – LATTS – IFRIS, 2, bd Blaise Pascal, 93160 Noisy Le Grand (France)

Patricia Laurens — p.laurens@esiee.fr / Université Paris-Est, CNRS – LATTS – IFRIS, 2, bd Blaise Pascal, 93160 Noisy Le Grand (France)

Christian Le Bas — christian.lebas@univ-lyon2.fr / Université Lyon 2, GATE, 93 Chemin des Mouilles, 69130 Ecully (France)

Lionel Villard — l.villard@esiee.fr / Université Paris-Est, ESIEE – LATTS – IFRIS, 2, bd Blaise Pascal, 93160 Noisy Le Grand (France)

Philippe Larédo — philippe.laredo@enpc.fr / Univ. Paris-Est, ENPC – LATTS – IFRIS, 5, Bd Descartes, 77454 Marne-La-Vallée Cedex 02 (France)

Introduction

This research investigates empirically the analytical framework initially developed by Kuemmerle (1997, 1999) where internationalization of technology within MNCs is either motivated by "home base exploiting" (HBE) or "home base augmenting" (HBA) factors. In the former case, the company's goal is to exploit already developed assets, delivering inventions that are mostly of the adaptive type; in the latter one the company aims at obtaining abroad strategic assets that are complementary with those already available at home.

Patel and Vega (1999) have established that a comparison of the technological advantage of the company at home and the advantage of the location shows that in a large majority of cases, firms tend to locate their technology abroad in their core areas where they are strong at home. Le Bas and Sierra (2002) have confirmed these findings: in a large majority of cases, companies locate their activities abroad in technological areas or fields where they are strong at home. Moreover these two studies show that HBA strategy dominates HBE.

Research questions

The main objective of our paper is to update these results ten years later. We will study two related research questions, the first dealing with the extent of internationalisation of corporate inventions; the second with the pattern of this internationalisation. In the first place, we will assess if corporate globalisation does apply to R&D matter (Guellec 2001) as heavily as it does for finance, sales and manufacturing. Second, we will assess to what extent, the "HBA" strategy still represents in the 2000's, as it was in the 1980's and 1990's, the main motivation for internationalization of invention within MNCs. We will more specifically investigate if the

rise of new actors – especially China – results in a shift for the choice of foreign locations for inventions.

Methodology

We use an original database that includes the largest industrial firms (the Corporate Invention Board (CIB)) that identifies the consolidated patent portfolios for the 2000 world largest R&D investors in 2008, Schoen et al. 2013), our study computes at country level:

- the technological specialization of corporate domestic inventions (i.e. patents applied for by CIB firms that include an inventor address in the corporation's home country – as defined by headquarters' location);
- the technological specialization of countries where corporate non-domestic inventions were produced (i.e. patents applied for by CIB firms that include an inventor address outside the corporation's home country).

Our study – as did by Patel and Vega (1999) – uses the (priority) patents IPC codes and inventors' addresses for computing countries' technological specialization – taking as a reference all inventions made across the world. Then we calculate Revealed Technological Advantages (RTA) reflecting the countries' technological profiles.

Results

Internationalisation rate

The first set of results shows that globally the whole set of corporations from the CIB have only slightly increased their internationalisation rate, which has gone up from 5,2% for the 1994 to 1996 period to 7,2% for the 2003 to 2005 period. This modest global increase encompasses actually different trends – including a decrease of internationalisation – depending on the population of firms taken into account.

In the first place, due to the large share of Japan in the whole stock of world inventions (60%) and the peculiarities of internationalisation pattern of Japanese firms, excluding this country from the analysis produces a radically different picture: (non Japanese) corporations have globally slightly decreased their internationalisation rate. The share of their non-domestic inventions went down from 18,3% in the 1994 to 1996 period to 16,4% for the 2003 to 2005 period. Thus, we don't witness any major pattern of R&D globalisation or R&D off shoring within the largest industrial corporations.

In the second place this slight decrease in internationalisation is mostly due to corporations from large European countries, which have reduced their internationalisation rates, whereas Asiatic

countries and the US have increased theirs. The table below presents this diversity of trajectories.

Table 1. Evolution of internationalisation rate in corporate inventions over 10 years

	94_96	03_05	Change of %	Patent nb 03_05
World	5,2%	7,2%	39,4%	851985
World without Japan	18,3%	16,4%	-10,4%	333348
JP	0,6%	1,3%	130,9%	518637
KR	1,5%	6,3%	319,8%	119292
US	9,0%	16,6%	84,7%	101382
DE	15,3%	13,1%	-14,2%	51935
FR	48,2%	33,8%	-29,9%	14394
CN	0,0%	0,4%	NA	11314
TW	10,6%	27,0%	154,2%	8666
GB	88,6%	80,0%	-9,7%	4232

Internationalisation strategies

The second set of results (table 2) shows a slight change in the corporate strategies with an increase of the HBE strategies whose share progresses from 35,8% for the 1994 to1996 period to 39,2% to 16,4% for the 2003 to 2005 period.

Table 2. Evolution of corporate internationalisation strategies over 10 years

	HBA	HBE	Market Seeking	Techno Seeking
1994–1996	43,3%	35,8%	9,3%	11,6%
2003–2005	43,3%	39,2%	8,0%	9,5%

The role of China as an emerging place for invention for foreign (i.e. non Chinese) corporations could explain the evidenced increase of strategy towards HBE. China has become during the 2003 to 2005 period the third most important location (behind the US and Germany) for corporate international inventions; when it was hardly visible 10 years before. This emerging giant feeds mainly (for 53,6%) the HBE strategy; HBE strategy-oriented inventions stemming out China weight more than one fifth (22,2%) of all inventions belonging to this type of strategy (with 27,7% for the US and 15,7% for Germany).

Consistency check

The quality of the CIB data set has been checked against the one used by Le Bas and Sierra(2002). This comparison has produced satisfying results. For the 170 institutions (out of a total of 350) from Le Bas and Sierra's corpus that belong as well to the CIB corpus, very similar internationalisation rates have been compiled for the 1994 to1996 period. Moreover, the distribution of internationalisation strategies compiled for the 1994 to 1996 period with both datasets produces very similar results, as displayed in the table below.

Table 3. Comparison of internationalisation strategies between Le Bas et al. (2002) vs CIB

	HBA.	HBE	Market Seeking	Techno Seeking
Le Bas et Sierra (2002)	47,4 %	30,1 %	9,5 %	13,1 %
CIB	43 %	36 %	9 %	12 %

References

de Rassenfosse, G., Dernis, H., Guellec, D., Picci, L., van Pottelsberghe de la Potterie, B. (2013), The worldwide count of priority patents: A new indicator of inventive activity. *Research Policy*, 42 (2013), 704–719.

Kuemmerle, W. (1999). Foreign direct investment in industrial research in the pharmaceutical and electronics industries–results from a survey of multinational firms. *Research Policy*, 28(2–3), 179–193.

Le Bas, C. & Sierra, C. (2002). Location versus home advantages' in R&D activities: Some further results on multinationals' location strategies. *Research Policy*, 31 (4), 589–609.

Patel, P. & Vega, M. (1999). Patterns of internationalisation of corporate technology: location versus home country advantages. *Research Policy*, 28 (2/3), 145–155.

Schoen, A., Heimeriks G., Villard, L., Laurens, P., Alkemade, F., Larédo, P. (2013). The internationalization of corporate invention: national and sectoral contributions. Paper submitted to the *EU SPRI 2013* conference.

The predictability of the Hirsch index evolution

Michael Schreiber schreiber@physik.tu-chemnitz.de / Institute of Physics, Chemnitz University of Technology, 09107 Chemnitz (Germany)

Abstract

The h-index can be used as a predictor of itself. However, the evolution of the h-index with time is shown in the present investigation to be dominated for several years by citations to previous publications rather than by new scientific achievements. This inert behaviour of the h-index raises questions, whether the h-index can be used profitably in academic appointment processes or for the allocation of research resources.

Introduction

Due to its simplicity the so-called Hirsch index or h-index has become attractive as a frequently used metric for describing the scientific achievements of a researcher. It was introduced by Hirsch (2005) as the largest number h of publications of a scientist which have received at least h citations each. This means $h = \max\{r \mid c(r) \geq r\}$. Here c($r$) denotes the number of citations to the paper at rank r, after the papers have been sorted according to decreasing $c(r)$.

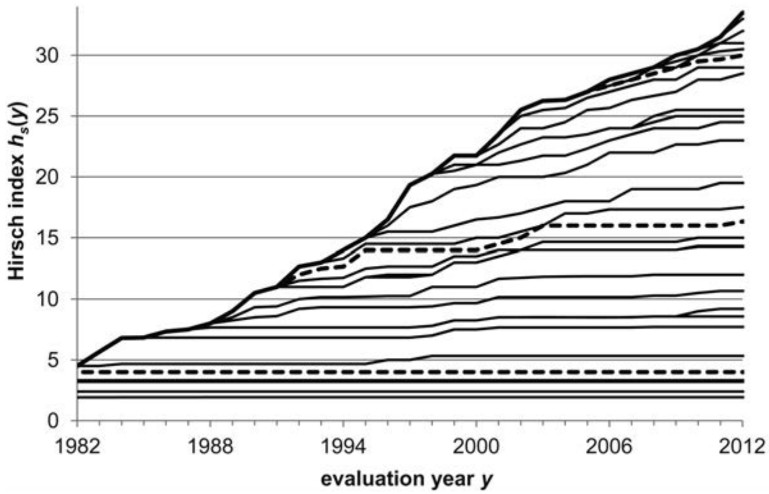
In spite of several weaknesses and in spite of doubts how representative this measure is, the h-index has become popular and is often used in academic appointment processes and evaluation procedures of research projects. Hirsch (2007) determined a high correlation comparing the h-index values of researchers after 12 years and after 24 years of their careers. He concluded that "the h-index has the highest abibility to predict future scientific achievement".

Utilizing a complicated fit with 18 parameters, Acuna, Allesina, and Kording (2012) were able to predict the future h-index rather accurately for several years. It is controversial (Rousseau and Hu, 2012) whether such an approach is meaningful. In a validation study Garcia-Perez (2013) showed that the predicted h-index trajectories overestimate the future h indices more and more for later and later target years. I have recently demonstrated (Schreiber, 2013) that the increase of the h-index with time after a given point of time (e.g., the appointment year or the evaluation year) is for several years not related very much to the scientific achievements after that date. Rather the growth of the h-index is nearly the same for several years, irrespective of whether further work had been published or not after that date.

In my previous publication I have presented examples for the rather smooth increase of the h-index with time thus confirming that the h-index is a good predictor of itself. However, I have

also presented evidence in 4 examples that the growth of the h-index does not depend very much on the factual performance for several years in the future but rather results mostly from previous, often rather old publications. As evidence I had selected the most impressive example years so that the deviations were small for a particularly long time interval and not representative. It is the purpose of the present investigation to analyze quantitatively the duration for which the h-index remains unchanged or only slightly changed.

Figure 1. The Hirsch index $h = h_y(y)$ *for the publication record of the present author (top line). The dependence of* $h_s(y)$ *is shown for* $s = 1976, 1977, \ldots$ *(from bottom to top). For the years* $s = 1980, 1990$, *and 2000 thick broken lines are utilized.* y *is the year of evaluation,* s *is the last year from which publications are taken into consideration.*



The citation data and the calculation of the h-index

I harvested my own citation record from the ISI Web of Science database in March 2013. I determined the citations up to a given year y and counted the publications with high citation frequencies selectively, namely considering only publications up to a certain year $s \leq y$. This yields the selective h-index $h_s(y)$ for the year y, under the assumption that I had stopped publishing in year s. Of course, if $s = y$ then the usual h-index $h = h_y(y)$ is obtained. As h and likewise $h_s(y)$ are restricted to integer values, a graphical representation of $h_s(y)$ curves is very difficult to survey, because many values coincide. Therefore I had restricted by previous investigation (Schreiber, 2013) to selective values of s and discussed the resulting curves qualitatively.

I have repeated the analysis now with the updated data for the interpolated version of the h-index, which is obtained after a piecewise linear interpolation of the citation distribution $c(r)$ between r

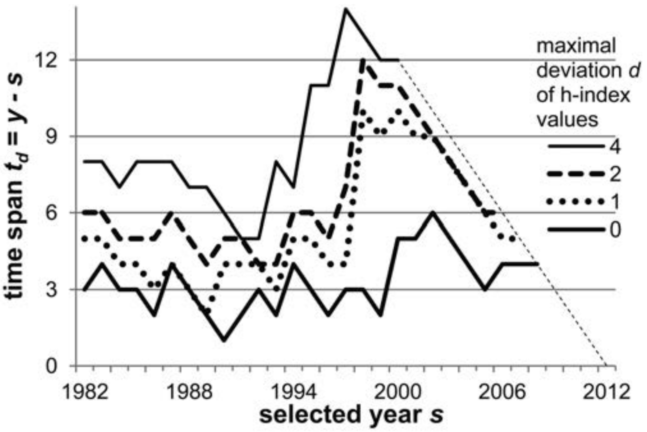
and $r + 1$, as suggested and utilized previously by Rousseau (2006), van Eck and Waltman (2008) and myself (Schreiber, 2008, 2009). The interpolated h-index then results from the solution of $c(h) = h$. By truncating the interpolated index values one obtains the usual integer results.

In Fig. 1 the determined $h_s(y)$ curves are presented, showing the expected rather smooth behaviour. The original h-index $h = h_y(y)$ shows a steady increase with a slope of approximately one index point per year. Selecting a particular year s, one can see that for a duration of several years $t = y - s$ the deviation $d = h_y(y) - h_s(y)$ remains rather small, indicating that the growth of the h-index after the year s is mostly due to the publications until the year s. In fact, for all reasonable values of s (values $s > y$ are not meaningful) the initial part of the curves $h_s(y)$ in Fig. 1 cannot be observed, because the values coincide with $h_y(y)$. Of course, the $h_s(y)$ curves level off for recent years, because the very old publications are either highly cited and belong already to the h-defining set of papers for several years, or they are so lowly cited that they have small chance to become relevant for the h-index. Nevertheless, there are some exceptions from this argumentation which lead to the prominent increases that some of the curves feature in Fig. 1. Furthermore, there is always the possibility of so-called sleeping beauties which have not received a significant number of citations for a long time, but then suddenly are cited frequently. This has happened to some of my papers about quasicrystals, because the subject which was a hot topic in the nineties has become topical again after the Nobel prize in chemistry was awarded to D. Shechtman in 2011.

The inert behaviour of the h-index

In order to obtain a more quantitative description of the observed behaviour, I have determined the time span t_0 for which the h-index does not differ, irrespective of whether I had performed as I did or whether I had not published any further work after the selected year s. The results are shown in Fig. 2. The bottom line indicates that for most values of s the h-index would not have changed for the next 3 years. Between 1982 and 2008 this inert behaviour is observed on average for 3.3 years, i.e. $\overline{t_0}\,0 = 3.3$, where the overbar denotes the average. I have not included more recent years, because for $s = 2008$ I get $t_0 = 4$, so that $y = s + t_0 = 2012$ equals already the last year covered by the dataset. Therefore for more recent years s the restriction $t_0 = 2012 - s$ effectively limits t_0 and it is quite likely that t_0 will grow further in the future, i.e. for larger values of y. In principle this could even happen already for $s = 2008$. I have also excluded the first six years from the average in order to avoid possible problems with a transient behaviour of the index at the beginning of my career. However, including this initial period, the average duration of coinciding $h_y(y)$ and $h_s(y)$ values would change only slightly to $\overline{t_0} = 2.9$.

Figure 2. Time span $t_d = y - s$ *for which the h-index* $h_s(y)$ *would have remained the same (*$d = 0$*) or deviated at most by* $d = 1, 2,$ *or 4 (from bottom to top) index points from the factual index values* $h_y(y)$*, if I had stopped publishing in the selected year* s*. The thin broken line indicates the border* $t_d = 2012 - s$*, which limits the curves for* $y = 2012$*.*



It is well known, that the accuracy of the h-index does not allow a meaningful distinction between researchers with nearly the same index values. In other words, small differences are not meaningful. Therefore I have also calculated and included in Fig. 2 the duration times t_d for which my values of $h_y(y)$ and $h_s(y)$ differ by not more than $d = 1, 2,$ or 4 index points. It is not surprising that the obtained values of $t_d(s)$ are considerably larger than $t_0(s)$. On average I obtained $\overline{t_1} = 5.4$, $\overline{t_2} = 6.6$, and $\overline{t_4} = 8.7$, excluding again values for $s \leq 1981$ and for $s \geq 2008, 2007, 2001$, respectively, because the border line $y = s + t_d = 2012$ was already reached. Again, including the early years does not change the averages much, in this case I determined $\overline{t_1} = 4.8$, $\overline{t_2} = 5.9$, $\overline{t_4} = 8.3$.

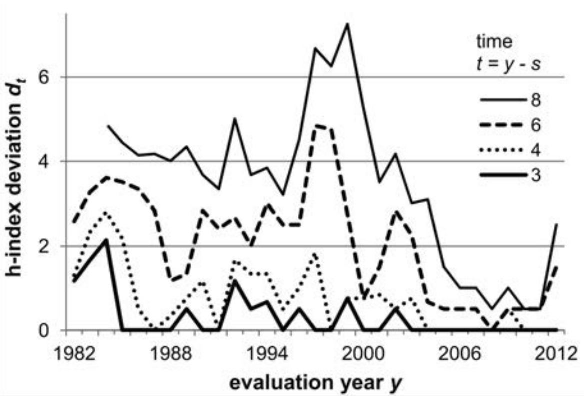
The decrease of the curves in Fig. 2 around 1990 indicates that rather recent publications influence the h-index evolution. The prominent increases after 1994 show that now the h-index is dominated by past scientific achievements. Of course with increasing values of h it is more and more difficult for new publications to obtain the number of citations necessary for becoming relevant for the h-index. The distinct drop of the curves on the right-hand side of Fig. 2 shows that now rather recent publications have had an effect. In fact, a closer investigation of my citation record revealed that this is due to my new hobbyhorse: My investigations on the h-index and other topics in Scientometrics have quickly made it into the set of h-defining publications.

The predictability of the h-index

Another possibility to describe the inert behaviour or the predictability of the h-index in a quantitative way is to consider a certain year y and to determine for different time spans $t = y - s$ the

deviation $d_t = h_y(y) - h_s(y)$, which is caused by excluding all papers after the selected year *s* from the evaluation of the h-index. Respective results are presented in Fig. 3. It can be seen that for a time span of $t = 3$ years the deviation is small, often there is no difference at all

Figure 3. Difference $d_t = h_y(y) - h_s(y)$ *for* $t = 3, 4, 6, 8$ *(from bottom to top), if the last* $t = y - s$ *years are excluded from the determination of* h *in the year* y*, i.e., if I had stopped publishing in the selected year* $s = y - t$*. For* $t = 8$ *the first two data points are missing, because my publication record starts in the year 1976 and these missing data points correspond to* $s = 1974$ *and 1975.*



as already observed in Fig. 2. The average deviation since $y = 1982$ is $\overline{d_3} = 0.31$. Excluding the publications for the last $t = 4$ years, somewhat larger deviations can be found but they are still small, on average $\overline{d_4} = 0.75$. Even for $t = 6$ years, the average difference is only $\overline{d_6} = 2.1$, which most people would probably consider as not being relevant. For $t = 8$ years of unproductivity, the average influence on the h-index is $\overline{d_8} = 3.5$.

For individual years there are some outliers in the curves presented in Fig. 3. The rather large deviations in the late nineties correspond to the low values of t_d in Fig. 2 around 1990 and reflect the high impact of several publications around 1990 which have quickly made an effect on my h-index and then dominated its evolution in the late nineties. Likewise the small values of d_t before and around 2010 correspond to the high values of t_d in the late nineties in Fig. 2, and can be likewise explained by rather old publications dominating the h-index evolution.

For a comparison between Figs. 2 and 3 it should be noted that the curves in Fig. 2 are plotted in dependence on *s*, while the curves in Fig. 3 are presented as a function of *y*. Thus there is a tilt of $t = y - s$ years between corresponding features in these two figures to take into consideration.

Discussion

It will be interesting to see whether other citations records show a similar behaviour, not only the overall inert behaviour, but also whether the predictability can be likewise attributed predominantly to relatively old publications. Therefore further studies of the persistence of the h-index values in terms of the duration *t* should be performed. It would also be interesting to see whether such features as discussed above, namely distinct maxima or minima and prominent increases or decreases of the curves as in Figs. 2 and 3 can be found and attributed to specific details of other citation records.

In spite of the discussed inert behaviour the obtained results corroborate the predictive power of the h-index (Hirsch, 2007). It is tempting to assume that a predicted growth of the h-index can be correlated with the future performance of a candidate. This would make the h-index a possibly useful measure in academic appointment processes and for the allocation of research resources. However, as the present investigation has shown, the future development of the h-index is dominated for several years by citations to previous publications. This means that a high h-index value of a candidate can be expected to increase after this person is hired, even if he or she goes to sleep after the appointment and does not publish any further work. On the other hand, the past evolution of the h-index does not automatically mean that the candidate has performed good work in recent years. The h-index would most likely have grown more or less as it did, even if the candidate had gone to sleep several years ago. In conclusion, the present investigation raises doubts about the usefulness of the h-index for predicting future scientific achievements.

On the other hand, the observed inert behaviour of the h-index bears testimony of the significance of a researcher's past achievements. This is certainly an aspect which should be taken into account also in appointment processes or for the purpose of allocating resources.

References

Acuna, D.E., Allesina, S., & Kording, K.P. (2012). Predicting scientific success. *Nature*, 489, 201–202.

García-Pérez, M.A. (2013). Limited validity of equations to predict the future h index. *Scientometrics*, in print, DOI 10.1007/s11192-013-0979-7.

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569–16572.

Hirsch, J.E. (2007). Does the h-index have predictive power? *Proceedings of the National Academy of Sciences*, 104(49), 19193–19198.

Rousseau, R. (2006). Simple models and the corresponding h- and g-index. http://eprints.rclis.org/archive/00006153

Rousseau, R. & Hu, X. (2012). Predicting the h-index. *ISSI Newsletter*, 8(3), 53–55.

Schreiber, M. (2008). An empirical investigation of the g-index for 26 physicists in comparison with the h-index, the A-index, and the R-index. *Journal of the American Society for Information Science and Technology*, 59(9), 1513–1522.

Schreiber, M. (2009). The influence of self-citation corrections and the fractionalised counting of multi-authored manuscripts on the Hirsch index. *Annalen der Physik* (Berlin) 18, 607–621.

Schreiber, M. (2013). How relevant is the predictive power of the h-index? As case study of the time-dependent Hirsch index. *Journal of Informetrics*, 7, 325–329.

Van Eck, N.J. & Waltman, L. (2008). Generalizing the h- and g-indices. *Journal of Informetrics*, 2(4), 263–271.

Bibliometric Evaluation Standards Debate: Introductory Presentation and Panel Discussion

Presentation:

Daniel Sirtes sirtes@forschungsinfo.de / iFQ Institute for Research Information and Quality Assurance, Berlin (Germany)

Ludo Waltman waltmanlr@cwts.leidenuniv.nl / Centre for Science and Technology Studies, Leiden University (The Netherlands)

Additional Panel Participants:

Éric Archambault Eric.archambault@science-metrix.com / Science-Metrix Inc., Observatoire des Sciences et des Technologies (OST), Montréal, Québec (Canada)

Wolfgang Glänzel Wolfgang.Glanzel@kuleuven.be / Centre for R&D Monitoring and Dept. MSI, KU Leuven (Belgium)

Stefan Hornbostel hornbostel@forschungsinfo.de / iFQ Institute for Research Information and Quality Assurance, Berlin (Germany)

Paul Wouters p.f.wouters@cwts.leidenuniv.nl / Centre for Science and Technology Studies, Leiden University (The Netherlands)

Introduction

The role of bibliometrics in research evaluation is becoming increasingly important, and bibliometric data and statistics are becoming more and more widely available. At the same time, a general consensus on how bibliometrics should be applied in research evaluation contexts and which types of bibliometric indicators should be used has not yet emerged. There is no set of explicitly formulated and commonly accepted 'good practices' for the use of bibliometrics in research evaluation, and instead of trying to reach consensus on a standard set of bibliometric indicators that work well for most evaluation purposes, many bibliometricians seem to prefer spending their time on inventing yet another new indicator.

Perhaps the current state of bibliometric research, in which a lot of attention is paid to technical innovations and in which there is less interest in reaching a broad consensus on what does and does not work well, is just a normal and healthy situation for a research field in a stage of quick development and expansion. At the same time, however, there is a serious risk that for many end users of bibliometric analyses and evaluations these controversial discussions are rather unsettling and alienating. University presidents and deans, their support staff, university librarians, research managers at funding agencies, etc. usually do not have the time or the right background in order to keep up with all the developments going on in the field of bibliometrics. In many cases, these people simply expect bibliometricians to tell them how bibliometrics can best be applied to their particular situation, which types of indicators should be used, and so on. All the more, these people expect that bibliometricians tell them a more or less consistent story. Why would one trust bibliometrics as a research evaluation tool if each bibliometrician advices to use

a different set of indicators, possibly leading to quite different outcomes in a research evaluation exercise?

From a practical research evaluation point of view, there appears to be a clear need for an increased level of standardization in bibliometrics. However, standardization could take place in a number of different ways, and some forms of standardization are perhaps more important or more relevant than others. For this reason, we believe that it is useful to make a distinction between three types of bibliometric standardization:

- Standardization of data sources
- Standardization of indicators
- Standardization of good practices

Below, we discuss each of these three types of standardization in more detail.

Standardization of data sources

A bibliometric analysis requires a bibliographic data source. For analyses that are limited in scope, discipline-specific data sources may be used (e.g., PsycINFO), but for large-scale analyses one usually needs a multidisciplinary data source. The most popular multidisciplinary data sources are Thomson Reuters' Web of Science, Elsevier's Scopus, and Google Scholar. However, only the first two are documented and stable enough in their methodologies, so that reproducible results are warranted.

Bibliographic data sources differ from each other in the literature they cover, and therefore analyses based on different data sources normally do not yield the same results. However, even if two analyses are done based on the same nominal data source, they do not necessarily produce identical outcomes. Web of Science, for instance, consists of a number of databases (Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Conference Proceedings Citation Index, Book Citation Index), and depending on which of these databases one includes in an analysis, different results will be obtained. Furthermore, the databases change (even concerning older entries) on a weekly basis. Thus, exactly the same citation scores between studies will be hardly attainable.

In addition, professional bibliometric centers often do not work with the original 'raw' data from a data source such as Web of Science or Scopus, but instead they attempt to enhance the data quality. For instance, they may perform their own citation matching (i.e., the linking of reference strings to the publications being referenced) and they may work with cleaned address data (i.e., address data in which the names of organizations, cities, countries, etc. have been made consistent as much as possible). Obviously, the way in which citation matching, address cleaning, and other related issues are handled affects the outcomes of a bibliometric analysis.

Exclusion and inclusion of parts of the database for different analyses are very common as well. Publication types like serials and proceedings books are sometimes included in publication counts but excluded in citation scores. Typically, not all document types are included in studies. Mostly, only articles, reviews, and letters, the so called 'citable document types', are part of the studies. However, even on these matters there is no consensus.

Another issue is the way in which scientific fields or disciplines are defined. Bibliographic data sources typically offer a classification system that assigns publications to fields. Different data sources provide different classification systems, but even when working with the same data source it is possible to use different classification systems. Some bibliometric centers have for instance developed their own classification systems. It is clear that the use of different classification systems leads to different results in a bibliometric analysis.

Standardization of bibliographic data sources would require agreement to be reached on how to deal with each of the above issues. It is unlikely (and for certain purposes disadvantageous) that all centers will agree to use always the same data source with all the same demarcation criteria. However, guidelines and transparency would constitute important progress..

Standardization of indicators

Especially since the introduction of the *h*-index in 2005, bibliometricians have paid a lot of attention to the development of new indicators, resulting in a large literature in which lots of proposals are reported for new approaches to counting publications and citations. Many indicators have been introduced for evaluating individual researchers, for assessing research institutions and research groups, and for measuring journal impact. A large variety of issues have been studied, ranging from the problem of how to deal with co-authored publications to the problem of weighting citations differently depending on their origin (e.g., PageRank) to the problem of normalizing indicators to correct for field-specific publication and citation practices.

A first step toward standardization may be to reach agreement on a minimum set of core principles that indicators should follow. These principles could for instance relate to the mathematical and statistical properties of indicators (e.g., averages vs. percentiles), the way in which different document types are dealt with and citation windows are chosen, the approach taken to correct for field differences (e.g., cited-side vs. citing-side normalization), and the way in which statistical uncertainty is handled. Reaching consensus on at least the core principles of the design of indicators may be a significant step forward.

Standardization of good practices

In addition to the more technical aspects of bibliometric research evaluation discussed in the previous two sections, standardization may also take place in terms of the development of a commonly accepted set of good practices for the use of bibliometrics in research evaluation. These good practices may for instance relate to:

– Situations in which the use of bibliometrics is considered appropriate and situations in which it is not, with special interest to for instance the use of bibliometrics at the level of individual researchers and the use of bibliometrics in the social sciences and humanities.
– The role bibliometrics should play relative to peer review and other methods of research evaluation.
– The transparency of bibliometric research evaluation..
– The way in which manipulation of bibliometric statistics (e.g., citation cartels, journal self-citations) is handled.
– The degree to which the producers of bibliometric statistics (e.g., professional bibliometric centers) have a responsibility to properly inform and educate the end users of these statistics (e.g., university managers).
– More generally, the way in which the responsibility for the use of bibliometrics in research evaluation is shared between the producers of bibliometric statistics on the one hand and the decision makers that use the statistics on the other hand.

Like in the case of indicator standardization, it may not be feasible to reach consensus on good practices for all aspects of the use of bibliometrics in research evaluation. However, standardization of good practices could start in those areas that are considered most important and in which there is most agreement. At some point, one could imagine that this will result in a commonly agreed code of conduct for the producers and the users of bibliometric statistics.

Issues for discussion

We propose that the following topics will be raised by the panel discussants:

– What are the advantages and disadvantages of bibliometric standardization?
– Given the advantages and disadvantages, should bibliometric standardization be considered as something desirable? If so, which types of standardization are most important, and which are less important?
– To what degree will bibliometric standardization be feasible in practice? Could it still be too early to achieve a high level of standardization?
– How could the process toward bibliometric standardization be organized? What should be the role of bibliometric research centers, and what role should be played by the major commercial parties (i.e., Thomson Reuters and Elsevier)? Which other parties should be involved?

- Do we need a professional code of conduct for providers of bibliometric analyses? If so, how should this code of conduct be designed and implemented?
- What role may recent developments such as the increasing interest in altmetrics and the recently published San Francisco Declaration on Research Assessment play in the process?

The notion of talent: What are the talents we are looking for in science?

Pleun van Arensbergen p.vanarensbergen@rathenau.nl / Science System Assessment, Rathenau Institute, Den Haag (the Netherlands)

Inge van der Weijden i.c.m.van.der.weijden@cwts.leidenuniv.nl / Centre for Science and Technology studies CWTS, Leiden University, Leiden (the Netherlands)

Peter van den Besselaar p.a.a.vanden.besselaar@vu.nl / Network Institute & Department of Organization Science, VU University, Amsterdam (the Netherlands)

Abstract

In this paper we study the process of scientific talent evaluation in more detail. By analysing the grant allocation process on both the individual and panel level, we will give more insight in the predominant notion of talent in external talent selection. We will compare this to the notion of talent that prevails when researchers identify talent in their own academic work environment. We interviewed 29 panel members involved in grant allocation. For improving the transparency, quality and legitimacy of grant allocation practices, we furthermore aim to uncover the details of the de facto (implicit and explicit) criteria applied in grant allocation.

Introduction

The quality of education and research is strongly connected to the quality of the people working in the academic sector. For excellent science, excellent scientists are needed. Therefore government and universities specifically aim at investing in the best people when investing public resources in education and research. This has led to an increased focus on talent and talent policy. In the staff policy of universities in the Netherlands the notion of talent has nowadays taken a central position. Several programmes and policy initiatives are currently implemented to attract and stimulate academic talent. However, due to restricted university resources, researchers depend to an increasing extent on external funding. Acquiring external funding forms a prominent criterion in academic recruitment and evaluation processes.

In the Netherlands personal career grants are strongly considered as necessary resource in order to be able to have an academic career. This leads to huge application pressure for this type of personal grants and – because of a fixed budget – to a relatively low success rate. The importance for researchers to acquire these grants combined with the low success rate, asks for clarity and transparency of how the funding decisions are made.

A recent quantitative study of talent selection through the allocation of career grants showed that the evaluation and selection of talent is highly contextual (Van Arensbergen & Van den Besselaar, 2012). The extent to which an applicant was considered to be talented varied during the selection process and there was moderate agreement between the reviewers. Unlike the face-to-face panel interviews with the applicants, the external peer reviews hardly influenced the final grant allocation decisions. While panel members have to review a broad range of proposals, therefore being no expert on most of the topics, external reviewers are considered to be the experts on the specific topic. Except for a few positive outliers (top talents), no evident pool of talents could be identified based on the various review scores. The differences in scores between just selected and rejected applicants are very small, while the consequences of these funding decisions are of great importance for the careers of (especially) young researchers. For improving the transparency, quality and legitimacy of grant allocation practices, it would therefore be important to uncover more deeply the details of the de facto (implicit and explicit) applied criteria.

In this study we will look at the process of talent evaluation in more detail. We will analyse how reviewers individually evaluate talented researchers in the context of grant allocation and how talent is reviewed and discussed within the decision-making panel. Which characteristics of the applicants do they value the most and how do they reach agreement on this within a group? This will give more insight in the predominant notion of talent in external talent selection. Furthermore we will investigate the notion of talent that prevails in a more general academic setting: how do (associate) professors recognize talents in their own department? By comparing these different contexts of talent evaluation we will clarify the notion of talent in science.

Theoretical background

The word 'talent' clearly has a positive connotation, but there is no general consensus on the exact meaning of it. A highly debated issue, for example, is the origin of talent: is talent innate or acquired (e.g. Baron-Cohen, 1998; versus Howe, Davidson & Sloboda, 1998)? In order to get a better grip on the term 'talent', Tansley (2011) studied the historical and linguistic development of the notion of talent over the years. The first dictionary definition of talent refers to a unit of weight, used by the Greek and Romans. Later in the bible it acquired in English the meaning of a monetary unit, which the Greek thereafter translated in terms of 'capital'. In a way this meaning of talent is currently still used in HR practices, as they often use the term 'human capital'. From the thirteenth century onwards the meaning of 'talent' changed relating more to a disposition, special ability or aptitude.

In this paper we consider talent as a form of 'symbolic capital', according to the concept of Bourdieu (1986). He discerned several forms of capital, most importantly economic (financial resources), cultural (education and upbringing) and social (relations and networks). In every societal field there is competition for accumulating as many of these types of capital as possible. Also within science. The position someone holds within the academic field depends on the types

of capital one can obtain and subsequently converse into symbolic capital. Symbolic capital involves reputation and prestige, the way one is valued by others. In science this is mainly determined by the judgement of peers. Scientific quality is what peers qualify as such. This also holds for the identification of talent: it is all about which qualities the scientific community value the most.

Following the example of Van den Brink (2009), who applied the concept of symbolic capital to academic careers, we will in this paper differentiate between professional, individual and social capital. In short, professional capital involves the track record in terms of education and publications. Individual capital is about personal traits and social capital about networks.

Evaluation of scientific quality is often carried out in panels, e.g. in the context of grant allocation. To understand how talent is evaluated and selected in these panels, it is not enough to only study the characteristics of the talents and the reviewers. The panel decisions are influenced by and the result of group interaction, making group and context variables important to include. From literature reviews on this type of panel reviewing, we learn that, for example, group composition, group dynamics (e.g. discussion, sharing of information, power relations), characteristics of the procedure and contextual factors (e.g. budget, time pressure, accountability) can strongly affect the decision outcomes (Olbrecht & Bornmann, 2010; Van Arensbergen et al., *forthcoming*).

These factors impede the transparency and predictability of the decision-making process. However, as this type of panel evaluations involves interaction between human beings, it needs to be considered as a social and emotional process. Therefore, it is impossible to completely rule out any form of subjectivity (Lamont, 2009).

In this paper, we approach the process of talent selection as a strongly subjective process. We aim to study the process of talent selection as fully as possible by investigating the decision-making process on both the individual and on the group level. We look at how reviewers use the formal procedures and interpret the formal criteria in their own way when evaluating grant applications. Furthermore, we analyse the panel discussions and the way the panels reach their final allocation decisions. By studying the process of personal grant allocation, we aim to get a better understanding of the notion of talent that prevails within the scientific community.

Data and methods

We conducted 29 semi-structured interviews with members of grant committees within the Talent Scheme of the Dutch Research Council. They were involved in reviewing and allocating personal career grants within two funding programs:

– The early career grant scheme (ECG) for researchers who received a Ph.D. within the previous three years. The grant offers them the opportunity to develop their ideas further.

– The intermediate career grant scheme (ICG) for researchers who completed their doctorates within a maximum of 8 years and who have already spent some years conducting post-doctoral research. The grant allows them to develop their own innovative research line and to appoint one or more researchers to assist them.

The interviewees (16 males, 13 females) are predominantly associate or full professor and come from various scientific domains, from social sciences to life sciences (see table 1 for more details on the interviewees). Most of the interviewees have been involved in this type of grant allocation for several years, have experience in internal selection processes at the university, and have been active as (national and international) peer reviewers.

Table 1. Overview of the interviewees per program, domain and gender

	ECG		ICG		Total
	Male	Female	Male	Female	
Social sciences, humanities	4	5	2	3	14
Science, life sciences	7	3	2	2	14
Total	**11**	**8**	**5***	**5**	**29***

* including one panel member for cross disciplinary applications

The interview mainly involved questions on the selection of talent within the personal career grant scheme. E.g. how do they review the grant applications; which criteria do they use in the different phases of the selection process; how do they recognize the top talents; how are the applicants discussed within the panel; how does the panel reach the final allocation decisions? Part of the interview also focused on talent selection in their own academic environment.

All interviews were transcribed and coded using the software programme Atlas.ti.

Results

Because of the high numbers of applications in relation to the councils budget, the grant allocation procedure often involves a preselection. In some cases panel members have to reject almost half of the applications in this first round. In this stage the applications are not sent to external reviewers who are considered to be the experts on the specific topic. Because the panel members are responsible for reviewing a large and broad set of applications, they tend to primarily focus on the curriculum vitae and to a smaller extent on the abstract of the research proposal. The curriculum vitae, a form of professional capital, is found to be generally easy to review for all the applicants, regardless of disciplinary proximity.

In the next round when the panel has to select which applicants to invite for the interview, external reviewers are involved. In an earlier study the external reviews were found to play a modest role in the selection process, correlating moderately ($\tau = 0.52$) with the panel reviews, which are eventually decisive (Van Arensbergen & Van den Besselaar, 2012). This can be explained by the finding that the panel members don't automatically take over these expert reviews, but evaluate and weigh them. They primarily want to understand how the external reviewers arrived at their given score. When the external review lacks a clear motivation or when the panel members disagree, they can decide not to take it into account when formulating their own score.

The most important criteria the individual panel members apply in evaluating the quality of the applicant are: number and type of publications (first author or other), international experience, acquired funding and awards, all forms of professional capital. These last two criteria serve as an indicator for talent. It shows that others have recognized them as talent in the past, *so it has to be a very talented applicant.*

When evaluating the quality of the research proposal the focus predominantly is on originality and innovativeness of the research topic, and on the elaboration and feasibility of the proposal. Furthermore panel members want to be convinced that the proposal is really the applicants work. It should be part (or the beginning) of his or hers own line of research, not of their promoter. However, the most important criteria panel members use is: do I understand the proposal? It should be written in a very clear way, generally understandable for everyone in the panel, not just for the experts on the topic. The skills related to these criteria can all be considered types of individual capital.

In the next stage a selection of the applicants is reviewed in a face to face interview. This is found to be a very decisive phase in the selection process (see also Van Arensbergen & Van den Besselaar, 2012). It serves as an opportunity to test several of the criteria earlier applied to the written application. Mainly the authenticity of the proposal is tested. The way applicants answer the questions posed by the panel, better reflects their personal skills and ideas. Other more subjective criteria of great importance used during the interview are enthusiasm, ambition, perseverance and persuasiveness, all forms of individual capital.

After the reviewing process on the individual level, we now turn to decision-making process on the group level. The panel discussion gives reviewers the opportunity to explain criteria specific to their discipline, e.g. publication practises or research methods. In general the panel goes along with the opinion of the expert within the panel. However, the influence of the level of expertise on the evaluation varies. On the one hand experts generally have affinity with the topic, which can make them put extra effort in convincing the other panel members of the strengths of that application. On the other hand their knowledge can make them more critical, identifying more easily the weaknesses of the application. Panel members are also found to be quicker enthused by topics they don't know much about.

The panel spends most of its deliberation time on the large middle area in between the few applicants at the top and bottom. Quality differences within this group are very small, leading to a rather arbitrary boundary between just selected and just rejected applicants. On average their scores are almost the same, but they generally vary on different criteria. Therefor it is very important which criteria the panel members emphasize. Random factors are found to play a decisive role in this. For example, the strong or weak points of the applicant the first speaker mentions to start the discussion, is strongly supported by the other panel members. The composition of the panel is also found to be very decisive, as this determines the available expertise. Personal preferences and atmosphere within the panel too determine final selection decisions.

Although the panel members are convinced the top talents are granted, they indicate to have strong doubts about part of the allocation decisions. Since the quality differences are small and random factors influence decisions in the middle group, many rejected applicants could have been granted as well.

Turning to the identification of talent by experienced researchers in their own academic work environment, we found that they use a rather broad concept of talent. Talents excel on various dimensions, combining all forms of symbolic capital. They have a broad expertise, excellent writing skills and a high productivity. Furthermore, they can work really hard, have a strong ambition, enthusiasm and perseverance. Social skills are also highly valued: communication skills, a proactive and social attitude and team spirit. Finally, talents are not always recognized as such immediately. Some time is needed for talent to develop and become visible.

Conclusion

In the context of grant allocation, the notion of talent is narrowed down to mainly professional capital. Number of publications and acquired grants are key criteria used in the selection process, especially in the first phase of the procedure. In a later phase where a selection of applicants is interviewed face to face, elements of individual capital come in. Besides testing the applicant's scientific expertise, reviewers now evaluate motivation, enthusiasm, ambition and uniqueness. The notion of talent is narrower in formal talent selection compared to informal talent selection, in which all three forms of symbolic capital are highly valued.

Within the panels a lot of time is spend on discussing the applicants in the so called 'grey area', in which there are minimal differences between the applicants. Despite of the thoroughness of the procedure and the sincere efforts of the panel members, random factors play an important role in part of the allocation decisions. Because of the nature of the selection process, involving human interaction, these factors can never completely be excluded. But it is important to be aware of the subjectivity and extent of randomness of these allocation decisions, since they have important consequences for the careers of many young researchers. Not only because they directly provide

researchers with resources to conduct research, but also because they are indirectly provide career opportunities as indicators of talent.

References

Arensbergen, P. van, Besselaar, P. van den & Weijden, I. van der (*forthcoming*). The Selection of Talent as a Group Process. Den Haag: Rathenau Instituut.

Arensbergen, P. van & Besselaar, P. van den (2012). The selection of scientific talent in the allocation of research grants. *Higher Education Policy*, 25, 381–405.

Baron-Cohen, S. (1998) 'Superiority on the embedded figures test in autism and in normal males: evidence of an "innate talent"?' Behavioral and Brain Sciences 21(3): 408–409.

Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.) *Handbook of Theory and Research for the Sociology of Education* (Vol. 1st edition, p 241–258). New York: Greenwood.

Brink, M. van den (2009). *Behind the scenes of science. Gender practices in the recruitment and selection of professors in the Netherlands*. PhD thesis, Radboud Universiteit Nijmegen.

Howe, M.J.A., Davidson, J.W. and Sloboda, J.A. (1998) 'Innate talents: reality or myth?' Behavioral and Brain Sciences 21(3): 399–407.

Lamont, M. (2009). *How professors think. Inside the curious world of academic judgment*. Cambridge: Harvard University Press.

Olbrecht, M. and Bornmann, L. (2010) 'Panel peer review of grant applications: what do we know from research in social psychology on judgment and decision-making in groups?' Research Evaluation 19(4): 293–304.

More competition, better science? The predictive validity of grant selection

Peter van den Besselaar p.a.a.vanden.besselaar@vu.nl / Department of Organization Science & Network Institute, VU University Amsterdam, De Boelelaan 1081, Amsterdam, 1081 HV (Netherlands)

Abstract

Does career grant competition result in the selection of the best young talents? In this paper, the predictive validity of grant decision-making is investigated, using a sample of 250 early career grant applications in three social science fields. We measure output and impact of the applicants about nine years after the application to find out whether the selected researchers perform ex post better than the non-successful ones.

In this first test, we find that predictive validity varies sometimes is high (psychology) but sometimes low (economics), when comparing grantees with *all* non-successful applicants. However, when comparing grantees with the *best* performing non-successful applicants, predictive validity varies between low (psychology) to negative (economics). Furthermore, our findings suggest that predictive validity decreases with increasing competition, which has theoretical implications for the understanding of how panels work, and for how scholarly talent is (or cannot) recognized. The analysis is based on publications in social science journals only. In a following step, we will extend the analysis to publications in science journals.

Introduction

An important question about peer and panel review is the predictive validity: does the post-performance of selected researchers legitimize their selection: do they outperform those that were not selected? Unfortunately, data to investigate this are scarce. Not surprisingly, a recent review of research on peer review (Bornmann 2011) could only identify six studies on the predictive validity of grant peer review (Armstrong 1997; Bornmann et al, 2005, 2008, 2010; Hornbostel et al, 2009; Melin & Rickard 2006; Van den Besselaar et al 2009). Recently, a few other studies have been published, indicating the growing interest in the subject (Campbell et al, 2010; Li et al, 2010, Neufeld & von Ins 2011; van Leeuwen & Moed 2012).

Some of the available studies compare successful applicants with all unsuccessful applicants. Generally, the former are reported *in average* to outperform the latter. This holds for past performance (before the grant application) and post performance (a few years after the application). Despite this positive relation, not all selected applicants of course score higher than all rejected

ones. A significant number of false positives and false negatives is reported, implying that about one third to two third of all decisions can be considered wrong if one accepts the deployed performance criteria as meaningful.

Several of these studies compared successful applicants with (an equally large set of) the *best performing non-successful* applicants. Some found a higher post-performance for the grantees, despite the fact that this was not the case for past performance. This suggests would that grant decisions may not select the best, but produce the best through providing resources. Other studies found no differences between past and post performance between the two groups.

This study builds on earlier work (Van den Besselaar et al 2009; Bornmann et al 2010). Here we go beyond our own and other earlier studies in the following ways: (i) Mid career researchers and the advanced researchers do have a variety of funding possibilities. One cannot control for that, as information lacks about applications elsewhere, e.g., the ERC career grants. However, this does not hold for early career researchers, who have mainly one funding source available. By restricting the analysis in this paper to the early career program (here: the Dutch VENI program), we most probably have disjoint sets of successful and unsuccessful applicants, not influenced by other grants. (ii) Our previous study only included a rather short period of post performance, between two and four years after the *start* of the project. As the time between having results and having those published is fairly long in the social sciences, and citing that work takes some more years, our earlier citation data (early February 2007) hardly measure real post performance. (iii) Comparisons are generally made in terms of averages. As distributions are skewed, we switch here to non-parametric statistics. (iv) The context of decision-making is taken into account, as e.g., the way the decision-making is organized, influences the process and its outcomes (Langfeld 2001; Van Arensbergen et al 2012; Van den Besselaar et al 2013); (v) data collection is improved, such as better disambiguation of authors.

Data and method

The data consist of 400 early career researchers in the social sciences, including name, age, university, field and discipline, reviewer scores, and panel decision. The applicants obtained their PhD between 2000 and 2002, and grants were obtained in the period 2003–2005. Post performance of these early career researchers is measured by all publications between 2000 and 2012, and citations received at 31-12-2012.

Post performance is defined in terms of productivity (publications), overall impact (citations), and average impact (citations per publication). Similar to the methodology used in an earlier study (Van den Besselaar et al 2009), the analysis in this paper is restricted to publications indexed in the Social Sciences Citation Index (WoS-SocSCI). Although this will underestimate output of those researchers who also publish in e.g., mathematical and statistical journals (not uncommon for economists), or life science journals (not uncommon for psychologists), we expected that this

does not influence the comparison at group level. Given the size of the sample used in the 2009 study, this underestimation was expected to be equal for the successful applicants, the non-successful applicants and the best-performing non-successful applicants. However, we are currently testing this. Finally, data for this analysis were collected manually[i], in order to improve identification and disambiguation of authors.

The analysis was done for applicants within economics, psychology, and educational and behavioral sciences. In these fields publishing in international journals has become the normal practice, possibly with the exception of educational research. Other social science fields are either insufficiently covered by the WoS (anthropology, sociology, law), or too small in our sample for the current analysis (geography, communication science, organization science). Competition (success rate) is rather different between the three fields under study (Table 1).

Table 1. The sample[ii]

	Applicants	Success rate
Psychology	95	28.4%
Educational and behavioral research	48	20.1%
Economics	101	12.9%
Total	**244**	**20.5%**

We will compare the publications, citations, and citations per publication of successful applicants (S), the equally large set of best performing rejected applicants (BPNS)[iii], and all rejected applicants (NS). Earlier studies were generally based on comparing means. However, as distributions are non-normal and outliers cannot be discarded, one needs to turn to non-parametric statistics. We will test whether the medians between the relevant groups are different and whether the performance distributions of these groups are different (Mann-Whitney; sample median test). Based in this, we introduce a measure for predictive validity.

We will test the following four null hypotheses for each of the three disciplines:

There is no difference between

(1) *median post performance* (publications, citations, citations per publication) of the successful applicants (S) and the non-successful applicants (NS)
(2) *post performance distributions* (of publications, citations, citations per publication) of S and NS
(3) *median post performance* of S and the best performing non-successful applicants (BPNS)
(4) *post performance distributions* of S and BPNS

Testing these hypotheses will enable us to find out whether the successful applicants outperform the others by the end of 2012, implying that they either were better when applying, or have

become the better through the resources and possibilities a career grant offers. An earlier analysis suggested that the selected applicants did not (ex ante) perform better than the best performing non-selected researchers (Van den Besselaar et al 2009). Are the grantees better now? Finally, by comparing the findings between the three disciplines, we may find out as whether the level of competition influences the predictive validity.

Findings

The hypotheses about the medians and the distributions of *post performance* of the three groups (S, BPNS, NS) within the three disciplines are tested, using the SocSCI publications. Table 2 shows the first results. If we compare post performance of successful applicants with *all* unsuccessful ones, we do not find performance differences within economics. We do find the granted applicants having a better post performance distribution than the non-successful applicants in psychology. Within educational and behavioral research, the finding is mixed: for some indicators, grantees' distribution is above the others' distributions, for other indicators they do not.

And what when we compare the grantees with the *best performing* non-successful applicants? Within economics, the BPNS have a better post performance distribution than the successful applicants. In psychology, the picture is mixed: according to two indicators, the grantees do better, but the four other indicators show no difference. Within educational and behavioral research, no differences between the groups are found.

In order to draw conclusions from these findings, we calculate a *predictive validity score* (PVS), which averages the predictive validity of the different indicators used. It is calculated in the following way, using applications in economics as example. We have six comparisons (median and distribution of P, C, and C/P) between the groups. When a comparison supports predictive validity, it gets a score of 1. When a comparison contradicts predictive validity, it scores -1. And if there is no difference, the score is 0. The average of the scores is used: for publications in economics (granted versus all non-granted), we found no difference between the two groups, so the score is (6*0)/6 = 0. For publications in economics (granted versus best non granted), we found for two comparisons no differences, and in four we found that the non-granted performed better. Here the predictive validity score is ((4*-1)+(2*0))/6 = -.67. The PVS-scores are in the right column of table 4. We find a rather low predictive validity score for education and an even lower and negative for economics. Psychology has a high PVS if we compare the granted applicants with all others, but in case the comparison with the best performing others, the predictive validity score drops to 0.33.

Table 2. Success by post performance and discipline – some provisional results

Null hypothesis: no differences between:			Medians*	Distributions*	PVS**
S versus NS	economics N=101	Publications (P)	Retain	Retain	0/6 = 0
		Citations (C)	Retain	Retain	
		Citations/Publication (C/P)	Retain	Retain	
	Psychology N=95	Publications	S > NS	S > NS	6/6 = 1
		Citations	S > NS	S > NS	
		Citations/Publication	S > NS	S > NS	
	Education N=48	Publications	Retain	S > NS	2/6 = 0.33
		Citations	Retain	S > NS	
		Citations/Publication	Retain	Retain	
S versus BPNS	economics N=26	Publications	Retain	Retain	-4/6 = -0.67
		Citations	BPNS > S	BPNS > S	
		Citations/Publication	BPNS > S	BPNS > S	
	Psychology N=54	Publications	S > BPNS	Retain	2/6 = .33
		Citations	Retain	Retain	
		Citations/Publication	S > BPNS	Retain	
	Education N=20	Publications	Retain	Retain	0/6 = 0
		Citations	Retain	Retain	
		Citations/Publication	Retain	Retain	

S = successful applicants; BPNS = best performing non=-successful applicants; NS= non-successful applicants; N= number applicants

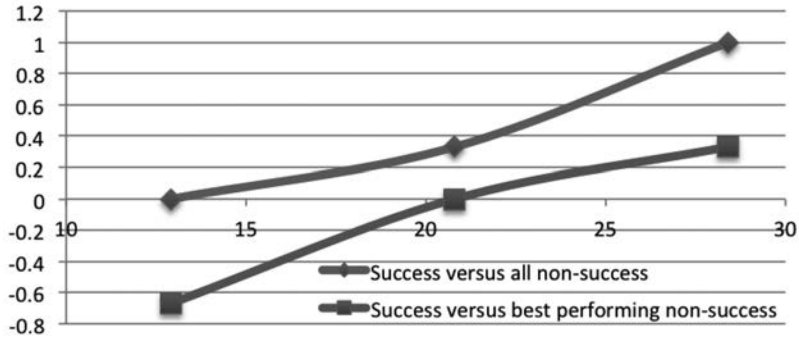
* Significance level = 0.10

** Predictive validity score (explained above in text)

Context and predictive validity – an initial analysis

In Fig. 1 gives the six predictive validity scores for the three fields, for S versus NS and for S versus BPNS by success rate. As this is a small sample, and the analysis does not include publications in non-social science journals yet, we cannot draw any definitive conclusions. Nevertheless, figure 1 suggests a positive relation between predictive validity score and success rate, confirming our expectation.

Figure 1. Predictive validity index score (Y-axis) by Success rate in percentages (X-axis)



Why would this be the case? When selecting scientific talent, panel members are influenced by scientific norms, and also by interests. What dominates may be dependent on the scarcity or abundance of resources. If resources are scarce and success rates very low, representing (disciplinary) interests may dominate. E.g., if there are not enough grants for every specialty, panelist may have highest priority to secure a grant for their own specialty over selecting the best applicants. And if selection is more based on interests than on the quality of the applicant and the proposal, predictive validity is expected to be lower. Or to put it differently, strong competition among applicants may lead to competition within panels; whereas lower competition among applicants enables collaboration in panels to select the best set of applicants. This suggests that although competition drives science, too much competition may destroy the normative order that is necessary for a good functioning science system.

Also another common belief has to be questioned. Interviews among panel members show that they are convinced that real talents are easily recognized. Within the larger sub-top, it becomes more difficult to differentiate between those that should be funded and those that should not (Van den Besselaar et al 2013). If that would be the case, the predictive validity would be better if only a few applicants have to be selected (the real top) than if a larger selection has to be made. Our findings, however, suggests exactly the opposite. The 'evidence' of talent may be a mere myth.

Conclusions and next steps

This analysis suggests that predictive validity is only high when (i) comparing grantees with all non-grantees, and when at the same time (ii) success rate is rather good. For the other situations (comparison with BPNS applicants; low success rate), predictive validity is low. We suggested some possible theoretical explanations above. However, further empirical analysis is needed too. The following next steps are on the agenda:

(1) Extending the performance measures from only SocSCI to also SCI data, to cover the output and impact more comprehensible;

(2) Testing the hypotheses in other fields;

(3) Exploring different indicators for performance. Are the (common) indicators, also used in this paper, valid for measuring post performance? As argued elsewhere, success may be related not so much to the common output and impact measures, but more to proper indicators of scholarly quality, reflecting independence (Van den Besselaar et al 2012), and scientific innovation. It is worthwhile, theoretically and from a policy perspective, to explore this further.

References

Armstrong PW, Caverson MM, Adams L, Taylor M, Olley PM (1997). Evaluation of the Heart and Stroke Foundation of Canada Research Scholarship Program: research productivity and impact. *Canadian Journal of Cardiology* 13, May, 507–16.

Bornmann L (2011). Scientific peer review. *An Rev Inf Sci Tech* 45, 199–245.

Bornmann L & Daniel H-D (2005). Committee peer review at an international research foundation: predictive validity and fairness of selection decisions on post-graduate fellowship applications. *Research Evaluation* 14, 15–20.

Bornmann L, Wallon G, Ledin A (2008) Does the Committee Peer Review Select the Best Applicants for Funding? An Investigation of the Selection Process for Two European Molecular Biology Organization Programmes. *PLoS ONE* 3 (10): e3480.

Bornmann, L, Leydesdorff L, Van den Besselaar P (2010). A Meta-evaluation of Scientific Research Proposals: Different Ways of Comparing Rejected to Awarded Applications. *Journal of Informetrics* 4 (3) 211–220

Campbell D, Picard-Aitken M, Cote G, Caruso J, Valentim R, Edmonds S, Archambault, E (2010) Bibliometrics as a performance measurement tool for research evaluation: the case of research funded by the National Cancer Institute of Canada. *Am J Eval* 31, 66–83.

Hornbostel S, Böhmer S, Klingsporn B, Neufeld J, von Ins M (2009). Funding of young scientist and scientific excellence. *Scientometrics* 79, 171–190

Langfeldt L (2001). Decision making constraints and processes of grant peer review, and their effect on review outcome. *Social Studies of Science*, 31, 820–841

Li J., Sanderson M, Willett P, Norris M & Oppenheim C (2010). Ranking of library and information science researchers: Comparison of data sources for correlating citation data, and expert judgments. *Journal of Informetrics* 4 (4) 554–563.

Melin G, Rickard D (2006). The top eight percent: Development of approved and rejected applicants for a prestigious grant in Sweden, *Science and Public Policy* 33, 702–712

Neufeld J & von Ins M (2011). Informed peer review and uninformed bibliometrics? *Research Evaluation* 20, 31–46.

Van Arensbergen P & Van den Besselaar P (2012). The selection of scientific talent in the allocation of research grants, *Higher Education Policy* 25, 381–405

Van Arensbergen P, Van der Weijden, Van den Besselaar P (2013), The notion of talent, What are the talents we are looking for in science? In: *STI 2013 Proceedings – this volume*.

Van den Besselaar P & Leydesdorff L (2009). Past performance, peer review, and project selection: a case study in the social and behavioral sciences. *Research Evaluation* 18, 273–288.

Van den Besselaar P & Van Arensbergen P (2013). Talent selection and the funding of research. Academic talent selection in grant review panels. *Higher Education Policy* 26.

Van den Besselaar P, Sandström U, Van der Weijden I (2012) The independence indicator. In Archambault E, et al, eds., *Proceedings STI 2012*. Montreal: OST & Science Metrix.

Van Leeuwen TN & Moed HF (2012). Funding decisions, peer review, and scientific excellence in physical sciences, chemistry, and geosciences. *Research Evaluation* 21, 189–198.

i In contrast, the 2009 study with Leydesdorff is based on automatic mapping between WoS data and application data, using family name and first initial. Spelling variants leading to under- and overestimation were accepted, because of the large number of cases. However some other issues have come up since. (i) The first initial was sometimes taken from the title (PhD, Dr, MSc, LLM, etc.). (ii) Some applicants use different initials in applications compared to publications. Re-analyzing after (partly) correcting (i) and (ii) suggests that this does not influence outcomes of the 2009 study.

ii Sixteen applications were removed because of missing data.

iii Best performing rejected applicants is defined in terms of total citations received.

Personal Research Grants and Career Paths: a Cohort Study of Dutch Life Science & Health Researchers

Inge van der Weijden — i.c.m.van.der.weijden@cwts.leidenuniv.nl / CWTS, Leiden University, P.O. Box 905, Leiden, 2300 AX (The Netherlands)

Rosalie Belder — r.belder@rathenau.nl / SciSa Department, Rathenau Instituut, P.O. Box 366, The Hague, 2509 CJ (The Netherlands)

Ingeborg Meijer — i.meijer@cwts.leidenuniv.nl / CWTS, Leiden University, P.O. Box 905, Leiden, 2300 AX (The Netherlands)

Abstract

As academic career opportunities are to a growing extent dependent on research funding, grant selection outcomes are expected to have an important impact on the careers and motivation of early career researchers. We study the careers paths of Life Sciences & Health researchers who received a prestigious personal research grant in 2007. These grants are awarded to creative and talented early career researchers to give them the opportunity to conduct their own research. We assume that in this way they gain entry to or promotion within institutions conducting academic research in the Netherlands. But what are the implications of this performance-based funding system on the academic careers of these Life Science scientists? First, we mapped the career trajectories of the grantees. Second, we conduct open interviews. In the interviews we focus on individual motivation for career steps, the effect of mentorship and career breaks, as well as on future career plans, including their opinion of successful scientific work. First results show that most grantees continue their career at the same university or University Medical Center. We also report gender differences.

Introduction

Research and researchers are increasingly vital resources in modern society. A powerful and internationally competitive research base depends fundamentally on strong cohorts of highly productive and creative researchers, and therefore on the capacity to attract some of the best minds in each generation from the global pool of talent (LERU 2010). Opportunities for intellectually stimulating work, passion for a field of study, and the opportunity to contribute to new knowledge are attracting people to the academic profession (Bexley, Arkoudis, James 2012). As academic career opportunities are to a growing extent dependent on research funding, selection outcomes are expected to have an important impact on the careers and motivation of early career researchers. In this work-in-progress report we study the impact of research funding agencies schemes, using the case of Dutch early-career grants.

Data and Methods

It is a longitudinal, retrospective and prospective study, from university degree to current position and beyond. We study the careers paths of a cohort of 43 Life Sciences & Health researchers (22 males and 21 females) who received a prestigious early career grant in April 2007 to develop their scientific ideas further for a period of three years. These career grants are allocated via competitive, externally peer-reviewed mechanisms. The cohort under study consists of outstanding researchers, who rank among the top 10–20% of their peer group. First, we mapped the individual career trajectories. We collected the current employment status (both organization and position) of the 43 grantees by searching for a Web CV and/or personal WebPages. In the spring of 2013 we will conduct open interviews with the grantees. In these interviews we will focus on individual motivation for career steps, the effect of mentorship and career breaks, as well as on future career plans, including their opinion of successful scientific work. We will also analyze how this cohort of Life science applicants experience the evaluation and selection processes.

Results so far

The prestigious early career grantees in this sample were in 2007 (time of grant award) between 29 and 40 years, with an average age of 33 years. In 2013, six years after the awarding of the grant, four grantees (1 male, 3 females) have left academia and moved to a non-academic career. 39 grantees stayed in academia; most of them (37) are employed in the Netherlands. Job mobility among this cohort is low. Most grantees (n = 31) continue their career at the same university or University Medical Center. Only 8 scientists started working for a different university after finishing working on their prestigious personal grant. Two of them (both males) went abroad.

Looking more closely at our cohort (see table 1), it is interesting to note that there are large differences in the faculty rank. Five grantees (2 males, 3 females) are still employed as a postdoctoral researcher in 2013. Already nine grantees (6 males, 3 female) are employed in senior positions, as an associate professor or as a staff member. A position as assistant professor is most frequently reported (18 grantees: 12 males, 6 females).

Table 1. Positions of prestigious early career grantees in 2013

	Postdoc	Assistant professor	Associate Professor	Staff member	Other	Non-academic	Unknown
Male n = 22	2	12	3	3	1	1	0
Female n = 21	3	6	1	2	3	3	3
Total n = 43	**5 (11.6%)**	**18 (41.9%)**	**4 (9.3%)**	**5 (11.6%)**	**4 (9.3%)**	**4 (9.3%)**	**3 (7.0%)**

Grant applications are derived from a subset of researchers that have high ambitions and high potential in basic research, which could be gender independent. In this cohort the number of male grantees equals the number of females. This gender distribution is in line with recent studies that show a growing share of women through all phases in academic careers (Gerritsen, van der Sanden et al. 2012). However, our research indicates that within the cohort of prestigious early career grantees, six years after awarding, there are fewer women in senior faculty positions. The number of female associate professors and staff members (3) in our cohort is remarkable lower compared to the males who hold these positions (6). Also the three of the four grantees that left academia are female. This lack of women in senior positions in science, the so called 'leaky pipeline' (Weber 2008) has been a well-known aspect of gender equality debates in higher education in European countries for many years. In the interview study we will elaborate on the gender dimension by studying the bottlenecks grantees perceived in their career steps after being awarded.

References

Bexley, E., Arkoudis, S., and James, R. (2012). The motivations, values and future plans of Australian academics. *Higher Education*. DOI 10.1007/s10734-012-9550-3.

Gerritsen, M., van der Sanden, J., Verdonk, T. and Visser, A. (2012) *Monitor vrouwelijke hoogleraren 2012*. Stichting de Beauvoir, Landelijk Netwerk Vrouwelijke Hoogleraren, Vereniging van Universiteiten, Nederlandse Federatie van Universitaire Medische Centra.

League of European Research Universities (2010) *Harvesting talent: strengthening research careers in Europe*. Leuven: LERU Office.

Recent developments in Dutch research assessment. Scholarly activity in SSH and Law in a new perspective

Thed van Leeuwen Leeuwen@cwts.nl / CWTS, Leiden University, Wassenaarseweg 62a, Leiden, PO Box 905 (The Netherlands)

Erik van Wijk wijk@CWTS.LeidenUniv.nl / CWTS, Leiden University, Wassenaarseweg 62a, Leiden, PO Box 905 (The Netherlands)

Paul Wouters p.f.wouters@CWTS.LeidenUniv.nl

Introduction and research question

Increasingly, scientific and scholarly research is periodically evaluated according to particular (national) protocols. In the Netherlands, the Standard Evaluation Protocol (SEP), designed and supported by the Association of Dutch universities (VSNU), the national Research Council (NWO) and the Royal Dutch Academy of Sciences and Arts (KNAW) is the framework. In many of these assessments, peer review is complemented by bibliometrics. The application of bibliometrics is not explicitly required by the SEP. Most bibliometric techniques use data derived from the journal literature and the currently available citation databases. Therefore, these techniques are less suitable for research fields in which international journal literature plays a less dominant role (Moed, 2005; van Leeuwen, forthcoming). For these fields, in particular the humanities, the more qualitative parts of the social sciences (such as anthropology, political science, sociology, etc.), and law, a detailed analysis of the various types of output coming from these fields is necessary as the basis for any form of performance measurement (Sivertsen & Larsen, 2012). This has been recognized internationally. For example, the protocols in Norway, Belgium, and Denmark are based on integral registration of all research outputs, across all fields of scholarly activity (Ossenblok et al, 2012).

Although these systems allow for a precise registration of all forms of output, the problem of assessing their influence on the scientific community, on professional practices, and on society at large still remains unsolved. This paper will provide a detailed analysis of the complete output of one university across all fields and an exploration of proper bibliometric webometric, and altmetric methodologies to assess the influence of this output. Where we cannot yet establish a solid solution, we propose a number of research questions as elements of a bibliometric research agenda.

Policy context

In the Netherlands, the KNAW has started initiatives to design a framework on how to organize research assessments in the arts and humanities and the social and behavioral sciences (KNAW 2011, KNAW 2012). This framework is based on two pillars: scientific quality and societal impact. Within each of these two domains, three criteria are distinguished: output, usage of the output, and the recognition of scholarly activities. For each of these three criteria various indicators are defined. These are not prescribed, the system is defined in a flexible way, so that indicators are selected by the field under review itself. Data for this type of research assessment can be extracted from local or national research information systems.

Data & Methodology

The data in this study are extracted from the research information system Metis. In Metis, one can include various types of scientific output, and output is labeled by a fine-grained document categorization system. The output data we work with in this study is the publication output registered on an annual basis over the period 2004–2009, as that indicates most clearly and directly what the units under study conceive themselves as their complete and unfiltered output. In order to be able to indicate the adequacy of current data systems providing a basis for bibliometric studies, we need to know the degree to which the output of university is processed within the WoS (the *external coverage*). In a next step we focus on the *internal coverage*, defined as the share of the references given by the university publications to other WoS covered publications. A relative high/low degree of referring to WoS journal publications indicates a high/low relevance of journal publications for the communication process, and if observed, allows a strong(er)/weaker focus on journal publications as a basis in the assessment of the research conducted by the units under study. In this study, the publication data retrieved from Metis were compared and combined with the indicators defined in the KNAW assessment scheme for the Humanities, as described above.

Conclusions

In this study, which started from the perspective of the total output of a Dutch university and the units therein, we show the possibilities and limitations of the current bibliometric methodologies and techniques. In an attempt to improve the situation for the social sciences, law and the humanities, we combine the output data of the university under study with the assessment scheme designed by the Humanities field itself. The study, of which the results will be presented at the conference, leads to a number of conclusions on various topics dealt with in the study:

– The relative low degree of relevance of current standard bibliometric techniques for the social sciences, humanities and law; the study clearly illustrates the dominance of (bio)

medical research over other disciplines present in the university when only journal literature is used for assessment purposes.

- The importance of External/Internal coverage analysis, particularly with respect to the social sciences, humanities and law; these type of analyses can serve as a shield against incorrect usage of metrics in an assessment context.
- The variety of the different types of scientific communication registered within an institutional research information system such as Metis; this various types of output offered insight in a wide variety of scholarly activities, both in the scholarly domain, as well as in the civil-society domain, thereby offering opportunities to assess also the societal quality aspects of scholarly work.
- The possibilities to link the registered output to the assessment scheme proposed by the KNAW advisory council for the Humanities;
- The dependence on the quality of the input into a research information system such as Metis; this relates to the element of standardization, in this case of the document types registered in the system, this should be done in a consistent and uniform way, as that guarantees a higher quality of information used in an assessment context.

References

Moed, HF, "Citation Analysis in Research Evaluation", 2005, Springer.

KNAW Report "Quality indicators for research in the humanities", 2011, Amsterdam http://www.knaw.nl/Content/Internet_KNAW/publicaties/pdf/20111024.pdf

KNAW Report "Quality indicators for research in the social sciences", 2012, Amsterdam (not yet public).

Ossenblok, T, T Engels, and G Sivertsen, "The representation of the social sciences and humanities in the Web of Science-a comparison of publication patterns and incentive structures in Flanders and Norway (2005-9), *Research Evaluation*, Vol. 21 (4), 280–290.

Sivertsen, G, and B Larsen, "Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: an empirical analysis of the potential", *Scientometrics*, 2012, vol. 91 (2), 567–575.

Van Leeuwen, TN, "Bibliometric research evaluations, Web of Science and the Social Sciences & Humanities: a problematic relationship ?", forthcoming.

The "Translational" effect in medical journals. Bridging the gap?

Erik van Wijk wijk@ cwts.leidenuniv.nl / Center for Science and Technology Studies (CWTS), Leiden University, PO Box905, 2300 AX Leiden (Netherlands)

Marc Luwel luwel@ cwts.leidenuniv.nl / Center for Science and Technology Studies (CWTS), Leiden University, PO Box905, 2300 AX Leiden (Netherlands)

Abstract

The objective of translational research is to harness knowledge from basic research to develop new drugs and treatments. For already two decades it is a 'hot' topic in academia and in policy-making circles. However few bibliometric studies have been carried out to make the characteristics of translational research visible. In this paper an analysis is presented of publications in journals covering translational medical research and processed for the Web of Science. These publications do not have more references to basic research than those in the other journals assigned to the same subject categories. Neither were significant differences observed between these two journal sets in the noun phrases used in the papers' abstracts , except for the abundant use of the adjective 'translational' in papers published in medical journals with the adjective 'translational' in the journal title.

Conference topic

Knowledge transfer, knowledge exchange and specific issues relating to translational research
Cross-sectoral research collaborations: fertilization or impediment?

Introduction

Over the last decades research in life sciences has progressed dramatically but it was perceived that the realisation of the possibilities they created to improve health, remained stubbornly slow. In the last decade of the twentieth century the idea took ground to overcome the death valley separating basic science from its clinical applications and to improve the 'bench-to-bedside' process of harnessing knowledge from basic research to develop new drugs and treatments that can be used clinically and commercialised (Butler, 2008).

The term 'translational research' was coined for this process and it entered the scientific jargon. Although the concept needs to be further clarified (Woolf, 2008), it became quickly at the forefront of science policy making. Confronted with increasing health care costs and an aging popu-

lation, governments in several countries supported the development of translational research by targeted investments and the establishments of new research institutes.

In the scientific literature a growing number of papers report on translational research or on results relevant for translational applications. Although powerful tools have been developed to study emerging research fields and the flows of knowledge between scientific disciplines, only a limited number of bibliometric studies have been carried out on the development of translational research. Cambrosio et al (2006) made a bibliometric analysis of the emergence of translational research, but limited to oncology. Mapping the evolving patterns of inter-citation relations between publications on cancer they found evidence of the emergence of a translational interface between basic and clinical research. Luwel (2008) analyzed for the period 1987 and 2006 the characteristics of papers published in journals processed for the Web of Science (WoS) containing the adjective "translational" in the title or abstract. Its growth rate was much lower than for nanoscience, another emerging research field. Based on a content analysis a fuzzy picture emerged indicating more the development of a methodology than the emergence of a new discipline. Using network analysis and mapping techniques Jones et al (2011) studied the translational links between basic and clinical research in cancer and cardiovascular medicine. For cancer research a translational interface composed of different techniques was made visible. For cardiovascular research no distinctive translational links were found, but these authors claim that translational research does occur in certain subdisciplines.

Starting in 1999 with the journal 'Clinical & Translational Oncology', journals with the adjective 'translational' in their title covering as stated in their editorial mission and guidelines translational research are established and some processed for the WoS. The main objective of this paper is to compare their characteristics with other journals assigned to the same ISI subject categories.

Methodology and data

Table 1 gives an overview of medical journals with the adjective 'translational' in the title processed for the 2012 edition of the WoS, the assigned subject categories and the first year they are covered by this bibliographic database. Five of the journals are dedicated to medical sub-disciplines such as oncology and the other four cover the broad fields of laboratory, clinical and public health research. The editorial boards state that publishing research spanning from bench to bedside is the scope and aim of these nine journals.

The flows of knowledge between disciplines and interdisciplinary work can be studied by analyzing cited and citing documents as well as the co-occurrence of noun phrases in documents. Most journals listed in table 1 are only fairly recently processed for the WoS, the feasible citation window is as yet too short to carry out a robust citation analysis. Therefore this paper reports on the study of the documents cited in publications in the journals listed in table 1 and on the preliminary results of the analysis of noun phrases in the abstracts of these papers.

In life and natural sciences authors get the ownership of the results of their work by publishing them in the open literature; publications in peer reviewed journals being the most widely used communication channel. References in these publications can be applied as a proxy for the knowledge used to get new insights or to develop new applications. The characteristics of the references mentioned in the publications in the journals listed in table 1 are compared with those of research work published in the other medical journals assigned to the same subject categories. As the aim of translational research is to bring laboratory results faster and more efficiently to clinical applications we assumed that in the reference lists of the former the basic disciplines are more prevalent than in the latter.

To test this hypothesis, reference profiles of the journals were constructed. From the references mentioned in the papers published in the journals listed in table 1 the subset published in journals processed for the WoS was identified and assigned to these journals' subject categories. This process was carried out for each year the journal listed in table 1 was processed for the WoS. As we are interested in original work on translational research, only the publication type "article" and "letter" were taken into account. Review articles were discarded as they attempt to summarize results and give an overview of the state of the research on a particular topic. For references to papers in journals assigned to more than one category a whole counting and fractional counting scheme was used. In this paper only results based on the whole counting scheme are presented. In a next step for the other journals assigned to the categories mentioned in table 1 and for the publication years 2006 till 2011, the references to papers published in journals processed for the WoS were assigned to these journals' subject categories. The categories were restricted to those assigned to the journals mentioned in the references lists of the papers published in the journals listed in table 1. With this limitation the two journal sets cover the same disciplines. Again only the above mentioned publication types were taken into account. For each of the journals in the two sets the fraction of references in the different subject categories was calculated.

To operationalize the concept 'basic research' a subset of subject categories was used. The results presented in this paper are based on a subset of 80 categories, further called 'the basic categories' (table 2). Finally for each journal listed in table 1 the total fraction of the references assigned to the basic categories is compared with the corresponding fraction for each of the journals in same subject category.

An alternative approach to study differences between the journals listed in table 1 and the other journals assigned to the same categories is a content analysis. From the abstracts of the publications in the journals listed in table 1 the 1000 most frequently used noun phrases were extracted and for each noun phrase the number of times it occurred was counted (its absolute frequency). Again only the publication type 'article' and 'letter' were taken into account. From all the other journals assigned to the categories listed in table 1 the same information was collected. For each of the two lists of 1000 noun phrases the relative frequency of each noun phrase was calculated. To test if there is a statistically significant difference between the two lists, an analysis of the variance of the relative frequencies is made (Altman, 1983). For each noun phrase the difference

between the relative frequency in both lists was calculated as well as the mean of these values. If the differences are normally distributed at an $\alpha = 0.05$ the confidence level of the differences will lie between $\langle d \rangle + 1.96^* \sigma$ and $\langle d \rangle -1.96^* \sigma$; $\langle d \rangle$ is the average difference and σ is the standard deviation of the differences. This method makes it also possible to identify the noun phrases lying outside the limits of agreements.

In the paper the noun phrases occurring only in the journals listed in table 1 and their frequency are also analyzed in an attempt to make conceptual differences between these journals and the other journals in the subject categories of table 1 visible.

Results

Table 1 gives for each journal the fraction of the references to papers published in journals assigned to the basic categories. This fraction is compared with that of the other journals in the same category and the journal's rank is given. For each category the highest fraction of the references to papers published in journals assigned to the basic categories is also given. Several journals listed in table 1 are only assigned to one category. To analyze the influence on the ranking of journals assigned to several subject categories, for each journal listed in table 1 the same ranking is made but limited to journals assigned only to its category or, in case of more than one category to each of its categories separately. For example based on the total fraction of references to the basic categories in journals in the subject category 'Oncology' the journal titled 'Clinical & Translational Oncology' ranked 102 out of 194 journals; 92 of these journals are only assigned to this category and among these journals 'Clinical & Translational Oncology' ranked 44.

Only the journals titled 'Translational Research' (in the subject category Medicine, General & Internal) and 'Journal Of Cardiovascular Translational Research' have (nearly) the highest fraction of references assigned to the basic categories. For the other journals and for the journal titled 'Translational Research' in the two other categories, this fraction is considerably lower. Moreover the outcome of the two ranking methods is fairly similar.

These results contradict the hypothesis that compared to the other journals in the subject categories the journals listed in table 1 would contain more references to work published in journals assigned to the basic categories.

A sensitivity test was carried out defining 'basic categories' by fewer categories than presented in table 2. Although the position of a journal in the ranking could somewhat shift, the overall picture remained unchanged.

To study the cognitive content of the papers, the average and the difference of the relative frequencies of the noun phrases mentioned in section 2 were analyzed. The differences are fairly normally distributed (number of noun phrases = 1287, mean = 0, median = 0,000020, standard deviation

= 0,00074 and Shapiro-Wilk W = 0.68). Both lists of noun phrases are very similar. Of the 1287 noun phrases occurring at least in one list, 38 or 2.9% lay outside the limits of agreement. Maybe this result should not be surprising as they are extracted from journals covering the same medical disciplines and the common underlying subgroup of concepts dominates the analysis.

Presenting a word cloud of the noun phrases out of a thousand most frequent occurrences that only occur in the journals listed in table 1 and not in the other journals assigned to the categories in this table is an alternative approach to make underlying differences visible (figure 1).

At first glance there seems to be some differences. The use of the term "translation research" appears to be paramount in the journals listed in table 1. Also more noun phrases are present in this word cloud that we would easily associate with journals on translational research and don't appear in the top frequency of the other medical journals. But this is only if we limit the noun phrases to the top 1000 most frequently used. If we compare all noun phrases and bring those terms to the front that are present in the journals listed in table 1 but not in other medical journals assigned to the same subject categories, very little to nothing remains of this difference. All the remaining noun phrases have a very low absolute frequency: 1 or 2. Even the term "translation research", of which we had high hopes, is then no longer present. We compared the relative frequency of the noun phrase "translational research" itself and found that is used significantly more frequent, by about 300 times as much, in the journals listed in table 1. The conclusion therefore is inevitable that the terminology used in both journal clusters is not so different after all, but that the journals listed in table 1 lean heavily on the use of the adjective "translational"; that in itself does not warrant an entirely different set of journals.

Discussion and ongoing work

Publications covering translational research cannot always easily be identified. In this paper the set of medical journals processed for the WoS with the adjective 'translational' in the title was used to study the characteristics of translational research. The characteristics of the papers in this journal set were compared with those in the more main stream clinical journals. Based on references in publications, the profile of the medical journals with the word 'translation' in the title is not different from the journals assigned to the same categories as the former. This is an indication that translational research does not rely more on basic knowledge than other clinical work.

A global analysis of the noun phrases in the abstracts leads to the same conclusion. The only peculiar observation is the abundant use of the adjective 'translational' in the journals covering as stated in their editorial mission and guidelines translational research.

The definition of 'translational research' itself seems to evolve reflecting the continuum from the results of basic research to its use in healthcare (Newly and Web, 2010). This process is hampered

by two bottlenecks: the translation of basis research results into clinical work and the translational of the results of clinical studies into clinical practice and healthcare (Cooksey Report, 2006). This paper is primarily focused on the study of the former. The latter is more complex and bibliometric tools are less appropriate.

In ongoing work we are comparing the key words assigned to the papers published in journals listed in table 1 and in the other journals assigned to the same categories and analyzing the institutional addresses in the byline of these journals' papers.

References

Altman DG & Bland JM (1983), Measurement in Medicine: the Analysis of Method Comparison Studies, *The Statistician*, 32, 307–317.

Butler, D. (2008), Translational Research: Crossing the Valley of Death, *Nature*, 453,840–842.

Cambrosio A., Keating P., Mercier S., Lewison G., Mogoutov A. (2006), Mapping the emergence and development of translational cancer research, *European Journal of Cancer*, 42, 3140–3148.

Cooksey D (2006), A review of UK health research funding, *UK*; *HMSO*

Jones DS, Cambrosio A, Mogoutov A (2011), Detection and characterization of translational research in cancer and cardiovascular medicine, *Journal of Translational Medicine*, 9, DOI: 10.1186/1479-5876-9-57

Luwel M. (2008) Translational Research: A New Research Discipline?, *10th International Conference on Science and Technology Indicators, Vienna, 17–20 September 2008*.

Newly D.E., Webb DJ (2010), Translational research: a priority for health and wealth, *Heart*, 96, 815–816.

Woolf, S.H.(2008),The meaning of Translational Research and Why it Matters, JAMA, 299:211–213.

Appendix

Table 1. Journals covering translation research. The first colon contains the journal name, the second the subject category/ies assigned to each journal, the third the first year the journal is processed for the WoS. In the fourth colon the rank of the journal is given compared to all journals assigned to the category; the rank 1 is given to the journal with the highest fraction of references to papers published in journals assigned to the basic categories. In the fifth colon the rank compared to the journals assigned only *to the corresponding category. In the last 3 colons the fraction of the references to journals assigned to the basic categories is given for the journal and the highest ranked journal in the two models.*

Journal name	Subject Category (1)	First year processed for WoS	Rank compared to all journals in (1) (2)	Rank compared to journals only in (1) (3)	Fraction of references in basic categories	Highest fraction of references in basic categories for journals in (2)	Highest fraction of references in basic categories for journals in (3)
Clinical & Translational Oncology	Oncology	2007	102 out of 194	44 out of 92	0,33	0,77	0,65
Cts-Clinical And Translational Science	Medicine, Research & Experimental	2008	56 out of 114	21 out of 46	0,48	0,85	0,85
Journal Of Cardiovascular Translational Research	Cardiovascular System	2008	7 out of 121	2 out of 67	0,51	0,60	0,57
	Medicine, Research & Experimental	2008	51 out of 114	19 out of 48	0,51	0,85	0,85
Journal Of Translational Medicine	Medicine, Research & Experimental	2005	61 out of 114	20 out of 46	0,47	0,85	0,85
Science Translational Medicine	Cell Biology	2009	144 out of 184	46 out of 50	0,55	0,96	0,94
	Medicine, Research & Experimental	2009	39 out of 114	12 out of 46	0,55	0,85	0,85
Translational Neuroscience	Neurosciences	2010	75 out of 239	28 out of 64	0,35	0,64	0,61
Translational Oncology	Oncology	2008	48 out of 194	20 out of 92	0,47	0,77	0,66
Translational Research	Medical Laboratory Technology	2006	21 out of 31	14 out of 19	0,45	0,85	0,85
	Medicine, General & Internal	2006	2 out of 162	1 out of 122	0,45	0,48	0,45
	Medicine, Research & Experimental	2006	62 out of 114	21 out of 64	0,45	0,85	0,85
Translational Stroke Research	Clinical Neurology	2010	13 out of 189	1 out of 45	0,34	0,55	0,34
	Neurosciences	2010	83 out of 239	33 out of 64	0,34	0,64	0,61

Table 2. Subject categories used to operationalized the concept 'basic research' for the analysis of the translation of the outcome of basic research into clinical work

Subject category	
Hematology	Computer Science Information Sys
Biochemistry Molecular Biology	Optics
Physiology	Mathematics Applied
Cell Biology	Materials Science Multidisciplin
Radiology Nuclear Medicine	Integrative Complementary Medici
Medicine Research Experimental	Computer Science Theory Methods
Endocrinology Metabolism	Chemistry Physical
Critical Care Medicine	Physics Multidisciplinary
Multidisciplinary Sciences	Mathematics Interdisciplinary Ap
Engineering Biomedical	Mechanics
Genetics Heredity	Physics Applied
Biophysics	Evolutionary Biology
Biology	Physics Fluids Plasmas
Medical Laboratory Technology	Polymer Science
Developmental Biology	Information Science Library Scie
Biotechnology Applied Microbiolo	Physics Atomic Molecular Chemica
Mathematical Computational Biolo	Instruments Instrumentation
Statistics Probability	Nanoscience Nanotechnology
Biochemical Research Methods	Engineering Multidisciplinary
Medical Informatics	Chemistry Inorganic Nuclear
Microbiology	Physics Condensed Matter
Acoustics	Engineering Mechanical
Chemistry Analytical	Engineering Environmental
Computer Science Interdisciplina	Computer Science Software Engine
Cell Tissue Engineering	Engineering Chemical
Materials Science Biomaterials	Electrochemistry
Chemistry Medicinal	Mathematics
Neuroimaging	Operations Research Management S
Chemistry Multidisciplinary	Construction Building Technology
Engineering Electrical Electroni	Metallurgy Metallurgical Enginee
Reproductive Biology	Mycology
Food Science Technology	Computer Science Cybernetics
Imaging Science Photographic Tec	Crystallography
Microscopy	Engineering Industrial
Chemistry Applied	Materials Science Characterizati
Physics Mathematical	Materials Science Coatings Films
Chemistry Organic	Physics Nuclear
Computer Science Artificial Inte	Engineering Manufacturing
Spectroscopy	Materials Science Composites
Nuclear Science Technology	Materials Science Textiles

Figure 1. Word cloud of the noun phrases out of a thousand most frequent occurrences that only occur in the journals listed in table 1.

[translational research] [clin trans sci] [dendritic cell] [translation] [technology] [translational psychiatry] [advance] [monocyte] [peripheral blood] [researcher] [discovery] [science translational medicine] [pbmc] [fibrosis] [promise] [peripheral blood mononuclear cell] [immunosorbent assay] [spain] [dcs] [epidermal growth factor receptor] [ovarian cancer] [clinical application] [gene therapy] [immunotherapy] [mirna] [normal tissue] [biology] [complete response] [expansion] [tumor tissue] [gemcitabine] [human cancer] [metastatic disease] [paclitaxel] [breast cancer patient] [chemokine] [resource] [toll] [docetaxel] [advanced disease] [clinical response] [crc] [ctl] [microrna] [neutrophil] [non small cell lung cancer] [plasma level] [recent advance] [viability] [better understanding] [host] [lymph node] [therapeutic potential] [validation] [adenocarcinoma] [adjuvant] [cardiac function] [clinical] [disease free survival] [doxorubicin] [gm csf] [hcv] [hepatocellular carcinoma] [immunity] [interferon] [myocardium] [solid tumor] [cancer therapy] [cardiomyocyte] [clinical research] [integration] [radiation therapy] [tissue microarray] [central role] [gene expression profile] [hypertrophic cardiomyopathy] [igf] [melanoma patient] [mir] [polymerase chain reaction] [potential biomarker] [regeneration] [schedule] [translational medicine] [autoimmune disease] [colon cancer] [consensus] [donor] [fluorouracil] [gastric cancer] [hcc] [hcm] [immunohistochemical analysis] [innovation] [lipopolysaccharide] [mesenchymal stem cell] [micrornas] [new approach] [new drug] [pancreatic cancer] [pilot study] [preclinical model] [prostate] [quantification] [regulatory t cell] [rodent] [stable disease] [t cell response] [therapeutic strategy] [tlr] [tumour cell] [advanced stage] [amplification] [blood vessel] [candidate] [caucasian] [cell survival] [centre] [chemotherapeutic agent] [clinical setting] [clinical use] [drug administration] [effective treatment] [epitope] [growth factor beta] [ihc] [immune system] [inflammatory cytokine] [inflammatory response] [insulin resistance] [melanoma cell] [murine model] [nuclear factor kappa b] [nude mouse] [progression free survival] [sera] [targeted therapy] [therapeutic efficacy] [therapeutic target] [thrombosis] [tolerance] [adjuvant chemotherapy] [autoimmunity] [cell therapy] [collagen] [early diagnosis] [hepatitis c virus] [human disease] [ifn] [interferon gamma] [irradiation] [major cause] [median overall survival] [messenger rna] [metabolite] [mononuclear cell] [national institute] [neutropenia] [new therapy] [progenitor cell] [proof] [rat model] [spectrum] [therapeutic agent] [treg] [tumor microenvironment] [tumor necrosis factor] [vasculature] [xenograft] [bevacizumab] [carcinogenesis] [carrier] [cell viability] [clinical benefit] [colleague] [combination therapy] [complexity] [cyclophosphamide] [fold increase] [g csf] [genetic alteration] [healthy donor] [heart disease] [hemoglobin] [high dose] [highlight] [hla] [immune cell] [iron] [mab] [median survival] [metastatic melanoma] [molecular basis] [oxaliplatin] [prognostic value] [sequencing] [solution] [staging] [sunitinib] [surgical resection] [taxane] [therapeutic] [therapeutic effect] [african american] [array] [arrhythmia] [cad] [cancer stem cell] [carboplatin] [cardiomyopathy] [clinical development] [collaboration] [glioma] [hdl] [her2] [local control] [methylation] [microenvironment] [mirnas] [modality] [murine] [neck cancer] [overview] [platinum] [primary endpoint] [primary tumour] [qrt pcr] [regenerative medicine] [regression] [stage iii] [staining] [subpopulation] [survival rate] [tnf] [vessel] [vitro study] [whole blood] [cd34] [cerebral ischemia] [colon] [first line treatment] [gene expression profiling] [glioblastoma] [immunofluorescence] [infiltration] [investigator] [ischemic stroke] [last decade] [manuscript] [mediator] [msc] [normal cell] [positron emission tomography] [possible role] [preclinical study] [preparation] [prognostic significance] [psa] [quantitative real time pcr] [real time] [responsiveness] [science] [small interfering rna] [stage ii] [standard treatment] [stem cell therapy] [useful tool] [albumin] [antitumor activity] [blood brain barrier] [cancer progression] [clinical parameter] [cytotoxic t lymphocyte] [erlotinib] [genetic] [healthy individual] [hepatocyte growth factor] [high expression] [hla class] [homeostasis] [immunization] [immunogenicity] [ldl] [leukemia] [lymph node metastasis] [mass spectrometry] [mixture] [monkey] [monotherapy] [mu g ml] [multidisciplinary team] [necrosis] [neoadjuvant chemotherapy] [neuroprotection] [new insight] [nk cell] [oncologist] [organization] [oxygen] [paraffin] [particle] [pfs] [primary tumor]

F1000 recommendations as a new data source for research evaluation: A comparison with citations[1]

Ludo Waltman waltmanlr@cwts.leidenuniv.nl / Centre for Science and Technology Studies, Leiden University (The Netherlands)

Rodrigo Costas rcostas@cwts.leidenuniv.nl / Centre for Science and Technology Studies, Leiden University (The Netherlands)

Abstract

F1000 is a post-publication peer review service for biological and medical research. F1000 aims to recommend important publications in the biomedical literature. In this study we present a comprehensive comparison between F1000 recommendations and citations. We find that about 2% of the publications in the biomedical literature receive at least one F1000 recommendation. Recommended publications on average receive 1.30 recommendations, and over 90% of the recommendations are given within half a year after a publication has appeared. There is a clear correlation between F1000 recommendations and citations, but the correlation is weaker than the correlation between journal impact and citations. More research is needed to identify the main reasons for differences between recommendations and citations in assessing the impact of publications.

Introduction

The recent introduction of the so-called 'altmetrics' (Priem & Hemminger, 2010; Priem, Piwowar, & Hemminger, 2012; Priem, Taraborelli, Groth, & Neylon, 2010) is leading to the development of new instruments for research evaluation, although altmetrics still needs to overcome important problems in order to become a robust and stable instrument for research evaluation (Wouters & Costas, 2012).

Among the various altmetrics tools, there is one that deserves special attention, particularly because of its innovative use of peer review. This is Faculty of 1000 (F1000), recently renamed as F1000Prime (see http://f1000.com/prime). F1000 is a commercial online post-publication peer review service for biological and medical research. It was launched in 2002[i] and so far it has

1 We would like to thank Omer Gazit, Jane Hunter, Ian Tarr, and Kathleen Wets from F1000 for providing us with the F1000 data set, for helping us with the interpretation of the data, and for providing feedback on an earlier version of this paper. The views expressed in this paper are those of the authors and are not necessarily shared by F1000.

collected reviews of over 100,000 biomedical publications. Reviews are produced by more than 5,000 peer-nominated researchers and clinicians (F1000 faculty members). Faculty members are requested to select the most important publications they read and to provide reviews of these publications. A review of a publication consists of a recommendation ('good', 'very good', or 'exceptional') along with an explanation of the strengths and possibly also the weaknesses of the publication. Faculty members can choose to review any primary research article from any journal, without being limited to recent publications or publications indexed in PubMed. From a research evaluation point of view, F1000 is a quite unique service, offering peer review judgments on individual publications in a large-scale, systematic fashion, with reviews being available to anyone with a subscription to the service.

In this paper, we present a large-scale analysis of F1000 recommendations, focusing on comparing recommendations with citations. Our analysis aims to provide insight into the potential value of F1000 recommendations for research evaluation purposes. We are interested to see to what extent recommendations correlate with citations and whether recommendations can be regarded as predictors of citations or highly cited publications. F1000 recommendations have been studied before (Allen, Jones, Dolby, Lynn, & Walport, 2009; Bornmann & Leydesdorff, 2013; Li & Thelwall, 2012; Medical Research Council, 2009; Mohammadi & Thelwall, in press; Priem et al., 2012; Wardle, 2010; Wets, Weedon, & Velterop, 2003), but earlier studies were based on relatively small data sets. In the present study, F1000 recommendations are analyzed in a much more comprehensive manner.

We note that this paper is a shortened version of a considerably more extensive working paper (Waltman & Costas, 2013). We refer to this working paper for the full details of our analysis.

Data and methodology

F1000 provided us with data on all 132,844 recommendations in their system. For each recommendation, we received a score (1 = 'good'; 2 = 'very good'; 3 = 'exceptional'), the date at which the recommendation was given, and some bibliographic data on the publication being recommended. Of the 132,844 records, 182 were excluded from our analysis[ii], thus leaving 132,662 recommendations. These recommendations relate to 102,360 unique publications, which means that the average number of recommendations per publication equals 1.30 (taking into account only publications with at least one recommendation).

We linked the publications in the F1000 data set to publications in Thomson Reuters' Web of Science (WoS) database, either by matching the digital object identifier (DOI) or by matching the journal title, volume number, issue number, and first author name. Overall, there are 95,385 publications (93%) for which a link could be established between the F1000 data set and the WoS database. The matched publications have been recommended 124,320 times[iii].

Our comparison between recommendations and citations is based on a more restricted data set including only publications from the period 2006–2009 and of the document types *article* and *review*. Also, we only take into account recommendations given in the year in which a publication appeared or in one of the next two years. There turned out to be 38,369 publications that satisfy our criteria and that have at least one recommendation. For each of these publications, we determined the 'microfield' to which the publication belongs in a newly developed publication-level classification system of science (Waltman & Van Eck, 2012). 38,327 publications were found to belong to 5,908 different microfields. The overall number of publications in these 5,908 microfields in the period 2006–2009 is 1,707,631. Our comparison between recommendations and citations is based on these 1.7 million publications, for which we counted citations received within a three-year citation window. Hence, citations were counted within the same time window as recommendations, so that we can make a fair comparison between the two. We also determined a journal citation score[iv] for each publication.

Results

We study 1.7 million publications from the period 2006–2009 of which 38,327 (2.2%) have at least one recommendation. Recommendations are typically given within half a year after a publication has appeared. As we show in the working paper version of this paper (Waltman & Costas, 2013), this is the case for more than 90% of all recommendations. We examine two ways in which recommendations and citations may relate to each other. Firstly, we analyze the relation between the highest recommendation a publication has received and the number of citations of the publication. Secondly, we analyze the relation between the total number of recommendations of a publication and its number of citations. In the latter case, no distinction is made between 'good', 'very good', and 'exceptional' recommendations. We also compare recommendations with journal citation scores (JCSs).

Table 1 reports the average number of citations and the average JCS of publications with a maximum recommendation score of 0, 1, 2, or 3. The table also provides 95% confidence intervals calculated using bootstrapping (e.g., Efron & Tibshirani, 1993).

Table 1. Average number of citations and average JCS of publications with a maximum recommendation score of 0 (no recommendation), 1 ('good'), 2 ('very good'), or 3 ('exceptional'). 95% confidence intervals are reported between brackets.

Max. recommendation score	No. of publications	Mean no. of citations	Mean journal citation score
0	1,669,304	7.2 [7.1, 7.2]	6.9 [6.9, 7.0]
1	22,862	20.7 [20.4, 21.1]	17.4 [17.2, 17.6]
2	12,838	37.6 [36.8, 38.6]	27.9 [27.5, 28.3]
3	2,627	68.6 [65.5, 72.3]	44.6 [43.7, 45.6]

Table 1 indicates that both the average number of citations per publication and the average JCS per publication increase with the maximum recommendation score of a publication. The effect is quite strong, especially for the average number of citations per publication. Recall that on average the publications in our analysis have been cited 7.7 times. As can be seen in Table 1, publications that have not been recommended are somewhat below this average, publications with a maximum recommendation score of 1 are more than 2.5 times above the average, and publications with a maximum recommendation score of 2 are almost 5 times above the average. Publications with a maximum recommendation score of 3 even tend to be cited almost 9 times more frequently than the average. Results similar to the ones reported in Table 1 are obtained when looking at the relation between the number of recommendations (rather than the maximum recommendation score) of a publication and publications' average number of citations and average JCS (Waltman & Costas, 2013).

Table 2 reports Pearson correlations between on the one hand publications' maximum recommendation score and number of recommendations and on the other hand publications' number of citations and JCS. Correlations obtained for the number of recommendations are slightly higher than those obtained for the maximum recommendation score, but the difference is very small. Correlations of recommendations with the JCS are higher than correlations of recommendations with the number of citations. Hence, in terms of the Pearson correlation, recommendations are more strongly related to the citation impact of the journal in which a publication has appeared than to the number of citations of a publication. We note that we also tested the effect of applying a logarithmic transformation to the number of citations of a publication. This turned out to yield lower correlations between recommendations and citations than the ones reported in Table 2.

Table 2. Pearson correlations of publications' maximum recommendation score and number of recommendations with publications' number of citations and JCS. 95% confidence intervals are reported between square brackets.

	No. of citations	**Journal citation score**
Max. recommendation score	0.24 [0.23, 0.26]	0.33 [0.33, 0.34]
No. of recommendations	0.26 [0.24, 0.28]	0.34 [0.33, 0.34]

Recommendations may be seen as an alternative to the citation impact of the journal in which a publication has appeared, since both recommendations and journal citation impact could potentially serve as a predictor of the number of citations of a publication.

An obvious question is whether for the purpose of predicting the number of citations of a publication recommendations may be more accurate than journal citation impact. Based on the Pearson correlation, the answer to this question is negative. The Pearson correlation between publications' JCS and their number of citations equals 0.52. This is much higher than the correla-

tions between recommendations and citations reported in Table 2. Hence, according to the Pearson correlation, predictions of citations based on JCSs are substantially more accurate than predictions of citations based on recommendations.

Given the fact that 97.8% of the publications in our analysis have not been recommended at all, it is perhaps not surprising that the correlation between recommendations and citations is much weaker than the correlation between JCSs and citations. Because of the low percentage of publications with recommendations, one could hypothesize that recommendations mainly indicate the most highly cited publications in the scientific literature. To test this idea, we identified the top 1% most highly cited publications (i.e., all publications with at least 58 citations) among the 1.7 million publications included in our analysis. We then examined to what extent recommendations and JCSs are able to distinguish between these highly cited publications and the other 99% of the publications.

We produced precision-recall curves for several approaches for identifying the top 1% most highly cited publications in our analysis (due to space limitations, we do not show the precision-recall results; see Waltman & Costas, 2013). Precision is defined as the number of highly cited publications in the selection divided by the total number of publications in the selection. Recall is defined as the number of highly cited publications in the selection divided by the total number of highly cited publications. Of the different approaches considered for identifying the top 1% most highly cited publications, we find that the approach based on JCSs performs much better than the approaches based on recommendations. Hence, it can be concluded that JCSs are substantially more accurate than recommendations not only for predicting citations in general but also for the more specific task of predicting the most highly cited publications.

It should be noted that our results may be sensitive to the selection of publications included in our analysis. We also calculated results based on a different selection of publications. Instead of our initial selection, we selected 1.1 million publications from 3,044 microfields with at least three F1000 recommended publications. The results turned out to be similar to the ones presented above, indicating that the results of our analysis have only a limited sensitivity to the selection of publications.

Discussion and conclusions

About 2% of the publications in the biological and medical sciences receive one or more recommendations from F1000 faculty members, depending of course on how exactly one chooses to delineate the biomedical literature. If a publication is recommended, the number of recommendations is usually small, with an average of 1.30 recommendations per publication.

In line with earlier studies (Bornmann & Leydesdorff, 2013; Li & Thelwall, 2012; Medical Research Council, 2009; Priem et al., 2012; Wardle, 2010), our analysis shows a clear correlation

between F1000 recommendations and citations. However, in order to qualify the strength of this correlation, it should be noted that the correlation between recommendations and citations is substantially weaker than the correlation between journal citation scores and citations.

In a sense, F1000 recommendations cannot be expected to correlate better than journal citation scores with citations, simply because about 98% of all biomedical publications do not have any recommendation at all. A more reasonable idea may be that recommendations predict highly cited publications. Our analysis shows that also from this point of view recommendations have a lower predictive power than journal citation scores. On the one hand, we do find that recommended publications tend to be cited quite a lot, with for instance half of the recommended publications belonging to the top 10% most highly cited publications in our analysis. On the other hand, however, we also find that many highly cited publications have not been recommended. For instance, almost three-quarter of the top 1% most highly cited publications have not been recommended.

From the research evaluation perspective, how should one interpret this relatively weak correlation between F1000 recommendations and highly cited publications? On the one hand, one could interpret this as an indication that F1000 fails to identify the most important publications in the biological and medical sciences (in line with the conclusion drawn by Wardle (2010) for the field of ecology). However, the relatively weak correlation could be also because recommendations and citations capture different types of impact, as suggested by Li and Thelwall (2012), who suggest that recommendations measure the 'quality of articles from an expert point of view' while citations measure 'research impact from an author point of view'. Thus, one would expect F1000 recommendations to sometimes identify important publications that remain unnoticed by citation analysis.

Future research may focus on providing more in-depth analyses of the reasons F1000 faculty members have to recommend a publication and of possible biases in F1000's peer-nomination system for selecting faculty members (as suggested by Wardle, 2010).

References

Allen, L., Jones, C., Dolby, K., Lynn, D., & Walport, M. (2009). Looking for landmarks: The role of expert review and bibliometric analysis in evaluating scientific publication outputs. *PLoS ONE*, 4(6), e5910.

Bornmann, L., & Leydesdorff, L. (2013). The validation of (advanced) bibliometric indicators through peer assessments: A comparative study using data from InCites and F1000. *Journal of Informetrics*, 7(2), 286–291.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. Chapman & Hall.

Li, X., & Thelwall, M. (2012). F1000, Mendeley and traditional bibliometric indicators. In *Proceedings of 17th International Conference on Science and Technology Indicators* (541–551).

Medical Research Council (2009). *MRC e-Val* 2009. Retrieved March 3, 2013, from www.mrc.ac.uk/Achievementsimpact/Outputsoutcomes/MRCe-Val2009/Publications/index.htm#P110_9319

Mohammadi, E., & Thelwall, M. (in press). Assessing non-standard article impact using F1000 labels. *Scientometrics*.

Priem, J., & Hemminger, B.H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social Web. *First Monday*, 15(7).

Priem, J., Piwowar, H.A., & Hemminger, B.M. (2012). *Altmetrics in the wild: Using social media to explore scholarly impact*. arXiv:1203.4745.

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). *Altmetrics: A manifesto*. Retrieved February 28, 2013, from http://altmetrics.org/manifesto/

Waltman, L., & Costas, R. (2013). *F1000 recommendations as a new data source for research evaluation: A comparison with citations*. arXiv:???.

Waltman, L., & Van Eck, N.J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, 63(12), 2378–2392.

Wardle, D.A. (2010). Do 'Faculty of 1000' (F1000) ratings of ecological publications serve as reasonable predictors of their future impact? *Ideas in Ecology and Evolution*, 3, 11–15.

Wets, K., Weedon, D., Velterop, J. (2003). Post-publication filtering and evaluation: Faculty of 1000. *Learned Publishing*, 16(4), 249–258.

Wouters, P., & Costas, R. (2012). *Users, narcissism and control: Tracking the impact of scholarly publications in the 21st century*. SURFfoundation. Retrieved February 28, 2013, from www.surf.nl/en/publicaties/pages/users_narcissism_control.aspx

---

i In 2002, F1000 was referred to as F1000 Biology. F1000 Medicine was launched in 2006. Later on, the two services were combined.

ii These are 'dissenting opinions' in which a faculty member indicates a disagreement with a recommendation given by another faculty member.

iii Our procedure for matching publications is quite conservative. We therefore expect there to be almost no incorrect matches.

iv The journal citation score of a publication in journal X equals the average number of citations received by all publications in journal X in the period 2006–2009. In the calculation of journal citation scores, only articles and reviews were considered. Citations were also counted within a three-year citation window.

The diffusion of scientific literature on the web[1]

Xianwen Wang xianwenwang@dlut.edu.cn / WISE Lab, Dalian University of Technology, Dalian 116085 (China)

Wenli Mao* maowenli.1990@163.com / WISE Lab, Dalian University of Technology, Dalian 116085 (China)

Chunbo Zhang* zwchunbo@163.com / WISE Lab, Dalian University of Technology, Dalian 116085 (China)

Zeyuan Liu liuzy@vip.163.com / WISE Lab, Dalian University of Technology, Dalian 116085 (China)

* Contributed equally as second authors.

Conference theme and subtheme

Knowledge transfer, knowledge exchange and specific issues relating to translational research

Abstract

Combined with open access, the rapid development of the social web makes the dissemination of scientific literature on the web much easier, faster and broader than ever before. We collected the tweet counts of 231,707 papers submitted to arXiv from 2010 to 2012. About 22.29% of arXiv papers have received at least 1 Tweets. The number of Tweets for arXiv papers increases significantly month by month. Through individual cases, we specifically study the diffusing process and mechanism for scientific literature in web. Analyzing the diffusing route based on blog media, social media and traditional media reveals progressive inducing and feedbacking among blog media, social media and traditional media generate complicated interactions, which strengthen the diffusion of scientific literature continuously.

Introduction

When scientists publish their findings, they hope their work will be read by as many people as possible. Before the computer age, academic articles were printed and published in paper-based scientific journals. When a new paper is published, people read the paper from the print journal. With the advent of the internet, another common form of access is that papers are read directly from electronic and on-line journals.

1 This work was supported by the project of "Social Science Foundation of China"(10CZX011) and the project of "Fundamental Research Funds for the Central Universities" (DUT12RW309).

Do scientists often communicate with one another about their research and new findings? Do they simply read each other's work without any personal communication? There does exist an invisible college with a network of communications among scientists (Crane, 1972, 1989; Wagner, 2008).

How do scientific literatures diffusing in the academic community? Face-to-face communication among researchers is one important way. Other communications may happen by phone and through letters among researchers. In the era of internet, a more common way is that scientists may forward the newly published papers to colleagues from their mailing list.

Nowadays, with the fast development of social web, every day, every second, much information is shared in social media. For example, as early as in October, 2012, there were 500 million Tweets per day. The information also includes scientific information from academic papers. The social web provides an open and instant platform to discuss and distribute academic papers for scientists. Diffusing of scientific literature in social media makes the invisible college broader.

As new metrics based on the social web, which aims to make a real-time analysis of the scholarly impact of new research findings, altmetrics have attracted much attention (Bar-Ilan et al., 2012; Priem, Groth, & Taraborelli, 2012).

Combining with the social web, open access is another important factor that accelerates the diffusion of scientific papers. There are controversies about the effects of open access and citation (Craig, Plume, McVeigh, Pringle, & Amin, 2007; Davis, Lewenstein, Simon, Booth, & Connolly, 2008). Compared with toll access literatures, open access publishing may get significantly more downloads and reach more readers (Davis, 2011; Eysenbach, 2006). Shuai et. al (Shuai, Pepe, Bollen, 2012) examined the relationship between article downloads, Twitter mentions and academic citations of arXiv submissions from October 2010 to May 2011. Their found that Twitter mentions were significantly correlated with arXiv downloads and early citations, especially for highly mentioned articles. In Shuai et. al's study, Tweets data was collected from Twitter Gardenhose. According to them, their data "represents roughly 10% of the total Tweets from public time line through random sampling", only "10%" random sampling could have great bias for the results.

Diffusion of scientific paper in the social web

Here we also analyze arXiv data. As the first open access database of electronic preprints of scientific papers, arXiv has great influence in the academic community. In the single year of 2012, the number of paper archived in arXiv reached 84603. In this study, we do a thorough survey of Tweets of arXiv papers one by one from January, 2010 to December, 2012. All the data are harvested in two days of March 5–6, 2013. We use these two days as a reference to measure the number of Tweets irrespective of the submission month of arXiv papers. For example, for the

submission of arxiv.org/abs/1001.0001, which was published on January 1, 2010, all the Twitter mentions generated during the following three years and two months, until March 5, 2013 are considered. Recent submissions would have had less time to accumulate Tweets than older ones. However, even for the submission of arxiv.org/abs/1212.6974, which is the last one in 2012, the time span of two months is long enough to get most of its expected Tweets. Our data indicate that submissions get most of their expected Tweets within a short period (no more than one month) after the publication date.

Total Tweets for arXiv papers

Twitter's default search only goes back one week, so we conduct search in Topsy.com. We search the number of Tweets of each arXiv paper one by one. The total number of articles archived in arXiv from January 1, 2010 to December 31, 2012 was 231,307, so 231,307 searches were conducted in Topsy.com. We designed a SQL database to analysis the harvested data. When processing the data, we excluded the Tweets from the official Tweet accounts of arXiv, i.e. @QuantumPapers, @PhysicsPaper, and @AstroPHYPapers, etc. Because Tweets from these accounts are simply introduction to all newly arXiv submissions.

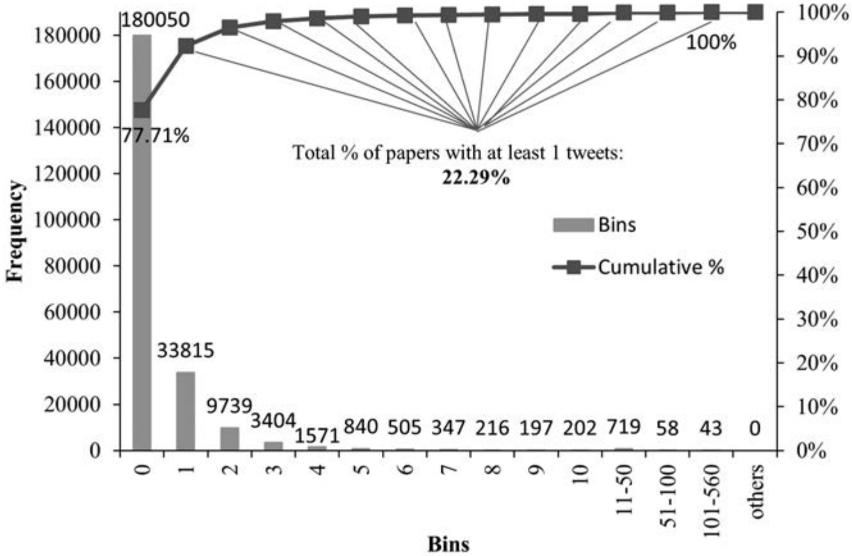
Table 1 shows the general distribution of Tweets for all arXiv articles. 180,050 papers never received any Tweets, which accounts for 77.71% of all papers, only 43 papers received more than 100 Tweets.

Table 1. Distribution of Tweets for arXiv papers

Range of Tweets	Number of articles	%
0	180,050	77.71%
1	33,815	14.59%
2	9739	4.20%
3	3404	1.47%
4	1571	0.68%
5	840	0.36%
6	505	0.22%
7	347	0.15%
8	216	0.09%
9	197	0.09%
10	202	0.09%
11–50	719	0.31%
51–100	58	0.03%
101–560	43	0.02%
Total	231,707	100%

Figure 1 shows the Tweet distribution histogram for arXiv papers. The thirteen ranges containing papers with at least 1 Tweet account for 22.29% of all arXiv papers.

Figure 1. Histogram of Tweets of arXiv papers

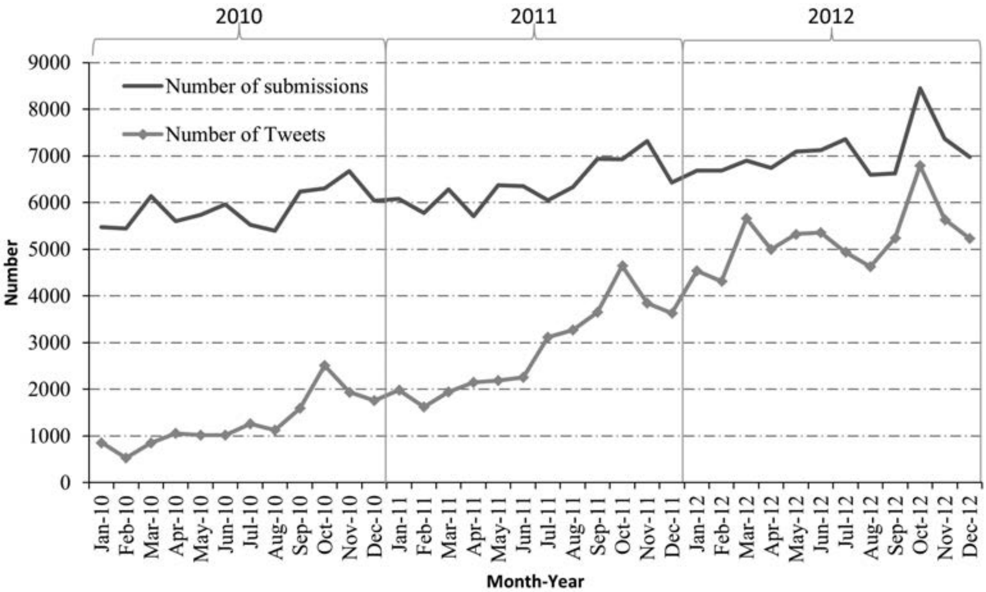


Monthly Tweet trends for arXiv papers

Figure 2 shows the monthly trends for arXiv submissions and Tweets. As the above curve shows, the number of papers submitted monthly is rather stable. In 2010, the curve fluctuates around a level of 6000. In 2012, authors submitted roughly 7000 articles every month, except in October, 2012, when there was a burst to 8450.

The curve below shows the monthly tweeting trend for arXiv articles. We see an obvious upward trend from 2010 to 2012. The 5471 papers submitted in January 2010 received only 838 Tweets. For October 2012, the 8450 submissions received 6341 Tweets.

Figure 2. Monthly submissions and Tweets for arXiv papers



Case Study of specific web diffusion process

Top articles with high attention in the social web

According to two recent ground-breaking summaries of the top 10 articles receiving the most online attention in the entire year of 2012 based on altmetrics data (Liu, 2012; Van Noorden, 2012), as Table 2 lists.

Table 2. Top 10 articles ranked by Altmetric data in 2012

Rank	Publish year	Blog media	Social networking media			Academic citation
		Science blog	Twitter	Facebook	Mendeley	Google scholar
1	2009	2	2223	0	14	13
2	2012	4	2170	134	70	3
3	2011	5	1647	43	64	10
4	2012	10	1595	53	97	17
5	1996	2	1419	91	7	203
6	2011	10	1243	20	166	42
7	2012	31	1103	108	330	23
8	2012	2	1125	125	9	0
9	2012	6	907	91	42	2
10	2012	5	924	13	26	4

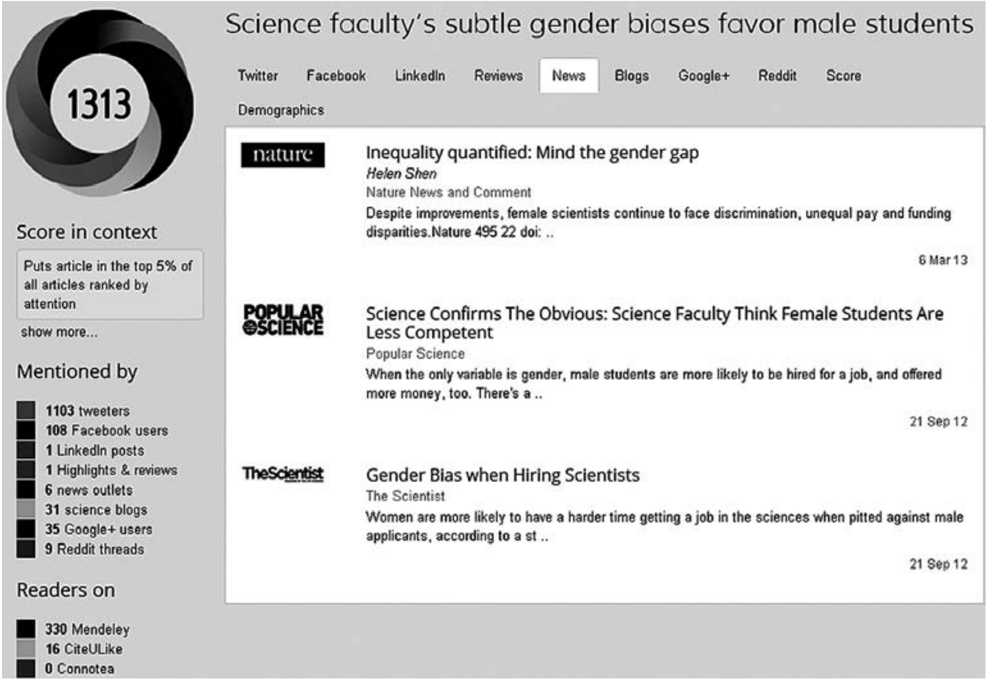
… Continuation Table 2

Rank	Publish year	Blog media	Social networking media			Academic citation
		Science blog	Twitter	Facebook	Mendeley	Google scholar
1	Food for thought. What you eat depends on your sex and eating companions					
2	The biological impacts of the Fukushima nuclear accident on the pale grass blue butterfly					
3	Bright Minds and Dark Attitudes Lower Cognitive Ability Predicts Greater Prejudice Through Right-Wing Ideology and Low Intergroup Contact					
4	Association of Coffee Drinking with Total and Cause-Specific Mortality					
5	Rape-related pregnancy: Estimates and descriptive characteristics from a national sample of women					
6	Higher social class predicts increased unethical behavior					
7	Science faculty's subtle gender biases favor male students					
8	Unilateral Dermatoheliosis					
9	Measuring the Evolution of Contemporary Western Popular Music					
10	Classic Nintendo Games are (NP-)Hard					

As shown in Table 2, most of the highly focused papers are published in 2012, with only two articles published in 2011, and one in 1996.

With altmetric methods and tools, altmetrics score can be calculated conveniently and quickly online, i.e. www.altmetric.com. As Figure 3 shows how the score is derived from http://www.altmetric.com/. However, do there exist any interactions among different media types? And what is the diffusion process and mechanism for scientific papers on the web? We need to explore this more deeply.

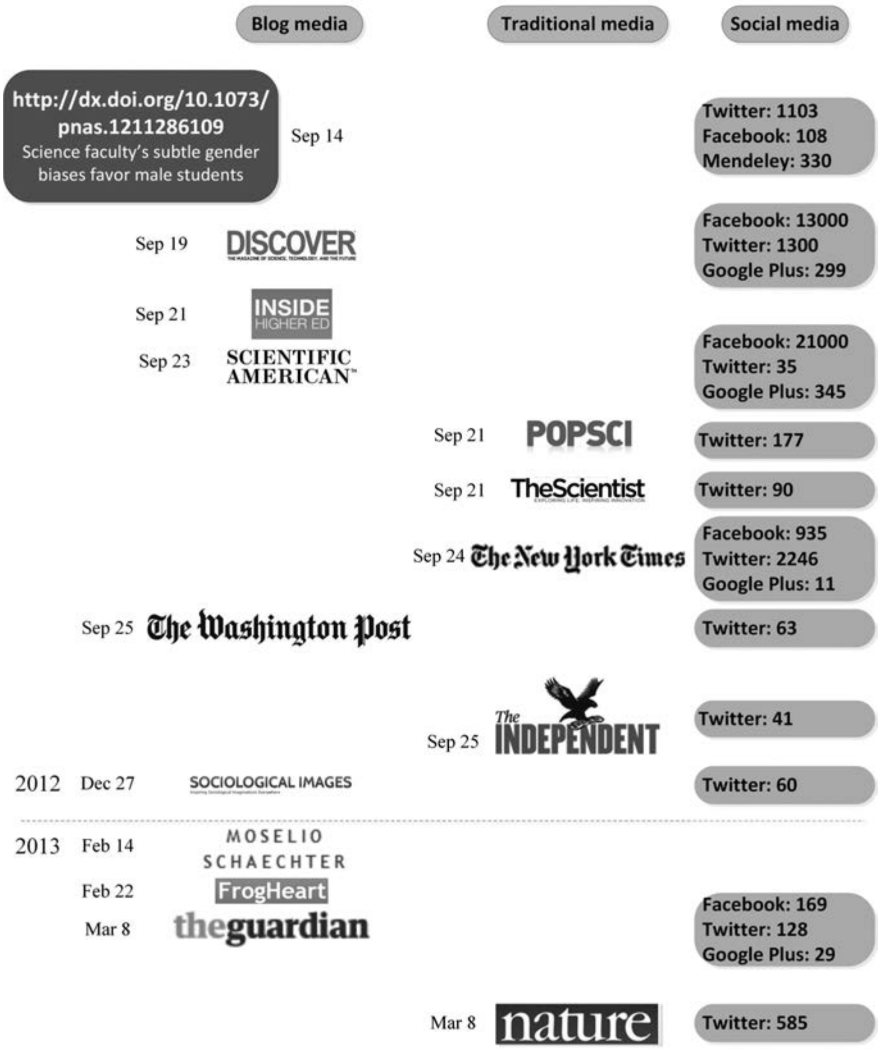
Figure 3. Screenshot of Altmetrics page for paper #7 listed in Table 2



Case I

How does a scientific paper diffuse on the web? Here we choose two cases to explore the detailed diffusion process. The first is paper 7 in Table 2, "Science faculty's subtle gender biases favor male students", published in the *Proceedings of the National Academy of Sciences*. The paper revealed hidden biases of research university faculties in favour of male students, with female students considered less competent, less hireable, and offered a significantly lower salary on average (Moss-Racusin, Dovidio, Brescoll, Graham & Handelsman, 2012). The paper prompted many discussions on the web. "Judging from the online attention, the article appeared to have reached its intended audience" (Liu, 2012). In Figure 4, for each column, we sort the media that refer to this paper according to the publishing date. Media are classified into three types, blog media, traditional media and social media. Blogs are sometimes considered a kind of social medium. However, in this study, we separate blogs from other social networking sites, i.e. Twitter, Facebook.

Figure 4. Diffusion of paper 7 on the web

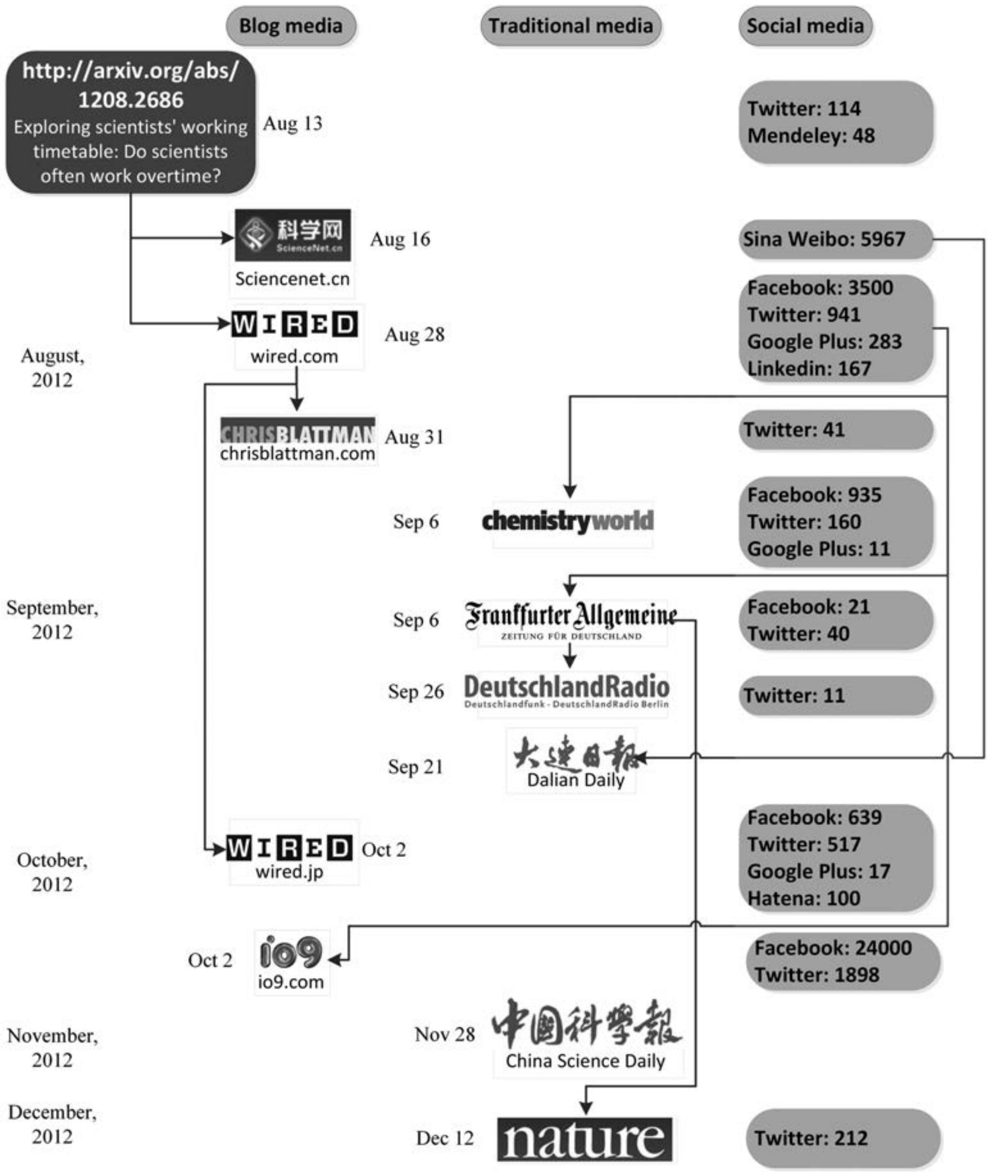


Case II

The second case is a paper published by ourselves. The paper "Exploring scientists' working timetable: Do scientists often work overtime?" was published in *Journal of Informetrics* on August 8, 2012 (Wang et al., 2012). On August 13, we submitted the preprint to arXiv. During the following months, more than 20 media, including newspapers, magazines, blogs referred to this article, including *Frankfurter Allgemeine Zeitung* (Frankfurt General Newspaper) (Anderl, 2012), *Deutschland Radio* from Germany (Böddeker, 2012), and *Nature* from UK (Schiermeier, 2012), etc. Figure 6 lists several representative media and shows the diffusion route of this paper on the web.

Compared to other papers, it's easy for us to analyze the dissemination process for our own paper, so the diffusion route is illustrated in Figure 5. The links with arrows demonstrate the source of the information.

Figure 5. Diffusion route of paper Wang et al. 2012 on the web

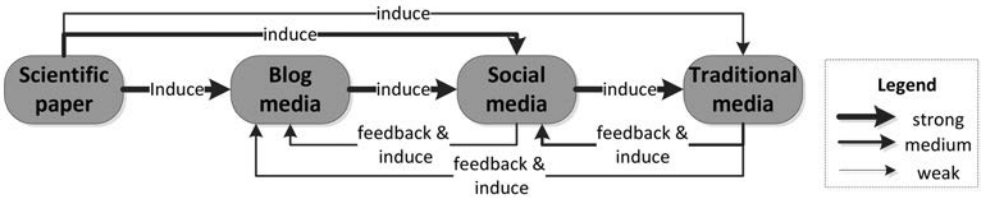


Wide diffusion of this paper on the web also brought it many downloads. According to the reports of SciVerse ScienceDirect's Top 25 Hottest Articles, this paper had the most downloads and ranked first during the whole year of 2012 in all papers published in *Journal of Informetrics* (http://top25.sciencedirect.com/subject/social-sciences/23/journal/journal-of-informetrics/17511577/archive/42/).

Conceptual Diffusion Model

The diffusion process of the two cases are summarized with a simple conceptual model, as shown in Figure 6. Line thickness is proportional to strength of influence.

Figure 6. A conceptual model of scientific paper diffusion on the web



When a new paper published, it is usually disseminated in the academic field first. Some scholars, particularly peers in the same or similar discipline, may review and recommend the paper in their personal blog. As Figure 7 shows, there is a direct link from scientific papers to blog media. Meanwhile, several factors, such as an eye-catching title and interest story, could support these blog articles, attracting others' attention and consequently pushing the scientific paper into the social web. Through various social media tools, the paper may draw increasing attention and become a hot topic in a wider community. The wide remit of the social web makes it possible to get the attention of reporters and journalists. Reports from traditional media may even help move the research toward a topic of social concern.

Alternatively, the paper may immediately induce comment and discussion in the social web after publication. On one hand, an increasing number of journals have provided direct connections to make published papers link directly to social networking sites. On the other hand, when readers catch sight of papers they are interested in, they may have no time to post long comments in a blog, but it is very convenient for them to share, forward and bookmark the paper in short text using social media tools.

It's not easy to get immediate attentions from traditional media for newly published scientific papers, except if the paper is published in the most prestigious multidisciplinary academic jour-

nals, i.e. *Nature*, *Science* and *PNAS*, or great discoveries published by famous scientists. Otherwise, it is very difficult to have a direct link from scientific paper to traditional media.

Although in much weaker immediate contact with papers, traditional media have an obvious advantage over other media for the influence of reporting. Based on the strict journalistic process and professional quality, the public prefers to believe in the authenticity and value of news in traditional media. Both blog media and social media belong to "we media", whereas traditional media are real "mass media". Therefore, a report from the traditional media can be viewed as setting an agenda to some extent and can spark public attention, and this applies also to scientific papers.

In the opposite direction, there are also feedback links among different media types. For the feedback link of social media to traditional media, because traditional media have broad audiences and social influence, most of them have official social media accounts. It's easy to understand that social media users respond a lot to scientific reports published in traditional media. For the feedback from blog media to social media, it could be that, for many social networking sites, the length of users' comments is restricted. i.e. in Twitter, the length is restricted to up to 140 characters. People may feel that Tweets have not given full expression to their views, so they may also like to post a longer comment in their blog media, as Figure 7 shows.

All in all, the loops formed by the inducing and feedbacking link progressively strengthen the diffusion of scientific literature continuously in web.

Discussion

The social web undoubtedly provides scientists with an open and instant platform to discuss and disseminate academic papers. Combining with open access, the social web helps spread research findings on the web much more easily, rapidly and broadly than ever before. This has also changed the communication and evaluation of scientific papers. However, diffusion of scientific papers on the web does have its unique characteristics and rules.

Compared with complicated studies, easier to understand but interesting findings, may get more buzz on the social web.

In the diffusion process for scientific papers on the web, different kinds of media play different roles. Progressive inducing and feedbacking links among blog media, social media and traditional media form complicated interactive loops, which strengthen the diffusion of scientific literature continuously.

A limitation of this study is that only two cases are considered to propose the diffusion model. In the future, with more case studies and more empirical evidence, we hope to summarize the diffusion patterns and mechanisms from a more quantitative prospect.

References

Anderl, V. S. (2012). Freizeit? Wissenschaftler haben Freizeit? (Leisure? Scientists have free time?). Retrieved December 28, 2012, from http://www.faz.net/aktuell/feuilleton/forschung-und-lehre/arbeitszeit-von-forschern-freizeit-wissenschaftler-haben-freizeit-11879056.html

Arbesman, S. (2012). The Results Are in: Scientists Are Workaholics. *www.wired.com*. Retrieved December 25, 2012, from http://www.wired.com/wiredscience/2012/08/the-results-are-in-scientists-are-workaholics/

Böddeker, V. M. (2012). Wenn Forscher zu lange forschen – Neue Studie deutet auf viele Überstunden und Wochenendarbeit hin (When researchers work too long- new study reveals a lot of overtime and weekend work). Retrieved December 26, 2012, from http://www.dradio.de/dlf/sendungen/forschak/1877153/

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). Beyond citations: Scholars' visibility on the social Web. *arXiv preprint arXiv*:1205.5611.

Craig, I.D., Plume, A.M., McVeigh, M.E., Pringle, J., & Amin, M. (2007). Do open access articles have greater citation impact?: a critical review of the literature. *Journal of Informetrics*, 1(3), 239–248.

Crane, D. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities*. Chicago: University of Chicago Press.

Crane, D. (1989). How scientists communicate–A citation classic commentary on invisible-colleges–Diffusion of knowledge in scientific communities by Crane, D. *Current Contents/Arts & Humanities*, 22, 14–14.

Davis, P. M. (2011). Open access, readership, citations: a randomized controlled trial of scientific journal publishing. *The FASEB Journal*, 25(7), 2129–2134. doi: 10.1096/fj.11-183988

Eysenbach, G. (2006). Citation advantage of open access articles. *Plos Biology*, 4(5), 692–698. doi: 10.1371/journal.pbio.0040157

Liu, J. (2012). Interactions: The Year in Conversations. *http://altmetric.com/blog/interactions-the-year-in-conversations/*. Retrieved January 15, 2013, from http://altmetric.com/blog/interactions-the-year-in-conversations/

Moss-Racusin, Corinne A., Dovidio, John F., Brescoll, Victoria L., Graham, Mark J., & Handelsman, Jo. (2012). *Science faculty's subtle gender biases favor male students*. Proceedings of the National Academy of Sciences. doi: 10.1073/pnas.1211286109

Priem, J., Groth, P., & Taraborelli, D. (2012). The Altmetrics Collection. *PloS one*, 7(11), e48753.

Schiermeier, Q. (2012). Lab life: Balancing act. *Nature*, 492(7428), 299–300.

Shuai X, Pepe A, Bollen J (2012) How the Scientific Community Reacts to Newly Submitted Preprints: Article Downloads, Twitter Mentions, and Citations. *PLoS ONE* 7(11): e47523. doi:10.1371/journal.pone.0047523

Van Noorden, R. (2012). What were the top papers of 2012 on social media? *Nature*. Retrieved January 15, 2013, from http://blogs.nature.com/news/2012/12/what-were-the-top-papers-of-2012-on-social-media.html

Wagner, C.S. (2008). *The new invisible college: Science for development*: Brookings Inst Press.

Wang, X., Xu, S., Peng, L., Wang, Z., Wang, C., Zhang, C., & Wang, X. (2012). Exploring scientists' working timetable: Do scientists often work overtime? *Journal of Informetrics*, 6(4), 655–660.

The discovery of 'introns': analysis of the science-technology interface

J. J. (Jos) Winnink winninkjj@cwts.leidenuniv.nl / Leiden University, Centre for Science and Technology Studies (CWTS), Wassenaarseweg 62A, 2333 AL Leiden (the Netherlands)/ Leiden University Dual PhD Centre The Hague, P.O. Box 13228, 2501 EE The Hague (the Netherlands) / NL Patent Office, P.O. Box 10366, 2501 HJ Den Haag (the Netherlands)

Robert J. W. Tijssen tijssen@cwts.leidenuniv.nl / Leiden University, Centre for Science and Technology Studies (CWTS), Wassenaarseweg 62A, 2333 AL Leiden (the Netherlands) / Leiden University Dual PhD Centre The Hague, P.O. Box 13228, 2501 EE The Hague (the Netherlands)

Anthony F. J. van Raan vanraan@cwts.leidenuniv.nl / Leiden University, Centre for Science and Technology Studies (CWTS), Wassenaarseweg 62A, 2333 AL Leiden (the Netherlands)

Abstract

The study addresses visible clues and empirical data that can be used to discover early stage breakthroughs in science that evolve into new technological developments. We analyse bibliographical information from scholarly publications, patent publications, and links between patent and scholarly publications. Our overall goal is to develop an analytical methodology for pinpointing the stage in which fundamental discoveries occur. This particular case study focuses on the discovery of 'introns' in chromosomes. We discuss two breakthrough events related to this discovery, classify them according to the 'Cha-Cha-Cha' theory, and show that the second breakthrough reflects a 'phase shift' where the scientific discovery moves into the first stages of technological development.

Introduction

Our longitudinal bibliometric analyses of 'breakthrough processes' aims at obtaining better insight into general patterns that characterise 'revolutionary' R&D dynamics (Winnink & Tijssen, 2011; Winnink 2012). We use the 'Cha-Cha-Cha' theory (Koshland, 2009) to classify breakthrough discoveries into distinctly different types. A short description of the 'Cha-Cha-Cha' theory is given in the next paragraph. The subject of this case study is the discovery of so-called 'introns'[i]. We categorise this discovery as a 'Challenge' because it can be characterised as posing a challenge that required a major 'paradigm shifting' adaptation of previous theories to explain new experimental observations (rather than solving an 'obvious' scientific problem).

The underpinning scientific discoveries were published in 1977 (see for instance Gelinas and Roberts, 1977; Sharp, 1993; Roberts, 1993) and revealed that chromosomes comprise of a mosaic

of 'exons', used to form new copies of genes, and of 'introns' that appeared to be non-functional 'interspaces'. These discoveries proved the assumption 'chromosomes for prokaryotic[ii] and eukaryotic[iii] organisms are similar in structure' to be wrong. The 1993 Nobel Prize in Physiology or Medicine was awarded to Philip A. Sharp and Richard J. Roberts for their pioneering work on 'introns'. Sharp (1993) describes in his Nobel Prize lecture the 'intron' problem as follows:

'By the late 70s the physical structure of a gene was firmly established from work in bacteria. The sequences of the gene, the RNA and the protein were colinearly organized and expressed. Since the science of genetics suggested that genes in eukaryotic organisms behaved similarly to those of prokaryotic organisms, it was naturally assumed that this bacterial gene structure was universal. It followed that if the gene structure was the same, then the mechanisms of regulation were probably very similar, and thus what was true of a bacterium would be true of an elephant. However, many descriptive biochemical aspects of the genetic material and its expression in cells with nuclei suggested that the simple molecular biology of gene expression in bacteria might not be universal'

The two main research questions we address in this study relate to how these revolutionary breakthroughs can be identified and monitored over time:

– "Can such a breakthrough, involving a paradigm shift, be identified and characterized in terms of bibliometric variables and indicators?"
– "Can one identify, bibliometrically, the stage in the development process where scientific knowledge is being used for work on science-based technologies?"

Conceptual framework

Our analytical approach builds on 'Cha-Cha-Cha'=theory by Koshland (2009), who's work on discoveries

'In looking back on centuries of scientific discoveries, however, a pattern emerges which suggests that they fall into three categories – Charge, Challenge, and Chance – that combine into a "Cha-Cha-Cha" Theory of Scientific Discovery.',

resulted in the following general typology that differentiates discoveries based on their nature:

'"Charge" discoveries solve problems that are quite obvious – cure heart disease, understand the movement of stars in the sky – but in which the way to solve the problem is not so clear.

"Challenge" discoveries are a response to an accumulation of facts or concepts that are unexplained by or incongruous with scientific theories of the time. The discoverer perceives that a new concept or a new theory is required to pull all the phenomena into one coherent whole.

"Chance" discoveries are those that are often called serendipitous[iv] *and which Louis Pasteur felt favoured "the prepared mind." In this category are the instances of a chance event that the ready mind recognizes as important and then explains to other scientists.'*

Charge discoveries can be considered 'normal science' in Kuhn's (1962) terminology. Challenge and Chance discoveries can be seen, within this particular context, as 'revolutionary science'.

Further work by Hollingsworth (2008) defines a 'major breakthrough' or a 'discovery' as:

'a finding or process, often preceded by numerous small advances, which leads to a new way of thinking about a problem.'

which is essentially equivalent to Koshland's 'Challenge' discovery, whereas 'small advances' can be seen as a 'Charge'. Hollingsworth's definition does not address 'Chance' discoveries as they are serendipitous and therefore have no precursory research directly linked to the discovery.

This discovery-oriented approach of describing knowledge creation dynamics differs from philosophy of science 'systems level' approaches, such as Kuhn's (1962) distinction between 'normal' and 'revolutionary' science, the latter being characterized by 'paradigm shifts'. Paradigm shifts start as a 'localized' diversion of an existing theoretical framework to explain observations that conflict with the theory. The consequence is that from a relatively small nucleus the new theory gains support of scholars resulting in growing numbers of publications. Paradigm shifts evolve gradually (Isnard and Zeeman, 1976) caused by the fact that scientists are reluctant to switch theories and concepts. During a paradigm shift increasing numbers of unique authors, institutions, journals, journal categories, and countries, where institutions are located, on publications related to the new paradigm can be observed.

Related studies, such as Andersen, Barker, and Chen (2006) focus on the cognitive structure of scientific revolutions and the changes in concepts during paradigm shifts. Worall (2003) discusses the various views on theory changes in scientific progress, whereas Isnard and Zeeman (1976) use catastrophe theory to explain the mechanisms behind the switching of views within societies.

Methodology

The various types of discoveries should manifest themselves by different types of events that can be externally detected through 'bibliographical signals'. The solution for a long-standing scientific problem, for instance producing free standing graphene (Winnink, 2012), can lead to an impulse-like sudden increase in the number of research publications. To select relevant scholarly publications we used the topic 'intron' in the on-line version of Thomson-Reuters Web-of-Science database (WoS). This search resulted in 31408 publications, articles and conference proceedings, for the period 1980–2012.

Patent publications were gathered by searching the EPO Worldwide Patent Statistical Database (PATSTAT), October 2011 version, for the terms “intron” and “non coding DNA” in titles and abstracts this resulted in 1,284 patent publications in 677 patent families for the period 1984–2012. We group patent publications describing the same invention in ‘simple patent families’ to prevent double counting[v].

The data set contains intron-related publications from 1980 (Figure 1) onwards. The figure shows a very significant increase for scholarly publications in 1990–1991. To link scientific discoveries and technology we use the citing-cited links between patent-publications and scholarly publications. These links are called non-patent-literature references (NPL).

Results

General trends

Figure 1 shows, on a logarithmic scale, the trends for the numbers of scholarly publications and patent applications. The sudden increase in the number of scholarly publications from 1990 to 1991 stands out. We show that there are two breakthroughs of different type involved. One is the discovery of ‘introns’ in 1977, a ‘challenge’ breakthrough. The second is a ‘charge’ breakthrough marked by the sudden increase in intron-related scholarly publications from 1990 to 1991. The figure also depicts a continuous increase in patent applications from 1990 onwards until reaching a plateau level in 1998. The number of scholarly publications reaches an output plateau level at about the same time.

Figure 1. Trends in scholarly publications and patent applications

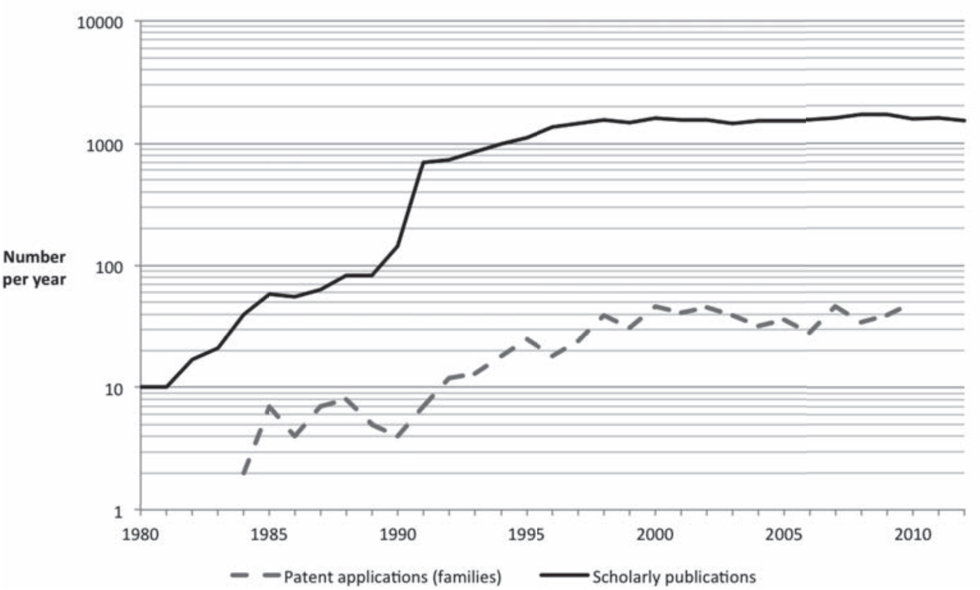
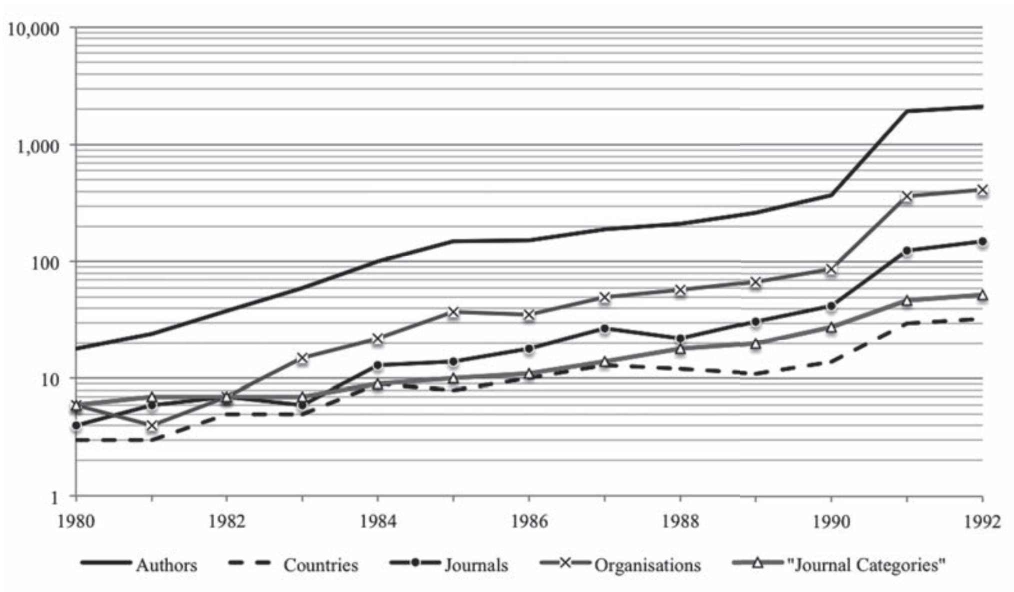


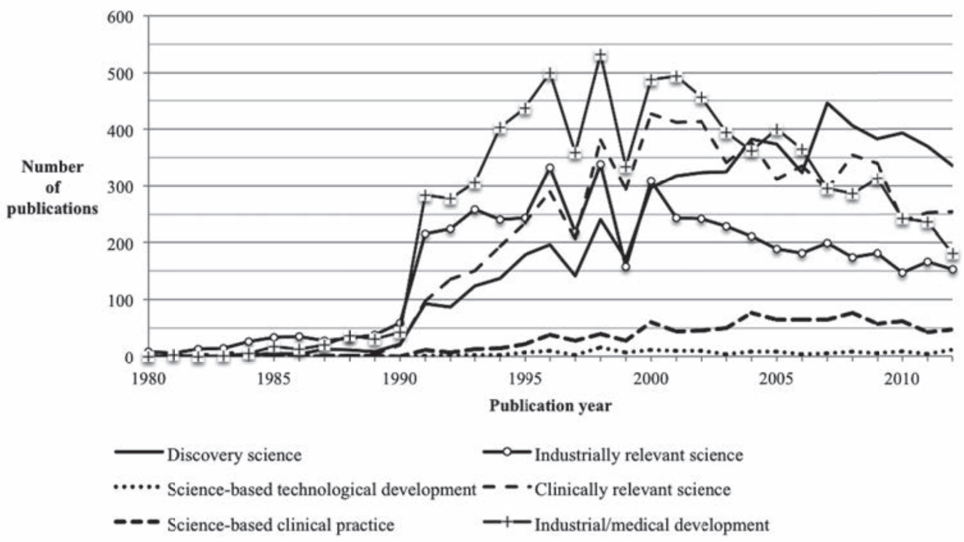
Figure 1 also reveals a low number of patent applications before 1992 indicating that intron related scientific knowledge was not yet used extensively in the development of new technologies. Figure 2 exhibits annual trends for the numbers of unique authors, journals, journal categories, organisations, and the number of countries where organisations that are 'active' in the intron-field at a certain moment are located; we presume that an actor that became active stays active. All these variables show a simultaneous and steady increase from 1980 onwards indicating that the theoretical concepts that form the basis for the selected scholarly publications spreads among the scientific community. These variables show the same 'phase shift' from 1990 to 1991 as in Figure 1.

Figure 2. Frequency counts of unique authors, journals, journal categories, countries, and organisations that produced intron-related scholarly publications.



In an attempt to explain the sudden increase 1990–1991 we classified the scholarly publications[vi] into six different categories, each encompassing a different type of science according to the cognitive-institutional environment in which the research was done (Tijssen, 2010). We could classify 85% of the publications in the document set. Figure 3 shows the results. The sudden increase is especially marked in 'application-oriented' categories 'Industrial/medical development', 'Clinically relevant science' and 'Industrially relevant science'. Our checks of the *Web of Science* indicate that this increase is not caused by new journals that were introduced into this bibliographical database. Although our inspection of the author affiliate addresses indicates that new institutional contributors (universities, research institutes, or other organisations) did enter the intron-field in 1991, the rise of publication output occurs mainly within incumbent institutes, especially the US universities.

Figure 3. Trends in scholarly publications differentiated by type of science



We constructed two subsets of scholarly publications; one for 1990 and one for 1991. Using the VOSviewer[vii] we created two maps (Figures 4 and 5) showing the spatial configuration of keywords and phrases as they occurred in the titles and abstracts[viii] of publications. Each of these content identifiers in these maps occurred in at least three publications. The same clustering and mapping algorithms were used to assure that the visible differences in the configuration reflect changes in content and relational structures.

In general, maps using terms from relevant documents to illustrate the evolution of a research area in science look for successive years quite similar. The differences between the two maps reflect considerable changes from 1990 to 1991 (Figure 3). Close inspection of the keywords and phrases in the maps shows a shift towards application of scientific knowledge for technological development.

Figure 4. Term map for scholarly publications published in 1990

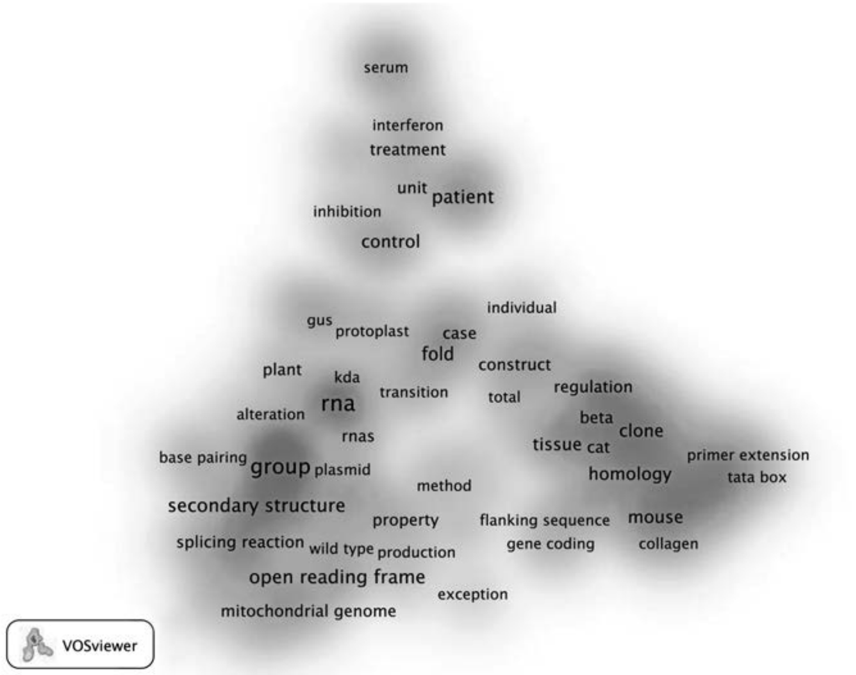
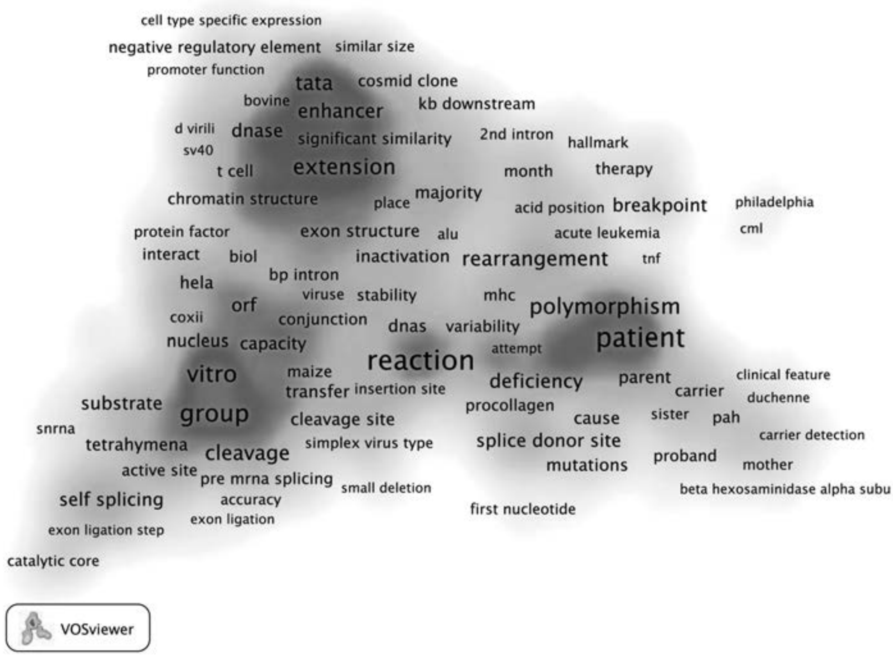
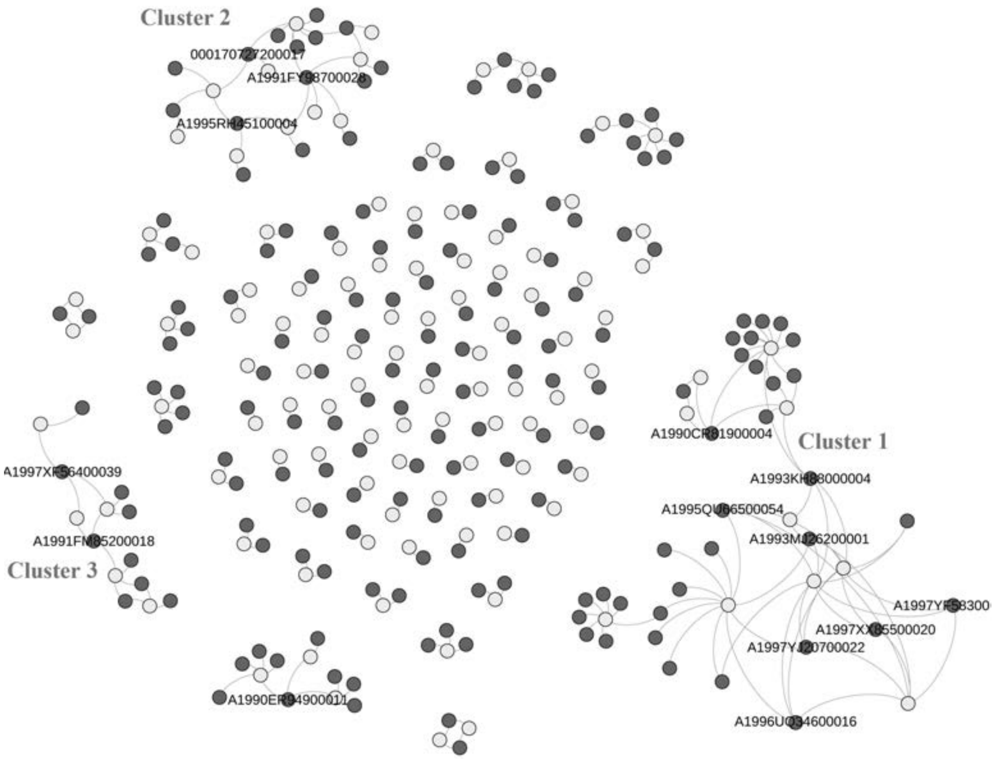


Figure 5. Term map for scholarly publications published in 1991



We find that 175 scholarly publications are cited within 112 patent families as non-patent literature (NPL) citations. These NPL citing-cited links are direct and visible connections between published science and patented technology. The network map in Figure 6 displays these NPL citing-cited links (citing-cited links between patents are omitted in order to highlight the science-technology interface) where the blue bubbles represent publications, while the white ones represent patent families. The graph shows a large number of dyads, mainly consisting of relations between a single publication and patent. These are the multitude of 'duplets' that occupy the centre. This patent-publication network contains three major patent-publication clusters. To place these clusters on a time-line we use the earliest application date of a patent family as application year for the whole patent family. For publications cited three or more times (maximum is 6 citations) as NPL the identification numbers from the Web of Science (UT's) are shown.

Figure 6. Patent-publication network according to non-patent-reference connections (1980–2010)



We used the patent classification codes assigned to the patent publications to determine the main 'technological' topic of a cluster. Based on the same scientific knowledge base (i.e. 'introns' science) the technologies represented in these three clusters diversify into three distinct technologies.

‘Recombinant DNA[ix]-techniques’ is the topic of Cluster 1 and the patent applications were applied for in the period 1992–2002. Cited research publications in this cluster were published between 1988 and 2005[x]. Cluster 2 focuses on ‘DNA vectors’[xi], patent applications in this cluster were applied for from 1998 until 2010, where the NPL-cited publications were published between 1991 and 2004. Cluster 3 covers the area of ‘Techniques to modify DNA to express or suppress genes’. The earliest patent application in this cluster is from 1995 and the most recent from 2005. The cited publications were published from 1986 until 2001.

Concluding remarks

A paradigm shift in science starts as a ‘localized phenomenon’, sometimes a discovery or breakthrough, the effects of which gradually spread throughout relevant research fields. This knowledge dissemination process reveals as simultaneous increases in numbers of unique authors, journals, journals categories, research organisations and the spread of different countries where those organisations are located. Figure1 shows that the number of patent applications is low prior to 1992, indicating that scientific knowledge shortly after the breakthrough in 1977 is not sufficiently mature for applications in new technologies. It also illustrates that the discovery in 1977 was a ‘Challenge’ breakthrough.

The sudden increase in the quantity of scholarly publications in 1990–1991 marks the stage in which intron-related science becomes mature in the sense of relevance for new technological applications; the increase is mainly visible in application-orientated scientific areas and marked by application oriented terms. The citing-cited network between scholarly publications and patents reveals three major clusters of publications. Patent applications in these three clusters start to appear after 1991, thus signalling a direct link between science and technology. The three clusters present technological differentiation based on the same scientific knowledge.

The observations we made support the two main research questions we addressed in this study. We are able to identify and characterize a breakthrough, involving a paradigm shift, in terms of bibliometric variables and indicators, and identified, bibliometrically, the stage in the development process where scientific knowledge is being used for work on science-based technologies. We furthermore conclude that the link between scientific knowledge and patented technology is established after a ‘Charge’ discovery successive to a ‘Challenge’ discovery.

Further research will focus on transformative developments in intron-related research since the ‘challenge’ breakthrough in 1977 that led to the ‘charge’ breakthrough in 1990–1991.

References

Andersen, H., Barker, P., and Chen, X. (2006). *The cognitive structure of scientific revolutions*. Cambridge University Press.

Van Andel, P. (1994). Anatomy of the unsought finding. serendipity: Origin, history, domains, traditions, appearances, patterns and programmability. *The British Journal for the Philosophy of Science*, 45(2): 631–648.

Hollingsworth, J. R. (2008). Scientific discoveries: An institutionalist and path- dependent perspective. In Hannaway, C. (Ed.), *Biomedicine in the Twentieth Century: Practices, Policies, and Politics*, 317–353. National Institute of Health.

Isnard, C. A. and Zeeman, E. C. (1976). Some models from catastrophe theory in the social sciences. In Collins, L. (Ed.), *The Use of models in the social sciences*, International Behavioral and Social Sciences Ser.: Classics from the Tavistock Press, 44–100. Tavistock Publications.

Koshland, D.E. (2007). The Cha-Cha-Cha theory of scientific discovery. *Science* 317(5839): 761–762.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. The University of Chicago Press.

Gelinas, R.E. and Roberts, R.J. (1977) One predominant 5'-undecanucleotide in adenovirus 2 late messenger RNAs. *Cell* 11, 533–544.

Roberts, R. J. (1993). Nobel lecture: 'Split genes and RNA splicing', December 8, 1993. (http://www.nobelprize.org/nobel_prizes/medicine/laureates/1993/roberts-lecture.pdf)

Sharp, P. A. (1993). Nobel lecture: 'An amazing distortion in DNA induced by a methyltransferase', December 8, 1993. (http://www.nobelprize.org/nobel_prizes/medicine/laureates/1993/sharp-lecture.pdf)

Tijssen, R. J. W. (2010). Discarding the 'basic science/applied science' dichotomy: A knowledge utilization triangle classification system of research journals. *Journal of the American Society for Information Science and Technology* 61(9): 1842–1852.

Winnink, J. J. and Tijssen, R. J. W. (2011). R&D dynamics in the development of HIV/AIDS drugs. In Noyons, E., Ngulube, P., and Leta, J., editors, *Proceedings of the 13th International Conference of the Insternational Society for Scientometrics & Informatics* (ISSI 2011), volume II, 855–860.

Winnink, J. J. (2012). Searching for structural shifts in science: Graphene R&D before and after Novoselov et al. (2004). In Archambault, E., Gingras, Y., and Larivière, V., editors, *Proceedings of the 17th International Conference on Science and Technology Indicators* (STI 2012), volume 2, 837–846.

Worrall, J. (2003). Normal Science and Dogmatism, Paradigms and Progress: Kuhn 'versus' Popper and Lakatos. In Thomas Nickles (Ed.) *Thomas Kuhn* (65–100). Contemporary Philosophy in Focus. Cambridge University Press.

---

i The word intron is derived from 'intragenic region'

ii Prokaryotic organisms are microscopic single-celled organisms, e.g. bacteria and cyanobacteria, that have neither a distinct nucleus with a membrane nor other specialized organelles.

iii Eukaryotic organisms are organisms consisting of a cell or cells in which the genetic material is DNA in the form of chromosomes contained within a distinct nucleus. Eukaryotes include all living organisms other than the eubacteria and archae.

iv See for instance Van Andel (1994))

v Because patents rights are national rights and patent procedures have several distinct phases patenting is a complicated process, in which several publications of the same invention can, and normally do, co-exist.

vi This classification is developed at CWTS and linked with the in-house version of the Web-of-Science database (TR-CWTS WoS). Linking with the on-line version of the WoS is based on 'Accession Numbers'.

vii For information on the VOSviewer see http://vosviewer.com

viii Before 1991 the Web-of-Science contained abstracts for only a fraction of the publications. We managed to obtain 138 abstracts for the 143 1990-publications, and 696 abstracts for the 703 1991-publications using additional sources.

ix DNA that has been formed artificially by combining constituents from different organisms

x Due to the complexity of the patenting process documents published after the filing of a patent application can show as cited non-patent-publication

xi An autonomously replicating DNA molecule into which foreign DNA fragments are inserted and then propagated in a host cell

Cross-Technology Diffusion of Software: A Patents-based Analysis[1]

Poh-Kam Wong PohKam@nus.edu.sg / NUS Entrepreneurship Centre, National University of Singapore, 21 Heng Mui Keng Terrace, S119613 (Singapore)

Yuen-Ping Ho yuenping@nus.edu.sg / NUS Entrepreneurship Centre, National University of Singapore, 21 Heng Mui Keng Terrace, S119613 (Singapore)

Abstract

This paper proposes a patents-based indicator which quantifies the extent of technological diffusion from a focal technology field to other fields. We use this indicator to examine diffusion trends in software, an area in which there has been a surge in patent registrations in the last decade. Using data on software patents granted in 2000–2012, we seek to establish how much of this rapid growth is due to increased application of software in non-software areas. The proposed indicator measures diffusion at point of invention and is developed by parsing the technology classifications of an invention as coded in its patent documentation. Four levels of cross-technology diffusion are identified: (i) Pure software or non-diffusion; (ii) Software + ICT; (iii) Software + non-ICT; (iv) Software + ICT + non-ICT. We find that more than half of software patents are non-diffused "pure software" inventions, and the proportion has grown over time. However, we also find that the advanced diffusion segment of Software + ICT + non-ICT has expanded, although still constituting only a small share of software patents. More detailed analysis reveals fast-growing diffusion of software into fields such as biotechnology and medical devices.

Introduction

The rapid development of digital technology in recent years, particularly quantum advancements in Information and Communications Technology (ICT) with the advent of the Internet as an enabler, has driven the rapid growth of the global digital economy. As a result, increasingly large numbers of innovations are in the form of applications of digital technologies. Inventions and tools that previously operated by mechanical means are now more reliant on computerized electronic controls. For example, UC San Diego computer scientist Professor Ingold Krueger stated that there are anywhere between 15 to 80 computerized units in a car, carrying out thousands of software functions (NPR, 2010).

1 This paper partially draws on data collated for a research project funded by the Intellectual Property Office of Singapore, as documented in Wong, Hang, Ho & Singh (2012).

Despite the pervasive phenomenon of the digital economy, there remains a dearth of systematic constructs to measure the extent of its diffusion. Existing indicators typically measure usage or adoption of technologies. Commonly cited examples of such indicators in the popular media are Internet adoption or broadband penetration rates, while academic studies have used measures such as capital expenditure or output share to proxy the diffusion of specified products or technologies (Comin, Hobijn & Ribito, 2006). While these are useful to show the general spread of digital technologies, they do not address the diffusion of digital technologies in the innovation landscape. In this paper, we develop a patents-based indicator which quantifies the extent of technological diffusion from a focal technology field to other fields. We use this indicator to examine diffusion trends in software, one of the fastest growing areas of digital technology.

Software Patents

In the digital economy, software is propelled to the forefront of innovation activities. In the creative content and digital media spaces, software content is proliferating as developers introduce new applications at a dizzying rate. Many hardware or machinery-based inventions also combine elements of software with hardware. Using empirical data and methods, we seek to establish the extent of the purported growth and influence of software technologies.

With software gaining prominence, the topic of software patents has garnered much attention. Traditionally, the intellectual property contained within software has been protected by copyright. For many years, patent laws in major jurisdictions regarded software as being outside the scope of inventions that are patentable. An early guideline issued in 1968 by the United States Patent and Trademark Office (USPTO) stated that computer programs were "mental steps" and did not constitute patentable processes or machines. Key judicial decisions have since considerably widened the scope for software inventions to be patented, beginning with a pivotal ruling by the specialist Federal Circuit Court in 1998 (*State Street Bank v. Signature Financial Group*) and culminating in the Supreme Court's ruling in *Bilski v. Kappos* in 2010. As a consequence, the USPTO now has among the most inclusive guidelines in the world for software patentability.

These changes at the USPTO have fuelled a heated debate in the international community over the desirability of allowing software to be patented. Much of the disagreement stems from different views on the macroeconomic impact of software patents, particularly the effects of a patents regime on the level of innovation in software development (Jaffe and Lerner, 2006). The debate is unresolved as there has been little empirical research on the topic (Bessen, 2011). At the firm level however, there are a number of persuasive arguments favouring patents over traditional copyright for software IP, the most important being the stronger protection and more clear-cut options for litigation against infringement afforded by patent laws. Empirical research shows that there has been a dramatic increase in the propensity to patent software since the pivotal US court decisions on the patentability of computer programs (Bessen and Hunt, 2007).

From a measurement perspective, the growth in software patenting is a welcome development and an improvement over the previous practice of using copyright protection. Because copyright registration is not mandatory in many jurisdictions, systematic data on software innovations were lacking. With software inventions being patented in large numbers, researchers now have access to detailed information coded in patent documents.

The starting point for the analysis in this paper is the stock of patents granted to software-related inventions. Prior research has established the rapid expansion of software patenting but the question of diffusion is largely unexamined. How much of the growth in software patents can we attribute to diffusion of software technology to other technology areas? By studying the technological composition of software patents, we develop an indicator which measures the degree of software diffusion.

Indicator of Cross-Technology Software Diffusion

In the past, software was commonly thought of as a standalone program running on a computer, for example a computer game or a word processing application. The field of software has since expanded beyond such "pure software" inventions. Many software applications are developed to support innovations in other fields of technology. In practice, software content is found in a multitude of products and industries beyond the traditional parameters of the ICT sector, such as medical devices, robotics and logistics.

It should follow that patents granted to software-related inventions are widely diffused to technology areas outside of the core software and ICT fields. We refer to this form of diffusion as "cross-technology" diffusion, which we quantify using an indicator based on the technology classifications of patented software.

Our proposed indicator draws on the concept of technological complexity, which is widely used in the literature (von Graevenitz , Wagner & Harhoff, 2008). The complexity of technologies is often framed in terms of to the industry or product in which technologies are applied. Cohen, Nelson & Walsh (2001) suggested a breakdown between discrete and complex industries/products and von Graevenitz et al. (2008) devised a concordance to technology classes using the OST-INPI/FhG-ISI nomenclature documented in the OECD Patent Manual. Globally, patenting in complex technologies has grown at a faster rate than discrete technologies (WIPO, 2011).

The concept of "complexity" is founded on the idea that technologies can be multi-faceted in nature, whether it is in their application or content. We expand on this idea to encompass the notion of diffusion. An invention with multi-technology content or applications is by definition diffused across technologies. In the context of software, a non-diffused "pure software" patent would only have software-related content and all its applications would be within the field of software. The corollary is a diffused software patent which has content from, or is applied in, other technology fields.

Our proposed indicator of cross-technology diffusion is based on the technological nature of patents themselves, rather than the technologies applied in patent-owning firms or industries. In scientometrics research, knowledge flows and diffusion have been measured by tracing the citation linkages between patents (Meyer, 2002; Hu, 2011). We differentiate from this approach by measuring diffusion at the point of invention and examining the diversity of technologies directly associated with the invention as coded within the patent document. In a patent document, the technologies germane to the invention are summarized in the technology class fields. For a patent granted by the USPTO, technology classes using both the IPC (International Patent Classification) and the USPTO's own US Patent Classification (USPC) are indicated.

We propose that a cross-technology diffused patent is one which is classified in multiple technological areas. A technological area is defined at the one-digit level of the NBER patent classification, which aggregates the detailed USPC schematic. At the one-digit level, the NBER classification identifies six technological areas: (i) Chemical, (ii) Computers & Communications, (iii) Drugs & Medical, (iv) Electrical & Electronics, (v) Mechanical, and (vi) Others. To identify diffused software patents, we further distinguish between Software and Other ICT in the area of Computers & Communications, making a total of 7 technology areas in the classification scheme.

A diffused patent is identified if it has technology classes spanning at least two areas. By definition, all software patents in our analysis will have Software as a technology class. If a software patent has at least one technology field in another (non-software) area, it is identified as a diffused patent. There are four possible levels of diffusion:

(a) Non-diffused: Pure software patent
(b) Software + ICT: A software patent which is diffused to other ICT area(s). This is regarded as a basic level of diffusion due to the proximity of software technology to other ICT.
(c) Software + Non-ICT: A software patent which is diffused to technology area(s) which are not related to ICT, such as Chemicals.
(d) Software + ICT + Non-ICT: A software patent which is diffused to both ICT and non-ICT. This is the most advanced level of diffusion, with the software patent interacting in multiple technology spaces.

A caveat to note is that this indicator imposes a restrictive definition on "diffusion", which is confined strictly to the point of invention. It is feasible, even likely, that "pure software" inventions have application potential in multiple fields. Such forms of diffusion are not captured at the point of invention and therefore not measured by this indicator.

The share of diffused patents indicates the proportion of the software patent stock which is diffused. We track the changes in this indicator over time to verify anecdotal evidence about the growing ubiquity of software applications.

Data and Methods

A key issue in studying software patents is the appropriate identification of patents for inclusion in the analysis. In the literature and at different patent offices, there are multiple definitions of what constitutes a "software patent." The USPTO's USPC system has devoted a section of classes to "computer implemented patents", spanning USPC 700 to 726. This is a broad categorization which was further expanded upon by Bessen (2011) to include classes of technologies that are "reliant on software".

In the academic literature, scholars have attempted to construct datasets of software patents through keyword searching (Bessen and Hunt, 2007) and identifying patents of top software or ICT firms (Graham and Mowery, 2003; Arora, Forman & Yoon, 2007; Hall and MacGarvie, 2010). In the latter approach, researchers examine the technology classes of patents assigned to top software/ICT firms and distil from these a list of IPC codes which are deemed to represent software-related inventions.

Wong et al. (2012) summarized the approaches and definitions that have been used in the literature, as shown in Table 1. Depending on the definition, the number of identified software patents granted by the USPTO in the 30-year period 1980–2010 varies from 157,761 (Bessen and Hunt's (2007) keyword search) to over 400,000 patents (Bessen's (2011) expansion of the USPTO's classification). In this paper, we adopt the definition used by Arora et al. (2007) which comprises 5 IPC classes, expanding on 3 IPC classes originally identified by Graham and Mowery (2003) based on their survey of patents in software firms.

Table 1. Definitions of Software Patents

	Identification Method	Total USPTO Patents 1980–2010
USPTO	"Computer Implemented Patents" USPC 700–726	355,927
Graham and Mowery (2003)	3 IPC Classes based on patents of selected software firms: G06F (Electrical Digital Data Processing) G06K (Recognition of Data), H04L (Secure Transmission of Digital Info)	221,466
Bessen and Hunt (2007)	Keyword search ("software" or "computer program", with exclusion words)	157,761
Arora et al. (2007)	5 IPC Classes: 3 as in Graham & Mowery + 2 others G06F (Electrical Digital Data Processing) G06K (Recognition of Data), H04L (Secure Transmission of Digital Info) G06T (Image Data Processing) G09G (Visual Indicators)	225,682
Bessen (2011)	USPC 700–707, 715–717 (data processing) + other selected classes ("reliant on software" and in which software firms patent)	429,326
Hall and MacGarvie (2010)	USPC subclasses based on patents on top ICT firms (details not revealed)	NA

Note: Number of USPTO granted 1980–2010 calculated by authors based on provided definitions

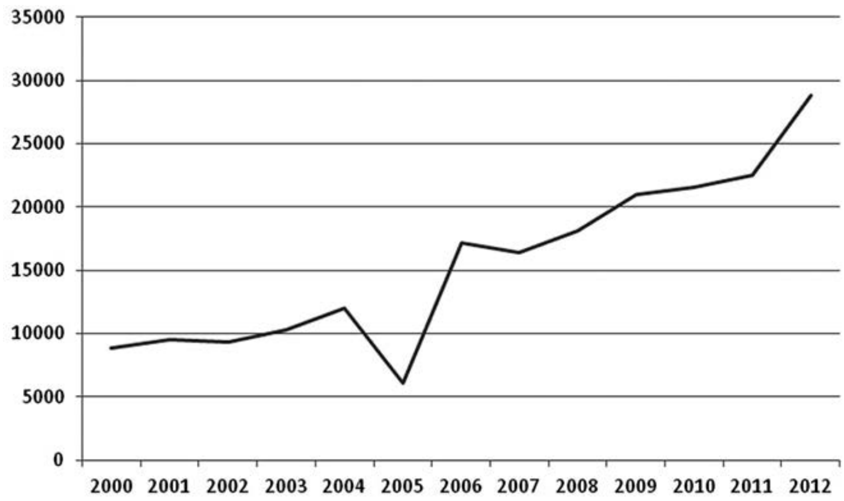
All analysis in this paper is based on patents granted by the USPTO. Using the 5 IPC classes identified by Arora et al. (2007) as the search parameter, software patents data from 2000 to 2012 were extracted via the Patsnap website, which provides an online database of USPTO patents. The post-millennium period was selected to reflect patenting trends after the US Courts liberalized the guidelines on software patentability in the late 1990s.

In total, 201,643 software patents were identified and extracted. The key field examined in our analysis is the technology classifications of each patent. Where there are multiple technology classes for a single patent, this field was parsed and each technology class was categorised into one of 7 technology areas. 12% of the patents only had a primary technology class and no secondary classes. The largest number of technology classes associated with a single patent in our dataset is 69.

Preliminary Findings

Annual grants of software patents by the USPTO has almost tripled in the last decade, as seen in Figure 1. This finding confirms and updates previous research on the recent escalation in software patent filings (Bessen and Hunt, 2007; Bessen, 2011).

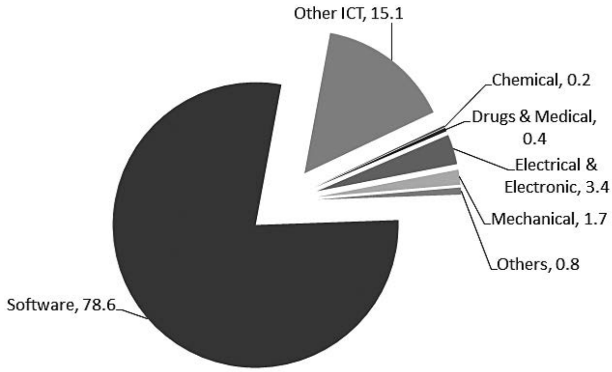
Figure 1. Number of Software Patents Granted by USPTO, 2000–2012



As shown in Figure 2, the overwhelming majority of software patents have Software as their primary technology classification, and 15% are classified primarily in a non-software ICT area. Very few software patents have primary classes outside of ICT. This provides an initial indication

that software patenting is still dominated by core ICT-related inventions, rather than inventions where software is incorporated or applied in other technologies.

Figure 2. Primary Technology Class of Software Patents, 2000–2012



Trends in our proposed indicator of cross-technology diffusion are shown in Figure 3. The proportions of software patents in the four levels of diffusions are represented by the four different shaded segments in each column. We see at once that more than half of software patents are non-diffused "pure software" inventions. This proportion rose from 53.5% in the period 2000–2002 to 58.5% in 2009–2011, before declining to 56% in 2012. This leads us to the slightly surprising conclusion that the growth in software patents has been driven by pure software inventions rather than software applied to products/services from other sectors. Of the patents that are diffused, the majority are diffused to other ICT fields, accounting for around one third of software patents. The share of patents in the Software + ICT segment has declined gradually between 2000 and 2011, before increasing slightly in 2012.

Higher levels of diffusion (Software + non-ICT and Software + ICT + non-ICT) form a relatively small share of software patents, hovering at the 11% to 12% mark throughout the last decade. Interestingly, we observe that the segment representing the most advanced level of diffusion, Software + ICT + non-ICT, has become larger. This is a notable result, suggesting the evolution of software-related inventions, where diffusion occurs through the interaction of software and hardware elements of ICT applied to non-ICT areas, an example of which would be biomedical devices.

Figure 3. Cross-Technology Diffusion of Software Patents, 2000–2012

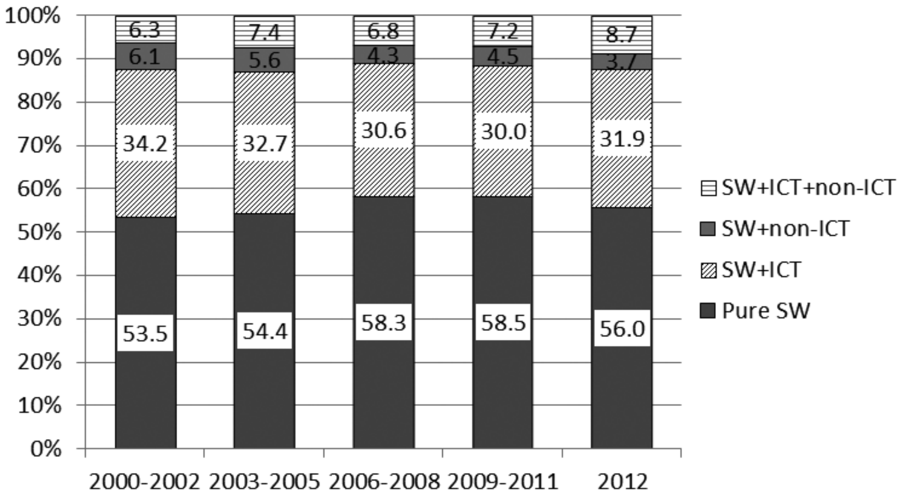


Table 2 breaks down in finer detail the shares of software patents diffused to non-ICT technology areas (collectively including diffusion to the SW + non-ICT and SW + ICT + non-ICT segments). The bulk of software patents diffused to non-ICT are found in Electrical & Electronics (around 7% of software patents), followed by Mechanical (around 3–4% of software patents). Only very small shares of software patents are diffused to the Drugs & Medical and Chemical areas. However, we observe that the diffusion to Drugs & Medical has almost doubled in the last 12 years, from 0.6% in 2000–2002 to 1.1% in 2012.

Table 2. Proportion of Software Patents Diffused to non-ICT Areas

	Chemical	Drugs & Medical	Electronic & Electrical	Mechanical	Others
2000–2002	0.5%	0.6%	7.1%	4.3%	1.5%
2003–2005	0.4%	0.8%	7.5%	4.4%	1.3%
2006–2008	0.3%	0.8%	6.8%	3.4%	1.1%
2009–2011	0.4%	1.0%	6.5%	3.0%	2.0%
2012	0.5%	1.1%	7.5%	3.4%	1.2%

We further drill down the Drugs & Medical area to examine software patent diffusion into two sub-classes, namely Medical Instruments and Biotechnology. As seen in Table 3, software patenting in these fields has expanded at a rapid rate, matching or outpacing the growth in total software patents. Table 3 also shows that growth in Software + Medical Instruments/ Biotech-

nology is much higher than in the Medical Instruments/ Biotechnology sub-class as a whole. Although the absolute number of patents is still quite small, this shows the strong potential for software diffusion to the biomedical and biotechnology sectors.

Table 3. Growth in Software Patents Diffused into Drugs & Medical Sub-Classes

	Software Patents Diffused to		All Medical Instruments Patents	All Biotechnology Patents	All Software Patents
	Medical Instruments	Biotechnology			
Average granted annually					
2000–2005	45	11	6,571	5,371	9,338
2006–2012	152	23	7,431	5,927	20,802
Average annual growth (%)					
2000–2012	19.6%	10.0%	4.8%	1.8%	10.3%

Table 4 examines software diffusion from another perspective, showing the penetration of software patents in non-ICT classes. While the proportion of software patents which are diffused has either declined or remained unchanged, there is evidence of increasing software content in non-ICT patents. We note from Table 4 the greatly increased propensity for Drugs & Medical and Electrical & Electronics patents to contain software-related technology.

Table 4. Proportion of Patents in non-ICT with Software as a Technology Class

	Chemical	Drugs & Medical	Electronic & Electrical	Mechanical	Others
2000–2002	0.07%	0.24%	1.37%	0.54%	0.10%
2003–2005	0.07%	0.43%	1.43%	0.57%	0.10%
2006–2008	0.09%	0.81%	2.35%	0.78%	0.15%
2009–2011	0.12%	0.99%	2.29%	0.69%	0.27%
2012	0.15%	1.07%	2.98%	0.83%	0.18%

Preliminary Conclusions

Subject to the caveat that our indicator restricts diffusion to the point of invention, we find that the majority of software patents are not diffused to other technology areas. The surge in software patenting in the last decade has been fuelled mainly by pure software inventions. This may point to non-ICT firms and inventors being slow to capitalize on the patentability of software elements in their innovations. We note however that there are growing shares of diffused patents in certain non-ICT sub-classes such as medical instruments and devices.

In the full version of this paper, we will expand our analysis to differentiate between patents with primary technology class in Software versus those that are primarily in non-Software fields. We will also complement our analysis by studying the contribution of software-related patents in the total patents portfolios of major firms in selected sectors with potential for software diffusion, such as the automotive, biomedical and healthcare sectors.

References

Arora, A., Forman, C. & Yoon, J. (2007). Software. In J. T. Macher and D.C. Mowery (Eds.), *Innovation in Global Industries: American Firms Competing in a New World* (pp 53–99). Washington DC: The National Academies Press.

Bessen, J. (2011). A generation of software patents. Boston University School of Law Working Paper, No. 11–31.

Bessen, J. & Hunt, R.M. (2007). An empirical look at software patents. *Journal of Economics and Management Strategy*, 16(1), 157–89.

Cohen, W.M., Nelson, R.R. & Walsh, J.P. (2000). Protecting their Intellectual Assets: Appropriability Conditions and why US Manufacturing Firms Patent (Or Not). NBER Working Paper No. 7522. Cambridge, MA: National Bureau of Economic Research.

Comin, D., Hobijn, B. & Rovito, E. (2006). Five facts you need to know about technology diffusion. NBER Working Paper No. 11928. Cambridge, MA: National Bureau of Economic Research.

Graham, S. J. H. & Mowery, D.C. (2003). Intellectual property protection in the U. S. software industry. In W. M. Cohen and S.A. Merrill (Eds.), *Patents in the Knowledge-Based Economy* (219–58). National Research Council, Washington: National Academies Press.

Hall, B. &MacGarvie, M. (2010). The private value of software patents. *Research Policy*, 39(7), 994–1009.

Hu, M.C. (2011). Evolution of knowledge creation and diffusion: the revisit of Taiwan's Hsinchu Science Park. *Scientometrics*, 88(3), 949–977.

Jaffe, A.B. & Lerner, J. (2006). *Innovation and its Discontents: How our Broken Patent System is Endangering Innovation and Progress, and What to do About it*. Princeton: Princeton University Press.

NPR (2010). Driving Cars Towards Complexity. Retrieved February 2013 from http://www.npr.org/templates/story/story.php?storyId=126418391

von Graevenitz, G., Wagner, S. & Harhoff, D. (2008). Incidence and Growth of Patent Thickets – The Impact of Technological Opportunities and Complexity. CEPR Discussion Paper number 6900. London: Centre for Economic Policy Research.

Meyer, M. (2002). Tracing knowledge flows in innovation systems. Scientometrics, 54(2), 193–212.

WIPO (2011). *World Intellectual Property Report 2011*. The Changing Face of Innovation. Geneva: World Intellectual Property Organization.

Wong, P.K., Hang, C.C., Ho, Y.P. & Singh, A.A. (2012). Intellectual Property Rights in the Global Digital Convergence Space. Report submitted to the Intellectual Property Office of Singapore. Singapore: NUS Entrepreneurship Centre.

The Influence of Group Research Support on Research Collaboration Networks: Center-level Evidence from the NRF Science/Engineering Research Center Program

Chang Hoon Yang cy8064@gmail.com / Department of Public Administration, Kwandong University, 522 Naegok-dong, Gangneung, Gangwon-do, 201–701 (South Korea)

Jungeun Heo prettyheo@nrf.go.kr / Division of National R&D Programs Planning, National Research Foundation of Korea, 165 Gajeong-ro, Yuseong-gu, Daejeon, 305–350 (South Korea)

Introduction

The group research support program in National Research Foundation of Korea (NRF) have provided extensive support to research collaboration networks between excellent researchers engaged in basic research in science and engineering (S&E) since 1990s. The implementation of the program has been initiated which should lead to the so-called Advanced Research Center Program (ARCP). The main goal of the ARCP is to organize researchers working at domestic universities into research groups by specific goals and fields and provide intensive financial supports to those groups so that the participating researchers may augment their competence. The ARCP is classified into three categories according to their research fields: Science/Engineering Research Center (S/ERC), Medical Research Center (MRC), and National Core Research Center (NCRC).

This paper is aiming at the identification and analysis of the research collaboration networks developed in the context of group research activities. In particular, it focuses on a special form of group research partnerships of S/ERC program. The S/ERCs are a group of cross-disciplinary, multi-institutional university research centers that join leading researchers in partnerships who are adept at innovation and primed for leadership in the basic research. The S/ERC is the first government-sponsored university research center program in Korea for the purpose of nurturing basic research and education in S&E by building stronger research relationships among faculty, students, and academic scientists and engineers and offering specialized cross-disciplinary master's and doctoral programs in S&E. We are aiming at the preliminary investigation of the collaboration networks of a new center-based model rather than traditionally department (or discipline)-based university research in basic science that emerge in consequence of group research activities in the area of human sensing system during the period 2009–2011. It is assumed that the formation of group research activities based on cross-disciplinary research center involves the cooperative partnerships of researchers from different fields that would exchange various levels of scientific and engineering capabilities and expertise and generates original research outcomes that could be classified into more than one field. Through the center-based cooperative

university research, capacity building among researchers as well as cross-disciplinary linkages among research fields may be enhanced and fostered.

Drawing from a range of literatures, cooperative research activities have increased rapidly in the field of S&E today (Altenkirch et al., 2003; Olson, Olson & Hofer, 2005; NSB, 2012). In particular, the increasing complexity of current basic research experiments and knowledge base needed for innovation with the fast advances in technology has promoted cooperative group research initiatives in S&E. In this regard, the concept of cross-disciplinary research collaborations can be defined as the level of integration (Rosenfield, 1992; Klein, 1996; Lattuca, 2003; Aboelela, et al., 2007; Stokols, et al., 2008). Multidisciplinary research refers to a research in which researchers from more than one discipline work independently on different areas of a project while remaining within their disciplinary boundaries. Interdisciplinary research is defined as research efforts whereby researchers from various disciplines work in partnership on a project using their discipline-specific perspectives. Transdisciplinary research is a field of research undertaken by researchers from different disciplines collaborating on a project using a shared framework that integrates various disciplinary approaches into a collective whole. Given these definitions, center-based cross-disciplinary university research is the collaborative efforts of researchers from different fields, affiliated with universities, in coordinating a project with cross-disciplinary perspectives to address and solve both shared and challenging scientific and engineering problems.

Boardman & Corley (2008) noted that university research centers tend to foster collaborations and interactions among researchers by forming multidiscipline, inter-organizational or cross-sector collaborations based on the scientific and technical goals of the center (Bozeman & Boardman, 2003). In other words, cross-disciplinary initiatives, involving multi-institutional research centers, serve to connect collaborators across disciplines in order to inform research results and methods, disseminate needed information and knowledge, and promote best practices as a means of eliminating traditional gaps in terminology, approach, and methodology and integrating the analytical strengths of various scientific disciplines that relate to a given problem (Aboelela et al. 2007). However, Stokols et al. (2008) pointed out other studies that cross-disciplinary research may cause problems of diverting resources from discipline-specific research and drawing research personnel into collaborative centers who otherwise might be more effective to research independently than in a collaborative environment. Correspondingly, understanding and assessing the extent to which group research activities influence the research collaboration network is the research question that must be answered in the evaluation of such university research centers and in the improvement of research funding program regarding cross-disciplinary, multi-institutional research collaborations in S&E.

Even though there is a rich related literature on cooperative research centers (Rogers et al., 1999; Bozeman & Boardman, 2003, 2004; Boardman & Corley, 2008; Boardman & Gray, 2010; Roessner, Manrique & Park, 2010; Rogers, 2012), relatively few studies have examined network topologies of cross-disciplinary relations within goal-oriented university research center. In this paper, we investigate research collaboration networks by using two categories in a university

research center, sub-projects and research fields. Specifically, this paper utilizes data from a sample of 12 sub-projects drawn from a cross-disciplinary S/ERC project (e.g., human sensing system). The center's sub-projects encompass a diverse range of disciplines, including materials science, chemistry, nanoscience and technology, physics, and biology, etc. In an effort to explore network topologies of cooperative research activities, we examine a matrix of the co-occurrence of sub-projects and research fields. In particular, the network method provides a systematic analytical tool to uncover the hidden structure of cross-disciplinary research relations and to monitor the effectiveness of group research support program designed to foster cooperative research across sub-projects and different academic fields.

Methods and data

What are the forms of relationships among research profiles that constitute networks which represent some essential features of cooperative research activity? How cross-disciplinary, multi-institutional relations might be formed across various research activities in different disciplines? There has been a rising interest in network research that has sought to examine these questions. This study focuses on network linkages by using the co-occurrence of the sub-project and research field identified by published journal articles. A co-occurrence matrix of the sub-project and research field may be taken as a basis for a relational linking process to identify clusters of cross-disciplinary group research activities. These clusters can be interpreted as the cooperative research network structures emerging between sub-projects and research fields, respectively. Due to the co-occurrence of sub-projects and research fields, we can assume the existence of certain relations among group research activities. Thus, a network analysis using a co-occurrence mapping method would allow us to uncover overall network topologies to assess cross-disciplinary, multi-institutional research efforts.

Examining network topologies for a given S/ERC project would require a dataset capable of characterizing collaborative group research activities. The dataset included 12 sub-projects that were interrelated to the main overall project (e.g., human sensing system) and funded by the S/ERC program for a 3 year period from 2009 to 2011. The dataset also contained 102 research articles that were published in SCI-listed journals and generated by the funding of this S/ERC project for the above mentioned period. For the elements of the research field categories, we used the classification method employed in the 2011 Journal Citation Report (JCR), a Thomson Reuters publication. In this study, the number of observations for the sub-project and the research article were 12 (n) and 102 (k) and for the research field and the research article were 39 (n) and 102 (k), respectively. A two-mode (n-by-k) matrix was utilized for the network analysis. Based on our data we analyze two different types of networks:

(i) *Collaborative relations*: for each research article in which authors worked on a same sub-project, we define a research network which is a reflection of the collaborative relations between the sub-projects. In this network, the nodes are the sub-projects, and these sub-

projects are considered to be linked if they contain the same research article. In the collaboration network, each research article *k* was examined to determine whether the article was observed in the sub-project *i* and the sub-project *j*. Thus, $R_{ij:k}$ is equal to 1 if the sub-projects *i* and *j* are classified in the same article *k*, whereas diagonal $R_{ij:k}$ is equal to the total number of articles where the sub-project *i* belongs to the research article *k*. In the case of a positive indication, both of the sub-projects *i* and *j* would place a check in the cell of the co-occurrence matrix that corresponds to a collaborative relation between *i* and *j*. Thus, the collaborative network is as follows: $R_k = 1$ if $R_{ij:k} = 1$ and 0 otherwise.

(ii) *Cross-disciplinary relations*: for each research article within a specific research field, we define a research network which is a reflection of the cross-disciplinary relations between the research fields. In this network, the nodes are the research fields, and these fields are considered to be linked if they contain the same research article. In the cross-disciplinary network, each research article *k* was examined to determine whether the article was observed in the research field *i* and the research field *j*. Thus, $R_{ij:k}$ is equal to 1 if the research fields *i* and *j* are classified in the same article *k*, whereas diagonal $R_{ij:k}$ is equal to the total number of articles where the research field *i* belongs to the research article *k*. In the case of a positive indication, both of the research fields *i* and *j* would place a check in the cell of the co-occurrence matrix that corresponds to a cross-disciplinary relation between *i* and *j*. Thus, the cross-disciplinary network is as follows: $R_k = 1$ if $R_{ij:k} = 1$ and 0 otherwise.

Based on the two-mode matrix, we constructed a symmetric one-mode matrix for the sub-projects as well as for the research fields respectively, by multiplying this two-mode matrix by its transposes. Data for the sub-project, the research field, and the research articles were assembled in Excel and imported into UCINET 6 (Borgatti et al., 2002). The visualizations were made with Net-Draw (Borgatti, 2002), which created a one-mode research network image representing cross-disciplinary collaborative research relations among sub-projects and various research fields, respectively.

Results

Overall network relationship between sub-projects

Understanding the relations between sub-projects and identifying the underlying cooperative network structure are essential features of defining cooperative research activities in a given network. We used network analytical tools to examine network structures indicating collaborative relations weighted by the location of sub-projects. With respect to the units of analysis, the nodes are sub-projects, and the links represent relations between sub-projects undertaken by university researchers whose papers are produced by the authors that work in sub-project groups. The cooperative research activities shown in Table 1 indicate the number of published research articles generated by collaboration between sub-projects. Each of the 12 sub-projects in the dataset was measured. There were 45 collaborative research articles out of 183 possible publica-

tions of research articles, confirming the presence of 24.6% of all possible publications of collaborative research articles.

Table 1. Number of articles between sub-projects.

	1(1)	1(2)	1(3)	1(4)	2(1)	2(2)	2(3)	2(4)	3(1)	3(2)	3(3)	3(4)
1(1)	18	13**	5	2	2	1	0	0	0	1	0	1
1(2)*	13	30	6	2	0	0	0	2	0	0	0	2
1(3)	5	6	12	1	0	0	0	0	0	0	0	0
1(4)	2	2	1	6	0	0	0	0	0	0	0	0
2(1)	2	0	0	0	11	1	3	0	0	0	0	0
2(2)	1	0	0	0	1	10	0	0	0	0	0	0
2(3)	0	0	0	0	3	0	13	0	0	0	0	0
2(4)	0	2	0	0	0	0	0	17	0	0	0	0
3(1)	0	0	0	0	0	0	0	0	9	3	0	0
3(2)	1	0	0	0	0	0	0	0	3	4	0	0
3(3)	0	0	0	0	0	0	0	0	0	0	4	0
3(4)	1	2	0	0	0	0	0	0	0	0	0	4

* 1(2) indicates sub-project 2 in project 1.
** Number of articles generated by collaboration between 1(1) and 1(2) is 13 (only off-diagonal counted).

The descriptive statistics for the collaborative research network shown in Table 2 demonstrate significant variation in the degree centrality of the research network. The coefficient variation of the degree centrality was 1.11 (Std. Dev. = 8.322 relative to a mean of 7.50), and the population exhibited a scattered distribution. The overall degree centralization was low at 14.69% of the purely centralized network. The network of collaborative relations was less centralized and might have been dominated more by various groups of sub-projects than by a single major sub-project.

The degree centrality describes the extent to which a sub-project may be integrated into a network of collaborative relations. In the degree centrality results shown in Table 2, sub-projects *1(1)*, *1(2)*, and *1(3)* had a higher percentage of degree links, indicating that they were the most central cooperative project portfolios because each had more than an 8% share of network centrality. Figure 1(a) shows these sub-projects visually positioned in the center of the network. For instance, the largest and most central node belongs to sub-project *1(1)*, and there are other sub-projects nearest to this node: [*1(1)*, *1(2)*, *1(3)*, *1(4)*, *2(4)*, *3(4)*] and [*1(1)*, *2(1)*, *2(2)*, *2(3)*, *3(1)*, *3(2)*]. In addition, Figure 1(b) shows sub-structures of collaborative relations as a dendogram found from k-plex analysis. K-plexes are sub-structures in which each sub-project is connected to all (*n*) but *k* of the other sub-projects, where *n* is the number of sub-projects (size) and the parameter *k* ranges from 2 to 5 geodesic distances from which sub-structures are found. The set of all sub-projects that are combined together into a single cluster is considered as a possible sub-structure. There were 25 maximal sub-structures presented in these relations. Sub-projects *1(1)* and *1(2)* shown in the

cluster had strong overlap with the sub-structures identified using k-plex analysis (a second cluster is formed by sub-projects *1(1)*, *1(2)*, and *1(3)*; a third by sub-projects *1(1)*, *1(2)*, *1(3)* and *1(4)*; a fourth by sub-projects *2(1)* and *2(2)*). It is apparent that sub-project *3(3)* was a complete isolate (with lower quality scores in research performance: the relative impact factor (RIF) = 0.58 in Table 2), and that sub-projects *1(1)* and *1(2)* were central in the sense of playing a bridging role among multiple slightly different collaborative relations in 7 of the 25 cliques in a given network of 2-plexes.

With respect to the betweenness centrality, which measures the extent to which a particular sub-project lies between other sub-projects in the network, Table 2 shows that the overall network centralization was relatively high (53.55%). Sub-projects *1(1)*, *1(2)*, 2(1), and *3(2)* played the role of an intermediary or a broker bridging the adjoining sub-projects in the collaborative research network. In other words, these sub-projects were considered to have more power than others in coordinating cooperative relations in a given network. The closeness centrality indicates the potential independence of a sub-project from the relation of research collaboration. With closeness, sub-project *1(1)* was the most close to all others in the research network, indicating that this sub-project would be able to create collaborative relations with other sub-projects in the network, while sub-project *3(1)* had the largest sum of geodesic distance to others in a given network (sub-project *3(3)* was placed in an isolated position, separated from the collaborative network).

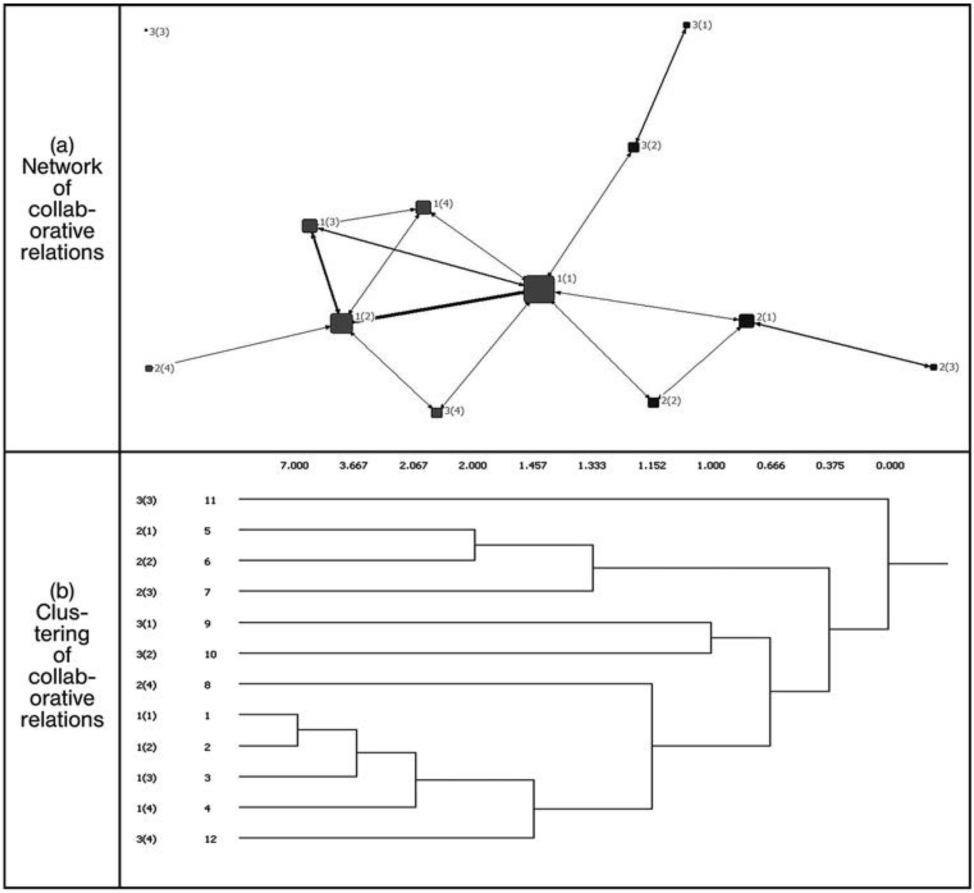
Finally, eigenvector centrality offers a measure of the diversity of a sub-project's network. In this manner, a sub-project that is collaborated with other sub-projects which are well collaborated has a high level of eigenvector centrality. On the contrary, a sub-project that is collaborated with other sub-projects which are less collaborated does not have a high level of eigenvector centrality. The result shows that sub-project *1(2)* had the highest eigenvector centrality measures because this sub-project had collaborated with sub-projects with high degree centrality such as sub-project *1(1)* and *1(3)*, indicating the important sub-project in the collaborative network (the degree of inequality was 142.58% of the maximum possible).

Table 2. Centrality measure scores.

	Degree	nDeg	Farness	nClose	Betweenness	nBet	Eigenvector	nEigen	RIF*
1(1)	25	17.483	25	44.00	32	58.182	0.524	74.121	0.70
1(2)	25	17.483	29	37.93	10	18.182	0.801	113.275	0.76
1(3)	12	8.392	31	35.48	0	0.000	0.259	36.663	0.87
2(1)	6	4.196	31	35.48	9	16.364	0.036	5.087	0.69
1(4)	5	3.497	31	35.48	0	0.000	0.083	11.767	0.72
3(2)	4	2.797	32	34.38	9	16.364	0.014	2.020	0.89
3(4)	3	2.098	32	34.38	0	0.000	0.058	8.134	0.76
3(1)	3	2.098	41	26.83	0	0.000	0.001	0.190	0.80
2(3)	3	2.098	40	27.50	0	0.000	0.004	0.546	0.62
2(4)	2	1.399	38	28.95	0	0.000	0.067	9.453	0.72
2(2)	2	1.399	32	34.38	0	0.000	0.018	2.558	0.81
3(3)	0	0.000	0	0.00	0	0.000	0.000	0.000	0.58
Mean	7.500		32.909		5.000		0.155		
Std. Dev.	8.322		4.601		9.065		0.243		
Network Centralization (Degree) = 14.69%									
Network Centralization (Betweenness) = 53.55%									
Network Centralization (Eigenvector) = 142.58%									

*RIF (relative impact factor) indicates a journal's place and impact within a specific discipline.

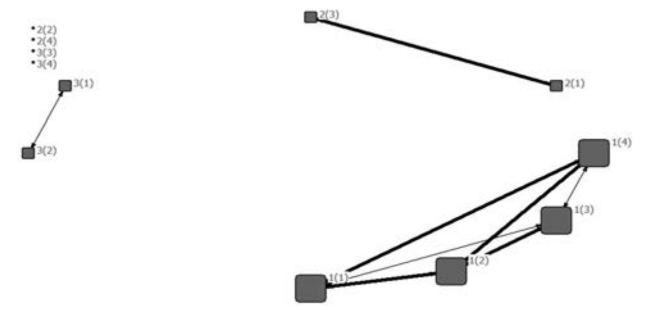
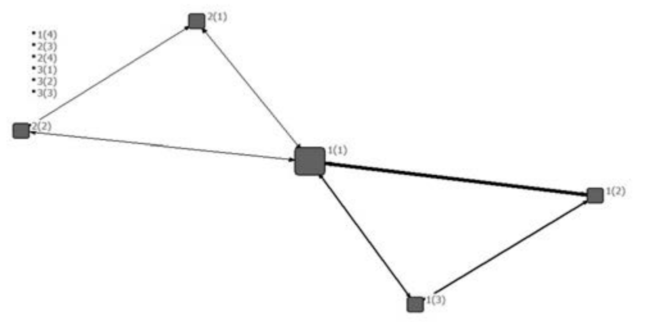
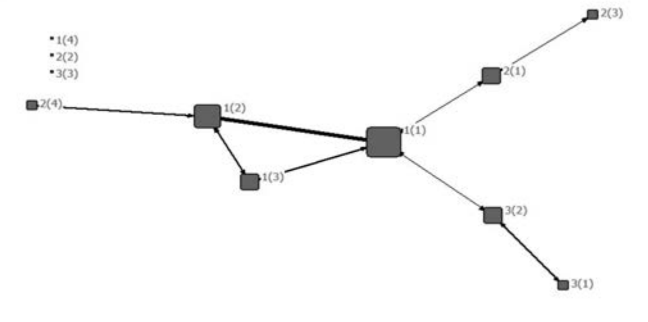
Figure 1. Network centralization in collaborative relations.



* The degree of research collaboration is represented by the size of the nodes, and the network ties indicate the strength of collaborative relations across the sub-projects.

The network centralization measures shown in Table 3 present a picture of the network structure of the research cooperation between sub-projects in different time periods. The types of networks are characterized by a different evolution of the collaborative relation structure. The analysis shows a trend towards decentralization. The overall degree centralization – a measure of the spread of collaboration across the network – decreased from 20.91 to 14.89% while the betweenness of the network centralization increased from 0 to 31.56%. In addition, the degree of inequality increased from 34.77 to 143.81% of the maximum possible. A glance at figures in Table 3 shows that these research linkages were significantly increased through cooperation. To summarize these observations, the overall network became less centralized, which was due to the collaborative relations possible between sub-projects which were not in the homogeneous research project field. It suggests that as the network grew over time, collaborative relations were spread more widely among sub-projects in different research project fields (with different disciplines).

Table 3. Network through collaborative relations for 2009–2011.

<table>
<tr><th></th><th>Network of collaborative relations</th><th>Network Centralization</th></tr>
<tr><td>2009</td><td>
</td><td>Degree = 20.91%
Betweenness = 0.0%
Eigenvector = 34.77%</td></tr>
<tr><td>2010</td><td>
</td><td>Degree = 16.67%
Betweenness = 8.89%
Eigenvector = 142.02%</td></tr>
<tr><td>2011</td><td>
</td><td>Degree = 14.89%
Betweenness = 31.56%
Eigenvector = 143.81%</td></tr>
</table>

* The degree of research collaboration is represented by the size of the nodes, and the network ties indicate the strength of collaborative relations across the sub-projects.

Overall network relationships of cross-disciplinary research activities

Understanding the co-occurrence of research fields and identifying the underlying cross-disciplinary network structure are essential features of defining research interdependence in a given network. We used network analytical tools to examine network structures indicating cross-disciplinary relations weighted by the location of cross-disciplinary research fields. With respect to the

units of analysis, the nodes are research fields, the links represent relations between research fields created by the activity of researchers whose papers cover different research fields. The descriptive statistics for the cross-disciplinary research network shown in Table 4 demonstrate significant variation in the degree centrality of the research network. The coefficient variation of the degree centrality was 1.13 (Std. Dev. = 10.623 relative to a mean of 9.385), and the population exhibited a scattered distribution. The overall degree centralization was relatively low at 8.51% of the purely centralized network. The network of cross-disciplinary relations was less centralized and might have been dominated more by various groups of research fields than by a single major research field.

The degree centrality describes the extent to which a research field may be integrated into a network of cross-disciplinary relations. In the degree centrality results shown in Table 4, *Materials Science*, *Multidisciplinary*; *Chemistry*, *Multidisciplinary*; and *Nanoscience & Nonotechnology* had a higher percentage of degree links, indicating that they were the most central cross-disciplinary research platforms (in the human sensing system project) because each had more than a 9% share of network centrality. Figure 2 shows these research fields visually positioned in the center of the network clusters: [*Chemistry*, *Multidisciplinary*; *Biotechnology & Applied Microbiology*; *Biochemical Research Methods*; *Biochemistry & Molecular Biology*; *Biophysics*] and [*Nanoscience & Nonotechnology*; *Materials Science*, *Multidisciplinary*; *Chemistry*, *Physical*; *Physics*, *Applied*] had the highest cross-disciplinary relations with each other.

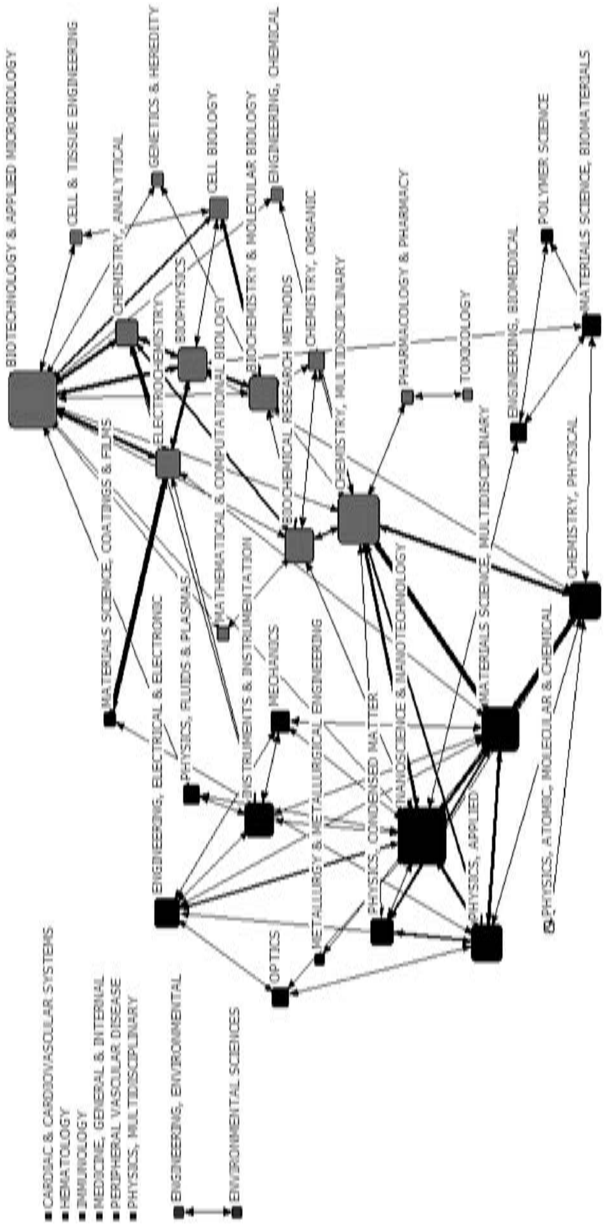
With respect to the betweenness centrality, which measures the extent to which a particular research field lies between other research fields in the network, Table 4 shows that the overall network centralization was relatively low (13.11%). *Nanoscience & Nanotechnology*; *Chemistry*, *Multidisciplinary*; and *Biotechnology & Applied Microbiology* played the role of an intermediary bridging the larger cluster of research fields with the smaller cluster in the cross-disciplinary research network. The closeness centrality indicates the potential independence of a research field from the cross-disciplinary relations. With closeness, Nanoscience & Nanotechnology; Chemistry, Multidisciplinary; and Biotechnology & Applied Microbiology were the most close to all others in the research network, indicating that these research fields would be in a relatively better position than others to create cross-disciplinary research opportunities across various research fields in the network.

Finally, eigenvector centrality offers a measure of the diversity of a research field's network. In this manner, a research field that is inter-linked to other research fields with a high degree of centrality has a high level of eigenvector centrality. The result shows that *Chemistry*, *Multidisciplinary* and *Materials Science*, *Multidisciplinary* had the highest eigenvector centrality, indicating the key research fields in the cross-disciplinary network (the degree of inequality was 82.90% of the maximum possible).

Table 4. Centrality measure scores – Top 20.

Rank	Research fields	Degree	nDeg	Between-ness	nBet	Far-ness	nClose	Eigen-vec	nEigen
1	Materials Science, Multidisciplinary	37	10.82	48.23	6.86	368	10.33	0.54	76.78
2	Chemistry, Multidis-ciplinary	34	9.94	105.30	14.98	362	10.50	0.56	79.02
3	Nanoscience & Nanotechnology	34	9.94	105.46	15.00	360	10.56	0.36	50.89
4	Electrochemistry	26	7.60	20.73	2.95	374	10.16	0.12	17.08
5	Biotechnology & Applied Microbiology	25	7.31	102.12	14.53	363	10.47	0.09	13.37
6	Physics, Applied	24	7.02	24.06	3.42	371	10.24	0.29	40.31
7	Biophysics	21	6.14	32.61	4.64	372	10.22	0.07	9.46
8	Chemistry, Analytical	20	5.85	5.06	0.72	377	10.08	0.12	16.33
9	Chemistry, Physical	20	5.85	52.47	7.46	368	10.33	0.31	44.40
10	Biochemistry & Molecular Biology	14	4.09	15.25	2.17	375	10.13	0.04	5.72
11	Biochemical Research Methods	12	3.51	18.85	2.68	371	10.24	0.09	13.12
12	Cell Biology	11	3.22	2.58	0.37	384	9.90	0.02	2.86
13	Physics, Condensed Matter	11	3.22	0.00	0.00	374	10.16	0.13	18.59
14	Engineering, Electrical & Electronic	10	2.92	3.17	0.45	380	10.00	0.07	9.44
15	Instruments & Instrumentation	10	2.92	10.65	1.52	378	10.05	0.05	7.65
16	Materials Science, Coatings & Films	9	2.63	0.74	0.11	389	9.77	0.04	6.05
17	Optics	6	1.75	0.00	0.00	386	9.84	0.04	5.55
18	Materials Science, Biomaterials	5	1.46	17.58	2.50	384	9.90	0.01	1.62
19	Cell & Tissue Engineering	4	1.17	0.00	0.00	391	9.72	0.01	0.89
20	Engineering, Biomedical	4	1.17	15.78	2.25	384	9.90	0.01	1.54
Mean		9.385		15.667		445.818		0.080	
Std. Dev.		10.623		28.806		253.826		0.139	
Network Centralization (Degree) = 8.51%									
Network Centralization (Betweenness) = 13.11%									
Network Centralization (Eigenvector) = 82.90%									

Figure 2. Network centralization in cross-disciplinary relations.

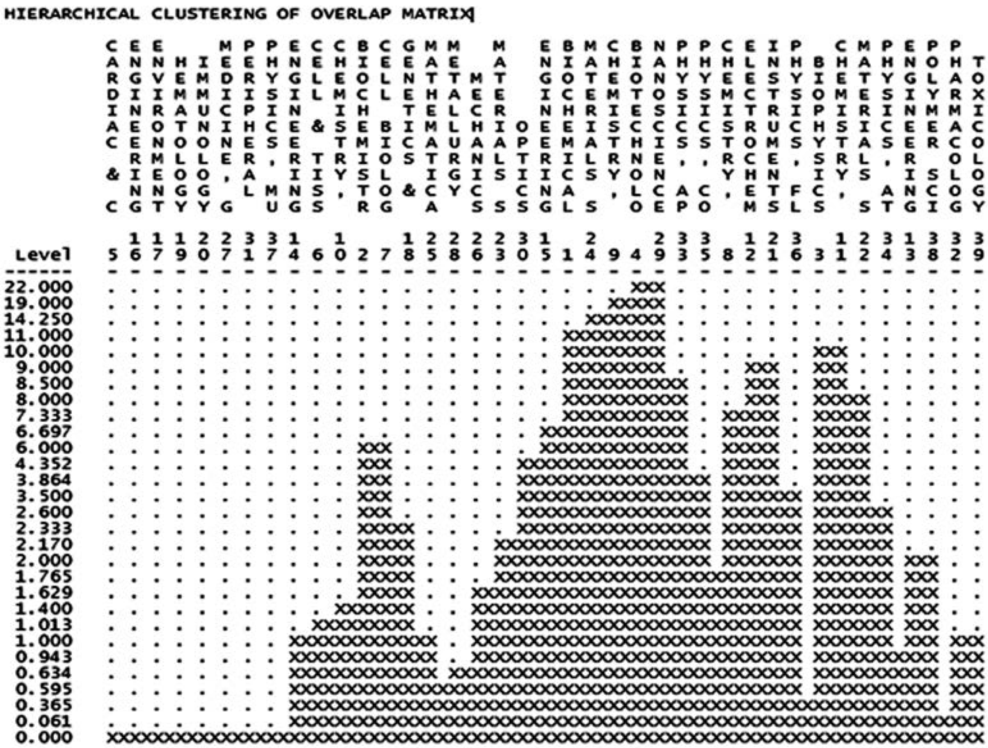


* The degree of research interdependence is represented by the size of the nodes, and the network ties indicate the strength of cross-disciplinary relations across the research fields.

For analyzing subgroups of cross-disciplinary relations, we look at clusters of the clique overlap matrix, suggesting a central / periphery structure in which the bridging research fields are at the

central, and other fields adjacent to them are in the semiperiphery (K-Plex analysis). Figure 3 contains an examination of the hierarchical clustering of the subgroups, revealing 206 maximal subgroups with some isolates in these relations. It is apparent that research fields *Nanoscience & Nanotechnology* and *Biotechnology & Applied Microbiology* were central in the sense of playing a bridging role among multiple slightly different cross-disciplinary relations in 22 of the 206 cliques. By contrast, *Engineering*, *Environmental & Environmental Sciences*; *Physics*, *Multidisciplinary*; and the field of medical sciences were the independent research areas not clustered with any of the other research fields in a given network of 2-plexes.

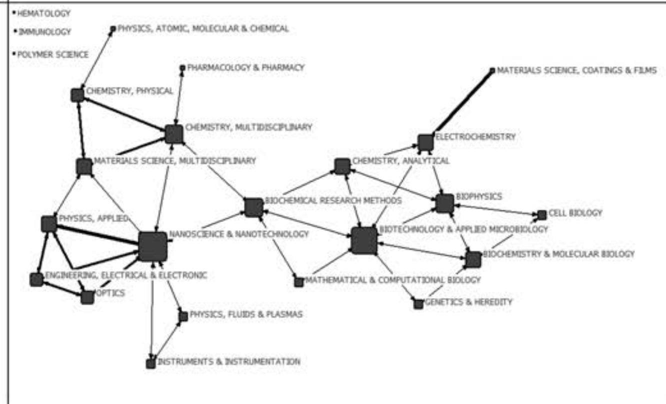
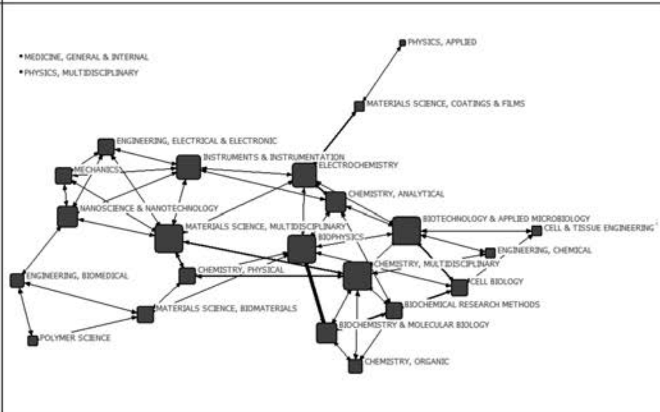
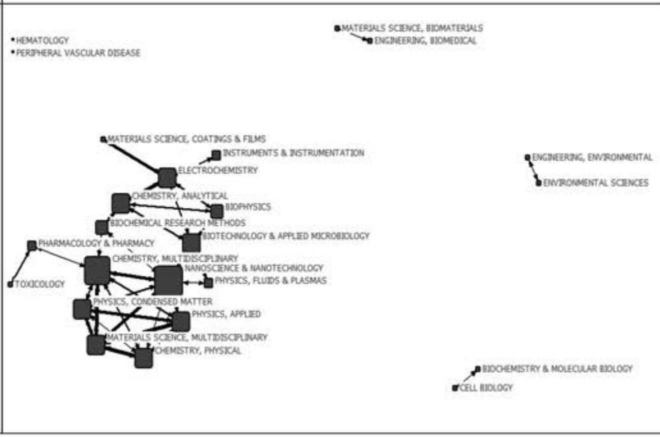
Figure 3. Clustering of subgroups.



The network centralization measures shown in Table 5 present a picture of the network structure of the cross-disciplinary research relation between research fields in different time periods. The types of networks are characterized by a different evolution of the cross-disciplinary relation structure. The analysis shows a trend of decentralization. The overall degree centralization – a measure of the spread of cross-disciplinarity across the network – was 12.65% in 2009, 6.23% in 2010, and 14.63% in 2011. The betweenness network centralization decreased from 34.51 to 13.56%, suggesting that the cross-disciplinary research network was formed without over

depending on intermediary research areas. In addition, the degree of inequality changed from 94.07% in 2009 to 67.63% in 2010 and 84.27% in 2011 of the maximum possible. A glance at figures in Table 5 shows that the overall network became less centralized, which was due to the cross-disciplinary research opportunities across various research fields. It suggests that as the network grew over time, cross-disciplinary relations were spread more widely across the research fields rather than dominantly driven by few major research fields.

Table 5. Network through cross-disciplinary relations for 2009–2011.

<table>
<tr><th></th><th>Network of cross-disciplinary relations</th><th>Network Centralization</th></tr>
<tr><td>2009</td><td>
</td><td>Degree = 12.65%
Betweenness = 34.51%
Eigenvector = 94.07%</td></tr>
<tr><td>2010</td><td>
</td><td>Degree = 6.23%
Betweenness = 14.73%
Eigenvector = 67.63%</td></tr>
<tr><td>2011</td><td>
</td><td>Degree = 14.63%
Betweenness = 13.56%
Eigenvector = 84.27%</td></tr>
</table>

Eigenvector = 84.27%

* The degree of research interdependence is represented by the size of the nodes, and the network ties indicate the strength of cross-disciplinary relations across the research fields.

Summary and conclusions

Using network analytical tools to examine collaborative and cross-disciplinary relations provides an important means to uncover topologies of cooperative research networks in a given cross-disciplinary S/ERC project (e.g., human sensing system). The top ranked sub-projects, *1(1)* and *1(2)*, were not only the leading sub-projects actively forming specific collaborative research relationships but also the intermediaries connecting other sub-projects through the improved dissemination of research activities. In addition, the three research fields that showed the most cross-disciplinarity were *Materials Science*, *Multidisciplinary*; *Chemistry*, *Multidisciplinary*; and *Nanoscience & Nonotechnology*. These were the central research fields that formed the cross-disciplinary research relationships that other research fields depended on.

The results shown clearly indicate that group research activities and their formation of research collaboration networks still depend on a framework that generates the right circumstances for cooperative and cross-disciplinary research activities. Each individual project and research field has a unique capability and specific knowledge assets that need to be carefully managed and monitored to accommodate its potential role in cross-disciplinary group research. Thus, when designing and managing group research support programs, research standards and evaluation criteria should avoid homogeneity by encouraging goal-oriented cross-disciplinary collaborative relations and by identifying potential benefits and costs of each project's and research field's specific expertise.

In this regard, investment in group research projects at university research centers should be made not only to maintain or foster highly competitive research fields but also to stimulate unexpected opportunities of cross-disciplinary research collaboration. As Yang et al (2010) noted, blind support for any particular research project or field could distort or discourage dynamic research interactions and capabilities of nurturing basic research and education in S&E for enhanced competitiveness. In addition, evidence-based evaluation baselines for assessing and exploring potential capabilities and competitiveness of group research activities need to be developed and implemented. By exploring potential vulnerabilities facing group research with regard to the interdependence among sub-projects and various research fields, capability assessments and performance measurements could enable program officials and/or S/ERC management to accurately assess current group research programs and design practical and feasible alternatives for achieving cross-disciplinary, multi-institutional research goals.

References

Aboelela, S.W., larson, E., Bakken, S., Carrasquillo, O., Formicola, A., Glied, S.A., Haas, J. & Gebbie, K.M. (2007). Defining Interdisciplinary Research: Conclusions from a Critical Review of the Literature. *Health Services Research*, 42(1), 329–346.

Altenkirch, R., Aung, W., Britocruz, C., Garside, J., Hinrichs, R., Johnson, W., Nieto, J., Roubicek, V., Thomas, G., Velez-Arocho, J., Wei, C. & Zielinski, W. (2003). What we must do together to increase quality and productivity. *Innovations: World Innovations in Engineering Education and Research, International Network for Engineering Education and Research* (1–11). Arlington, VA.

Boardman, C. & Corley, E.A. (2008). University Research Centers and the composition of research collaborations. *Research Policy*, 37, 900–913.

Boardman, C. & Gray, D. (2010). The new science and engineering management: cooperative research centers as government policies, industry strategies, and organizations. *Journal of Technology Transfer*, 35, 445–459.

Borgatti, S.P. (2002). *Netdraw network visualization*. Harvard, MA: Analytic Technologies.

Borgatti, S.P., Everett, M.G. & Freeman, L.C. (2002). *UCINET for Windows: software for social network analysis*. Harvard, MA: Analytic Technologies.

Bozeman, B. & Boardman, C. (2003). *Managing the New Multidiscipline, Multi-discipline University Research Center: Institutional Innovation in the Academic Community*. IBM Endowment for the Business of Government, Washington, D.C.

Bozeman, B. & Boardman, C. (2004). The NSF Engineering Research Centers and the University-Industry Research Revolution: A Brief History Featuring an Interview with Erich Bloch. *Journal of Technology Transfer*, 29, 365–375.

Klein, J.T. (1996). *Crossing boundaries: knowledge, disciplines, and interdisciplinarities*. Charlottesville, VA: University of Virginia Press.

Lattuca, L.R. (2003). Creating interdisciplinarity: Grounded Definitions from College and University Faculty. *History of Intellectual Culture*, 3(1), 1–20.

National Science Board (US). (2012). *Science and Engineering Indicators* 2012. Arlington Virginia: National Science Foundation (NSB 12–01).

Olson, J.S., Olson, G.M. & Hofer, E.C. (2005). What makes for success in science and engineering collaboratories?. In *Workshop on Advanced Collaborative Environments*.

Roessner, D., Manrique, L. & Park J. (2010). The economic impact of engineering research centers : preliminary results of a pilot study. *Journal of Technology Transfer*, 35, 475–493.

Rogers, E.M., Hall B., Hashimoto, M., Steffensen, M., Speakman, K.L. & Timko, M.K. (1999). Technology Transfer from University-Based Research Centers: The University of New Mexico Experience. *The Journal of Higher Education*, 70(6), 687–705.

Rogers, J.D. (2012). Research centers as agents of change in the contemporary academic landscape: their role and impact in HBCU, EPSCoR, and Majority universities. *Research Evaluation*, 21, 15–32.

Rosenfield, P.L. (1992). The potential of transdisciplinary research for sustaining and extending linkages between the health and social sciences. *Social Science and Medicine*, 35, 1343–1357.

Stokols, D., Hall, K.L., Taylor, B.K. & Moser, R.P. (2008). The Science of Team Science: Overview of the Field and Introduction to the Supplement. *American Journal of Preventive Medicine*, 35, 77–89.

Yang, C.H., Park, H.W. & Heo, J. (2010). A network analysis of interdisciplinary research relationships: the Korean government's R&D grant program. *Scientometrics*, 83, 77–92.

Citation impacts revisited: How novel impact measures reflect interdisciplinarity and structural change at the local and global level[1]

Michel Zitt Michel.Zitt@nantes.inra.fr / INRA-Lereco, Nantes (France)

Jean-Philippe Cointet Jean-Philippe.Cointet@polytechnique.edu / INRA-SenS, INRA, Marne-la-Vallée, France / Complex Systems Institute of Paris Ile-de-France (ISC-PIF), Paris (France)

Abstract

Citation networks have fed numerous works in scientific evaluation, science mapping (and more recently large-scale network studies) for decades. The variety of citation behavior across scientific fields is both a research topic in sociology of science, and a problem in scientific evaluation. Normalization, tantamount to a particular weighting of links in the citation network, is necessary for allowing across-field comparisons of citation scores and interdisciplinary studies. In addition to classical normalization which drastically reduces all variability factors altogether, two tracks of research have emerged in the recent years. One is the revival of iterative "influence measures". The second is the "citing-side" normalization, whose only purpose is to control for the main factor of variability, the inequality in citing propensity, letting other aspects play: knowledge export/imports and growth. When all variables are defined at the same field-level, two propositions are established: (a) the gross impact measure identifies with the product of relative growth rate, gross balance of citation exchanges, and relative number of references (b) the normalized impact identifies with the product of relative growth rate and normalized balance. At the science level, the variance of growth rate over domains is a proxy for change in the system, and the variance of balance a measure of inter-disciplinary dependences. This opens a new perspective, where the resulting variance of normalized impact, and a related measure, the sum of these variances proposed as a Change-Exchange Indicator, summarize important aspects of science structure and dynamism. Results based on a decade's data are discussed. The behavior of normalized impact according to scale changes is also briefly discussed. A shift towards a network-based definition of domains, more in the nomenclature-free spirit of citing-side normalization than database classification schemes, appears promising, albeit with technical challenges. An appealing issue is the connection with macro-level life-cycles of domains, and the dynamics of citation network.

1 The authors thank OST (S. Ramanana and G. Filliatreau) for help in providing aggregated data and bases of nomenclature. Modifications are the responsibility of the authors. They also thank E. Bassecoulard for her assistance.

Introduction

The use of citation measures in science (Garfield, 1955, 2006) is a controversial issue in research evaluation, as shown in the recurrent debates on impact factors[i]. Citations also shape a large-scale network (Price, 1965, Redner, 2005) which, along with collaboration, linguistic and web-communication networks, is a powerful tool for mapping science and understanding knowledge exchanges and self-organization of communities. A lasting issue is the variability of citation practices across fields, which prevents any sensible comparison between gross citation figures or h-indexes, say in mathematics vs. cell biology. A traditional way to deal with this variability is the normalization of citation figures based on fields baseline figures (Murugesan & Moravcsik 1976, Schubert & Braun 1986, Sen, 1992,. Czapski 1997, Vinkler, 2002, see also Raddichi et al., 2008). This "*ex post*" or "cited-side" statistical normalization is typically nomenclature-dependent, assuming an explicit delineation of scientific domains, usually from databases classification schemes. Forcing equality of cited domains, it sacrifices the consistency of the network and jeopardises multidisciplinary analysis. An alternative is the citing-side normalization ("*ex ante*", "source-level", "fractional citation"; Zitt & Small 2008, Moed 2010, Glänzel 2011, Leydesdorff & Opthof 2010, Waltman & van Eck 2012). The citing-side perspective (Zitt et al., 2005) is at the confluence of Garfield's insights on citation density (Garfield, 1979) and fractional weighting to reduce biases in cocitation mapping (Small & Sweeney, 1985). It corrects for the unequal propensity to cite amongst domains: in doing so, it keeps the best part of normalization – by removing undesirable sources of across-fields variability – while keeping the coherence of the citation network.

Especially, the partial normalization brought by the citing-side process, gives interpretable figures of domain-level average impact, which is true neither for usual "cited-side" normalized figures, forced to equality, nor for gross citation figures, blurred in magnitude by the effects of differential propensity to cite amongst fields. Focusing here on the analysis at the aggregate levels, we argue that citing-side approach opens new perspectives on interpretation of citation impacts at the domain level, and on structure and change of science insofar as it can be depicted by citation networks. We shall first establish two basic propositions on the decomposition of gross impacts and citing-side normalized impacts at the domain level. For the latter, we propose to summarize into a "Change-Exchange Index" (CEI) the variances over domains of its two factors at the domain level, growth rate and dependence. It may seem strange at first to come across a time-dependent variable such as growth, but the diachronic nature of citations implicitly carries information on change.

The CEI identifies with the variance of normalized impact when the two factors are independent, making the covariance term zero. We shall examine, on a decade's data on citation flows across science and a fixed nomenclature of domains, the empirical value of the variance and covariance terms calling for interpretations in terms of dynamics of the system, and discuss the challenges of shifting the nomenclature-based analysis to a bibliometric classification into topic/domains.

In section I, analytical, we shall state two propositions at the domain level: one on gross impacts,one on citing-side normalized impacts. Then considering the science level, we will define

the *CEI* and its relation with the normalized impact. Section II summarizes first results from an on-going empirical analysis on a decade of the Web of Science (hard sciences). The discussion section discusses several aspects: the shift from a database classification scheme (nomenclature) framework to a bibliometric classification of science; the relationship with various aspects of dynamics of science.

Analytical bases

If all variables are calculated at the same level of classification (whatever the level: for example the subject category) we get two basic propositions.

proposition 1: domain level, gross impact (not normalized)

The impact $I(A)$ of a domain A is defined as the average number of incoming citations per articles susceptible to get cited in A. If $\phi_{\leftarrow}(A)$ denotes the aggregated number of references citing domain A then : $I(A)=\frac{\phi_{\leftarrow}(A)}{|A|}$.

The growth rate $\rho(A)$ of a domain A is simply defined here by the ratio of publication volumes between the cited and the citing periods, volumes reduced to average volume over each period. We then introduce the balance which compares the total inflow of citation with total outflow emitted by A ($B(A)=\frac{\phi_{\leftarrow}(A)}{\phi_{\rightarrow}(A)}$). Finally we denote $\kappa(A)$ the average number of references in citing articles in A.

It is then straightforward to deduce the following equation[ii]:

$$I(A) = \rho(A)B(A)\kappa(A)$$

Equation 1

From this equation, it is useful to introduce the notation ^ transforming any domain level index into its relative version: the ratio to its science level counterpart. Given any domain-level measure $m(A)$ one can compute $\hat{m}(A) = \frac{m(A)}{m(S)}$.. Thus the relative impact $\hat{I}(A)$ is obtained by dividing the gross impact by the impact computed at the whole science level ($I(S)$). We will also denote $\hat{\rho}(A)$ the relative growth rate (i.e. ratio of growth rate by growth rate at the global science level) and $\hat{\kappa}(A)$ the relative number of references in citing articles in A.

$$\hat{I}(A) = \hat{\rho}(A)B(A)\hat{\kappa}(A)$$

Equation 2

proposition 2 – domain level: citing-side normalized impact

In order to neutralize the main source of variability, a normalization based on the relative number of active references (the "citing propensity") is introduced. It is implemented by weighting the

links of the original directed and unweighted citation network, with options fixing the granularity of the baseline. In a simple device, cited-side normalization weights links proportionally to average in-links by node within the citable set in a given domain's delineation while citing-side normalization weighs links proportionally to average out-links by node within the citing set in the domain. Those domains can be defined by some neighbourhood of the citing article: journal, cluster, or librarians/database categories. Here, for establishing basic propositions, we shall rely on subject categories as defined by Web of Science (Thomson Reuters). With such a weighting of the citation links it naturally appears that $\hat{\kappa}(A) = 1$. Neutralizing citing propensity variability then defines a new measure of impact which can be decomposed as:

$$\hat{I}_g(A) = \hat{\rho}(A)B_g(A)$$

Equation 3

These propositions generalize previous results on the journal impact factor (Zitt, 2011).

proposition [3] – science level: the deviation of citation impacts.

If the domain-level normalized impact is the product of two relative measures linked to interdisciplinary structure (asymmetry of exchange) and local dynamism (relative growth), what can we learn at the science level? All measures being relative, the signs of change are expected in the deviation indexes.

For a particular category A at a given level of breakdown $\hat{I}_g(A) = \hat{\rho}(A)B_g(A)$. With logarithmic transformation of variables, suggested by the distribution of impacts at the domain level:

$LI(A) = LG(A) + LB(A)$ where LI, LG, LB designate respective logs of normalized impact, growth rate and normalized balance. Over all domains:

$$\text{wVar}(LI) = \text{wVar}(LG) + \text{wVar}(LB) + 2\text{wCov}(LG, LB)$$

Equation 4

where wVar stands for variance weighted by the volume of publications of domains, expressed in number of publications. For comparison sake, the unweighted variance has also been used.

In Equation 4 the variance terms have a simple interpretation. wVar(LB) over domains is a proxy of global interdisciplinary dependences in the system, and wVar(LG) is a proxy for the intensity of "creative destruction" through differentiation of growth rates over domains. A scientific system where domains do not exchange and are in steady state will associate zero variance and covariance terms, giving a zero variance of impacts. At the opposite end, a scientific system combining a high proportion of growing and declining domains and a strongly asymmetrical balance of flows across fields (exporters and importers) will show a high level of variance terms, but the final value of wVar(LI) will also depend on the covariance term.

The relationship between growth and balance partly depends on the superposition of domains at various stages of their life-cycle, while the potential value of balance for individual domains, typically reached at maturity stages, can show great dispersion linked to the position of the domain in the cognitive chain. The variance of balance (compared to growth's) may play a dominant role in the shaping of impact dispersion. Domains in emergence both grow rapidly and are quite dependent on imports of knowledge/ information from their parent fields. Hence they are likely to enhance the variance of growth, and to yield negative covariance

In order to summarize asymmetry and growth effects, we propose then to consider only the sum of variance terms, the "structural-change and exchange-asymmetry index", abridged into Change-Exchange Index *CEI*

$$CEI = \mathrm{wVar}(LG) + \mathrm{wVar}(LB)$$

Equation 5

This index is closely related to the variance of impacts with $CEI = \mathrm{wVar}(LI) - 2\mathrm{wCov}(LG, LB)$

CEI is trivially equal to the variance of impacts if growth rate and balance are independent.

scale issues

If the level of calculation of impact and the level of normalization (at which balances and growth rate are computed) are different, factors of scale come into play. In such configurations where the level of definition of impacts and of other variables are not homogeneous (which is the case in many practical uses of normalization), the relations above should be altered by a correcting factor for the balance. This aspect is not detailed in the present submission. Scale irregularities in standard (cited-side) normalization had also already been stressed by Zitt, Ramanana, Bassecoulard (2005).

A first experiment within a fixed nomenclature

Data are based on OST aggregate figures at the category level, based on primary data and subject categories from the Web of Science (Thomson Reuters).

The citation framework is based on "cited years", 5-years citation window, on the period 1999–2010, giving an exploitable span 1999–2006 and with caution through 2008 (with reduced but acceptable citation window). In the database (OST-WoS), there are overlaps in assignment of journals and then papers to categories (WoS subject categories) at the lower level, handled by fractional counting. The nomenclature at the sub-discipline and the discipline level is derived from OST scheme, modified for simplicity sake, in order to get an embedded scheme disciplines/subdisciplines/categories.

In present version, we limited ourselves to data for hard sciences. Results including social sciences and humanities are reported elsewhere[iii]. The global results are quite similar, but trends are very slightly different. Let us summarize the main results.

Gross impact: As expected the gross impact heavily depends on the propensity to cite. The variance of the impact is essentially shaped by the variance of this factor, which jeopardizes any interpretation of its variance in terms of balance – the variance of which is by an order of magnitude lower – and growth, still behind.

Normalized impact: As soon as citing propensity is corrected, a new avenue is open to interpretation of citation impacts, in terms of dynamism and asymmetrical interdisciplinarity. Fig. 1A, B, C shows the time series of variance (weighted variance) of normalized impact, of its factors growth and balance, of the covariance term, and the series of CEI. A couple of striking points:

- the respective role of the two factors: within this citation window, the influence of growth is small, the CEI is mostly shaped by the asymmetry of exchanges. However, the dominant role of balance increases with the level of aggregation. In average over years 1999–2006 (hard sciences), the balance accounts for 89% of the CEI (ratio balance /relative growth about 8.35) at the category level, 91% (ratio 11.05) at the sub-discipline level, 93% (ratio 17.06) at the discipline level. With respect to the reduction of all variances in the aggregation process, it appears that the smoothing effect is stronger for differential growth than for balances. We cannot expand on this aspect in the present text.

  In terms of trend, the dispersion of balance slightly declines in the period 2000–2005 with a sign of reversal from 2006 on, remembering that information is partial for further years (citation window). The reversal is stronger at the category level. The dispersion of growth reaches a maximum in 2000 with a very slight down-trend in the next years.

- the covariance of growth and balance: covariance is almost always negative over the period, however get closer to the zero value at the end of the period. This trend is seen whatever the level of aggregation.

- at the sub-discipline and discipline level, the variance of impacts is less affected than the CEI by the down-trend, since the increase of covariance (from fairly negative to weakly negative) compensates for the reduction of variances of growth and balance.

- As far as an analysis based on a fixed nomenclature can be trusted, there is no sign of increasing differentiation of growth rates over the period, nor of increasing asymmetry in the system, whatever the scale.

Figure 1a – hard sciences – category level

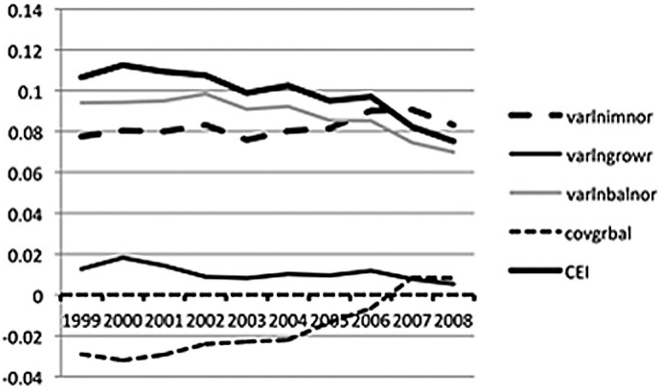


Figure 1b – hard sciences – sub-discipline level

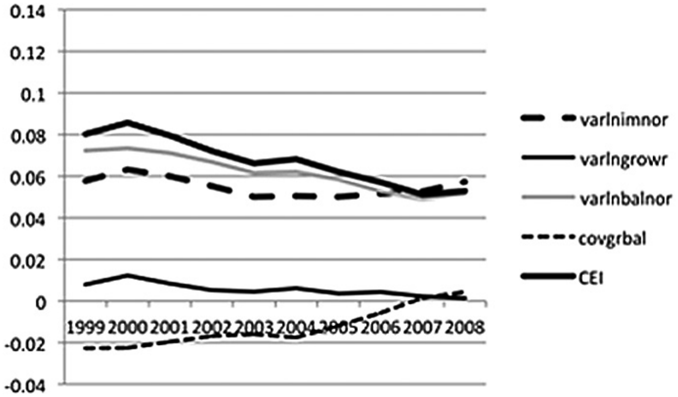


Figure 1c – hard sciences – discipline level

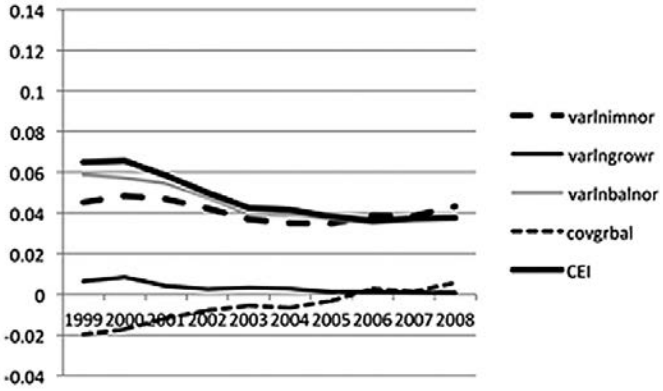


Fig. 1 – time series of normalized impact and CEI – variances and covariance
varlnimnor: variance of logarithm of normalized impact / varlngrowr: variance of logarithm of relative growth / varlnbalnor: variance of logarithm of citation balance / covgrbal: covariance (logs) growth-balance / CEI: Change-Exchange Index

Discussion

We emphasized that citation impacts corrected for citation propensity identify, at a given domain level, with the product of growth and export-import balance; over all these domains (science), the variance of those factors are markers of differential growth and of interdisciplinarity in a particular form, the citation dependence. These properties suggest that the variance of normalized impacts on the one hand, and a related index, the Change-Exchange Index, on the other hand, help to partly characterise structural dynamism of the scientific system at a particular level of aggregation. Without scale correcting factor, those measures hold iff all variables involved are defined at the same level of nomenclature (identical level for the calculation of impact and the normalization). Of course, practical applications of citing-side normalization for citation scoring don't have to comply with this condition.

From nomenclature to clusters

The experiment was conducted on a fixed nomenclature. This is clearly a limitation, especially when one remembers the nomenclature-free spirit of citing-side normalization, quite helpful in practical applications. Nomenclatures such as databases classification schemes suffer shortcomings: artefacts in the delineation of categories, low reactivity in the short term, sensitivity to national context. An alternative is to rely on networks, especially citations, with various transformations (Kessler 1963, Small 1973, Marshakova 1973, Chen 1999, Boyack, Klavans & Börner, 2005), which proved powerful tools for clustering and mapping science. Clustering reduces tensions by adjusting delineation of domain and trading topics. Substituting clusters/ neighborhoods to categories is therefore expected to yield more realistic representations, minimizing artificial exchange flows. Bibliometric clusters (co-authorship, citations, semantic content) enable scholars to track emergence and life-cycle phenomena (Scharnhorst et al., 2012; Morris, 2005; Chavalarias & Cointet, 2013). If citation approach is preferred, which is logical in our context, a sensible objective is to reach bipartite clusters encompassing both cited and citing items in close relation.

A challenge of network-based clustering is the loss of coverage, in contrast with nomenclature schemes or classifications based on editorial entities (journals, see the "audience factor"): in the latter, any citable article is classified, whether cited or not; any citing article is classified, whether its references are "active" (falling into the citation window) or not. There are various ways – easier for citing than for citable articles – to circumvent the problem on pools or clusters (Waltman & van Eck, 2012; see also on SNIP 2: Waltman et al., 2013).

It should be recalled that unlike the conventional cited-side normalization, the citing-side approach does not aim at a complete normalization. Usual quality tests assessing the performance of the various methods on the ground of the total reduction of variability are not therefore completely convincing. By limiting itself to the correction of propensity to cite, the citing-side approach is shown perhaps more fruitful.

Normalized Impact, Change-Exchange Index and dynamics of science

Further research is needed to explore the various aspects of these measures. A general issue is the linkage between macro and micro-models. The sign of growth-balance covariance, all things equal, may change over different phases in a domain's life-cycle. In this experiment (and with the limitations of database's scheme), it is typically negative in emergence phases, with a strong growth rate and a heavy dependence on external knowledge. Examples of such categories are computational biology, nanoscience and technology, integrative medicine, remote sensing. At the opposite, a lot of established categories (cell biology, embryology, genetics, immunology) exhibit a weak relative growth along with a strong position of exporter of knowledge. Very few categories associate strong positive deviations on both variables, such as biomaterials and development biology. A challenge is to connect model of life-cycle of areas, preferably delineated by citation-based clustering, with various mechanisms of networks dynamics (Powell et al. 2005), among them preferential attachment (Price 1963, Jeong et al., 2007, Eom et al., 2011).

Interdisciplinarity is only seen, here, through asymmetrical linkages: domains equalizing exports and imports of knowledge will tend to reduce the dispersion. Diversity, essential for understanding the science structure and dynamics, is not directly studied, but citing-side weighting of the network, beforehand, improve measures of diversity and more generally of interdisciplinary flows (Zitt, 2011, Rafols et al., 2012), with a significant improvement over gross flows analysis (e.g. Rinia et al., 2002). It should also be stressed that only relative changes were addressed here, by using relative variables. The absolute growth or the average impact over science is corrected for, in contrast with long-range analyses in the wake of Price (1963) which focus on volumes of publications and citations (see for example Larivière et al., 2008 in their study of aging).

To conclude, citing-side approach opens a new perspective for the analysis of knowledge flows, insofar as they can be sketched by citation networks. Here, at a macro-level, we have shown that basic relations connect the novel normalized impact and a derived measure, the CEI, to important features of dynamism and structure of science. The relation with the parallel and powerful "influence weighting" pioneered by Narin & Pinski (1976) with iterative weighting of citation sources, that has known a revival in the last decade (Palacio-Huerta & Volij, 2004, Bergstrom, 2007, whose algorithm claims correction for citing density) is also appealing.

References

Bergstrom, C. (2007). Eigenfactor: measuring the value and prestige of scholarly journals. *College & Research Libraries News*, 68, 5, www.ala.org/ala/acrl/acrlpubs/crlnews/backissues2007/may2007/eigenfactor.cfm.

Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, 64(3), 351–374.

Chavalarias D, Cointet J-P (2013). Phylomemetic Patterns in Science Evolution – The Rise and Fall of Scientific Fields. *PLoS ONE* 8(2): e54847. doi:10.1371/journal.pone.0054847.

Chen, C. M. (1999). Visualising semantic spaces and author co-citation networks in digital libraries. *Information Processing & Management*, 35(3), 401–420.

Cronin, B. (1984). *The citation process; the role and significance of citations in scientific communication*. London: Taylor Graham.

Czapski, G. (1997). The use of deciles of the citation impact to evaluate different fields of research in Israel. *Scientometrics*, 40(3), 437–443.

Garfield, E. (1955) Citation Indexes for Science. A new dimension in documentation through association of ideas. *Science*, 122(3159), 108–111.

Garfield, E. (1979). Citation Indexing. Its theory and applications in science, technology and humanities. New York: Wiley.

Garfield E. (2006). The history and meaning of the journal impact factor, *JAMA* 295(1), 90–93.

Glänzel, W., Schubert, A., Thijs, B., & Debackere, K. (2011). A priori vs. a posteriori normalization of citation indicators. The case of journal ranking. *Scientometrics*, 87(2), 415–424.

Jeong, H, Néda, Z, & Barabasi, A-L. (2003). Measuring preferential attachment in evolving networks. EPL (*Euro-Physics Letters*), 61(4), 567–572.

Kessler M.M. (1963). Bibliographic coupling between scientific papers, *American Documentation*, 14 (1), 10–25.

Larivière V., Archambault E., Gingras Y. (2008) Long-Term Variations in the Aging of Scientific Literature: From Exponential Growth to Steady-State Science (1900–2004), *Journal of the American Society for Information Science*, 59(2): 288–296.

Leydesdorff, L., & Opthof, T. (2010). Scopus's Source Normalized Impact per Paper (SNIP) versus a Journal Impact Factor based on Fractional Counting of Citations. *Journal of the American Society for Information Science & Technology*, 61(11), 2365–2396.

Marshakova, I. V. (1973). System of Document Connections Based on References, *Scientific and Technical Information Serial of VINITI*, 6(2),3–8

Merton R.K.(1968). The Matthew effect in science, *Science*, 159 (3810) 56–63.

Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, 4(3), 265–277.

Morris S.A.(2005). Manifestation of emerging specialties in journal literature: a growth model of papers, references, exemplars, bibliographic coupling, co-citation, and clustering coefficient distribution, *Journal of the American Society for Information Science*, 56(12), 1250–1273.

Murugesan, P., & Moravcsik, M. J. (1978). Variation of the nature of citation measures with journal and scientific specialties. *Journal of the American Society for Information Science*, 29(3), 141–155.

Palacio-Huerta, I., & Volij, O. (2004). The Measurement of Intellectual Influence. Econometrica, 72(3), 963–977.

Pinski, G., & Narin, F. (1976). Citation influence for journal aggregates of scientific publications: theory, with application to the literature of physics. *Information Processing and Management*, 12(5), 297–312.

Powell W.W., White D.R., Koput K.W., Owen-Smith J. (2005)Network Dynamics and Field Evolution: The growth of interorganizational collaboration in the Life Sciences, *American Journal of Sociology*, 110(4), 1132–1205.

Price D.J. de Solla (1963). Little science, big science, Columbia Univ. Press.

Price D.J. de Solla (1965). Networks of scientific papers, *Science*, 149(3683), 510–515.

Raddichi F., Fortunato S., Castellano C. (2008). Universality of citation distributions: Towards an objective measure of citation impact, *PNAS*, 105(45), 17268–17272.

Rafols I, Leydesdorff L., O'Hare A., Nightingale P., Stirling A. (2012) How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management, *Research Policy* 41, 1262– 1282.

Rinia E.D., Van Leewen T.N., Bruins E.W., van Vuren H.G., van Raan A.F.J., (2002) Measuring knowledge transfer between fields of science, *Scientometrics* 54, 3, 347–362.

Scharnhorst A., Börner K., van den Besselaar P., (eds., 2012). Models of science dynamics: Encounters between complexity theory and information sciences, Springer.

Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics*, 9(5–6), 281–291.

Sen, B. K. (1992). Normalized Impact Factor. *Journal of Documentation*, 48(3), 318–325.

Small, H., & Sweeney, E. (1985). Clustering the Science Citation Index using co-citations I – a comparison of methods. *Scientometrics*, 7(3–6), 391–409.

Small, H. (1973). Co-citation in the scientific literature : A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265–269.

Small, H., & Sweeney, E. (1985). Clustering the Science Citation Index using co-citations I – a comparison of methods. *Scientometrics*, 7 (3–6), 391–409.

Vinkler, P., (2002). Subfield problems in applying the Garfield (Impact) Factors in practice, *Scientometrics*, 53(2), 267–279.

Waltman L., Van Eck N.J., (2012). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison, *arxiv.org*/pdf/1208.6122.

Waltman L., Van Eck N.J., van Leeuwen T.N. & Visser M.S.(2013), Some modifications to the SNIP journal impact indicator, *Journal of Informetrics* 7 (2), 272–285.

Zitt, M., Ramanana-Rahary, S., & Bassecoulard, E. (2005). Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation. *Scientometrics*, 63(2), 373–401.

Zitt, M., & Small, H. (2008). Modifying the Journal Impact Factor by Fractional Citation Weighting: the Audience Factor. *Journal of the American Society for Information Science and Technology* 59(11), 1856–1860.

Zitt, M. (2010). Citing-side normalisation of journal impact: A robust variant of the Audience Factor. *Journal of Informetrics*, 4(3), 392–406.

Zitt M. (2011). Behind citing-side normalisation of citations: Some properties of the journal impact factor, *Scientometrics* 89(1), 329–344.

---

i in recent literature, see the dedicated issue of Scientometrics 92(2), (2012)

ii detailed proofs are shown in Arxiv document http://arxiv.org/abs/1302.4384

iii The extensive nomenclature is given in Appendix II of the Arxiv document http://arxiv.org/abs/1302.4384

Fortresses of books: A bibliometric foray into ranking scholarly publishers

Alesia Zuccala a.a.zuccala@uva.nl / Institute for Logic, Language and Computation, University of Amsterdam, P.O. Box 94242, Amsterdam, 1090 GE (The Netherlands)

Raf Guns raf.guns@ua.ac.be / University of Antwerp, IBW, City Campus, Venusstraat 35, B-2000 Antwerpen (Belgium)

Rens Bod rens.bod@uva.nl / Institute for Logic, Language and Computation, University of Amsterdam, P.O. Box 94242, Amsterdam, 1090 GE (The Netherlands)

Abstract

This research makes use of a unique dataset comprised of Scopus journal citations to books, in the field of *History*, and WorldCat® international library catalogue records of university presses/ commercial publishers. With this dataset, we take two approaches to ranking and indicating the prestige of book publishers. First we compiled all citations to a selection of books that were published between 2001 and 2006, and cited in journal articles published from 2007 to 2011. We matched the cited book titles to known publishers (recorded in WorldCat®), and obtained a basic ranking from all counts. Comparisons were then made between total citation counts for every publisher and mean citation per book (CPB) for every book title. To complement this ranking, we employed a network and mapping approach, which allows us to observe directed links between journal titles and book publishers. The results of our analyses are preliminary, but show that *Oxford University Press* and *Cambridge University Press* are (as expected) the most prestigious – i.e., highly centralized on the basis of directed network links – and rank highest also with respect to total citations. However, in terms of mean citation rate (CPB), the top ranking publishers are: *Belknap*, *Harvard University Press*, *Princeton University Press* and the *University of North Carolina Press*. Several commercial publishing houses are also in the top ranked 50, including *Basic Books*, *A. A. Knopf*, and *WW Norton & Co*.

Introduction

Bibliometricians do not know very much about academic book publishers. Unlike journals, which have been studied quite intensively, book publishers are a bit like fortresses; shadowy strongholds within the scholarly communication system. Researchers (especially Humanities scholars) rely on them to print 'the book' that will lead to their career promotion and tenure

1 The authors are grateful to both the Elsevier Bibliometrics Research Programme (http://ebrp.elsevier.com/) and OCLC WorldCat® for granting access to the data that were used to create the unique database required for this study.

(Cronin & La Barre, 2004; Williams et al., 2009). Universities are eager to promote their presses as great additions to their scholarly reputation, and evaluators are curious about whether or not a publisher's authority is equal to a proven measure of quality. Publisher 'quality' has previously been assessed by survey studies (Goodson et al., 1999; Lewis, 2000; Metz & Stemmer, 1996) and it has also been related to library catalogue holdings (Donovan & Butler, 2007; Torres-Salinas & Moed, 2009; White et al., 2009; Zuccala & Guns, 2013). The concept of 'prestige' is typically associated with academic research advice and the selection of a university press for a new book (Pasco, 2002; Pratt, 1993; Rowson, 1995). Often it is also linked to the age of the press and its original mission, including the rank of its housing university (Gump, 2006).

The objective of this paper is to determine how publisher 'prestige' can be indicated quantitatively using bibliographic citations. One can find lists on the Web that rank publishing houses in terms of A, B, C, and D categories based on international library holdings (e.g., SENSE, 2009). Recently, Garand and Gilles (2011) also developed a ranking system using a combination of open-ended surveys with readers and closed-ended lists of scholarly presses for the subject of Political Science. Other, somewhat 'general' ranking procedures give more attention to publisher sales and yearly revenues (e.g., Frankfurt Special, 2010; Publishers Weekly, 2012, June 25). In this study, we focus on ranking book publishers, both university presses and commercial houses, using a newly constructed Scopus-WorldCat® citation-'libcitation' relational database. Here we share some preliminary statistics produced from our unique datasets and highlight critical issues related to the standardisation of publisher names. Our first analysis is a ranking of presses/publishing houses based on raw citation counts and mean citations per book. In the second part of this research, we present an exploratory network analysis based on directed citation links between Scopus *History* journals (i.e. category code 1202) and book presses/publishing houses. The directed network serves as a complement to the ranking procedure, by demonstrating how publishers and journals tend to cluster on the basis of geographical regions/languages and sub-disciplines within the broader field of *History*.

Relational datasets

One drawback to preparing a study of cited books and press 'prestige' relates to the structure of current databases. Both the Thomson Reuter's Web of Knowledge and Elsevier's Scopus journal indices include cited books in tagged reference lists; however, each book lacks a distinct source identification code. The researcher is forced to grapple with a unit of information known as the reference 'string' [e.g., *Runge L*, *2005*, *Companion 18 Century, p292*]. Sometimes the book title appears in short form, and sometimes it is recorded in full, but almost always the publisher name is omitted. Moreover, both the author and book title itself can be recorded inconsistently from article to article depending on the person who made the original reference/citation. Thomson Reuters is in the process of rectifying this problem with its new Book Citation Index (BKCI). Since we do not have access to the BKCI, we only give attention to citations to books coming strictly from journal articles.

Table 1. Cited documents from Scopus History *and* Literature *matched in WorldCat®.*

	All Cited Docs	Sourced in Scopus only	Not in Scopus, but Matched in WorldCat®	In Scopus & Matched in WorldCat®	Not in Scopus or WorldCat®	Cited Docs w. Missing Values (?)
History						
1996–2001	882,155	6,945	303,048	368	564,773	7,021
2007–2011	2,858,005	117,789	**806,985**	2,251	1,915,002	15,978
Literature						
1996–2000	198,606	815	75,840	139	120,445	1,367
2007–2011	1,395,917	36,737	504,721	1,546	845,561	7,352

In June 2012 our project team constructed a unique Microsoft SQL database, with citation data from Elsevier Scopus journals (*History* and *Literature & Literary Theory* for the periods *1996–2001* and *2007–2011*) as well as metadata from the OCLC-WorldCat® International Library Catalog. All Scopus citation data were granted to us through the 2012 Elsevier Bibliometrics Research Program. With an API developer key we performed thousands of queries in the OCLC-WorldCat® to match book reference strings to their original publisher information (i.e., publisher name and location). With each book title matched and identified by a new OCLC source code, we also obtained corresponding library holding counts (i.e., 'libcitations') for the Association of Research Libraries (ARL) and non-ARL libraries. Table 1, above, presents the results of this matching procedure.

Publisher ranking

In our pilot analysis, we focused specifically on book titles cited in Scopus *History* Journals and matched in WorldCat® (see highlighted cell in Table 1). Here we have citation counts as well as 'libcitation' counts for 806,985 titles. We have identified the citations as being 'books' (i.e., monographs or edited volumes) because all were catalogued as such in one or more international libraries. From this original set, we then selected a much smaller sample of book titles (n=68,474) that were published between 2000 and 2006, and cited in a Scopus *History* journal between the period of 2007 and 2011.

Data cleaning/standardisation started at the level of the cited book (i.e., ensuring that all reference strings and corresponding citation counts were to individual books) and was followed by a process of standardising all publisher names. For every publisher name it was necessary to clarify, for example, that "*Oxford Univ. Press*" and "*OUP*", as well as "*Oxford U. Press*" were equal to the full name "*Oxford University Press*". Many publishers were not always recorded in their singular form, for instance: "*Oxford University Press for the British Academy*" or "*Polity Press in association with Blackwell*". Here we decided to accept and include all cases that were "*in association with*" another

publishing body or published "*for*" another organisation. We also did not alter the names of publishers that shifted from old names to new names (e.g., *J. Wiley & Sons merged with Blackwell publishing to become Wiley-Blackwell*) and we maintained all imprints of a larger publishing name (e.g. *Scribner is an imprint of Simon & Schuster*). The most difficult names to standardise were those written in a non-English language and transcribed in a non-Latin script (e.g., *ROSSPEN-Rossiiskaia Politicheskaia Entsiklopediia*).

Figure 1. Distribution of citations by publisher type
(Books published 2000–2006 and cited in Scopus History Journals 2007–2011).

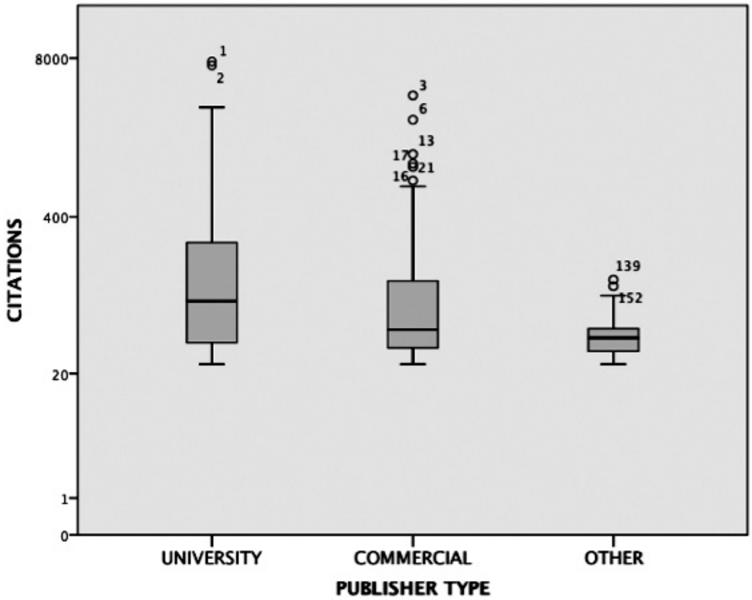


Table 2. Top 50 publishers based on raw citations and citations per book (CPB) in Scopus History *Journals (2007–2011). (Commercial publishing houses printed in italics)*

	PUBLISHER NAME	Raw Citation Counts	Mean CPB		PUBLISHER NAME	Raw Citation Counts	Mean CPB
1	Cambridge University Press	7510	3.118	1	*Belknap Press*	633	6.750
2	Oxford University Press	6992	2.972	2	Princeton University Press	3177	4.586
3	*Routledge*	3975	2.152	3	Harvard University Press	2184	4.334
4	Princeton University Press	3177	4.586	4	University of North Carolina Press	1455	4.320
5	University of California Press	2516	3.974	5	University of Chicago Press	2443	4.039
6	*Palgrave Macmillan*	2512	2.097	6	University of California Press	2516	3.974
7	University of Chicago Press	2443	4.039	7	Duke University Press	1623	3.803
8	Harvard University Press	2184	4.334	8	*Verso Books*	541	3.699
9	Yale University Press	1885	3.316	9	Johns Hopkins University Press	1128	3.397
10	Duke University Press	1623	3.803	10	Cornell University Press	1368	3.359
11	University of North Carolina Press	1455	4.320	11	*Basic Books*	418	3.357
12	Cornell University Press	1368	3.359	12	Yale University Press	1885	3.316
13	*Brill Academic Publishers*	1311	1.908	13	University of Pennsylvania Press	857	3.293
14	Stanford University Press	1128	3.190	14	Stanford University Press	1128	3.190
15	Johns Hopkins University Press	1128	3.397	15	*Penguin Group*	630	3.166
16	*Ashgate Publishing Company*	1097	1.789	16	Cambridge University Press	7510	3.118
17	*Blackwell Publishing*	1030	2.428	17	Oxford University Press	6992	2.972
18	MIT Press	910	2.585	18	*Alfred A Knopf*	670	2.913
19	University of Pennsylvania Press	857	3.293	19	*WW Norton & Co*	714	2.797
20	Columbia University Press	834	2.518	20	*Polity Press*	432	2.719
21	*Clarendon Press*	797	2.566	21	MIT Press	910	2.585
22	*WW Norton & Co*	714	2.797	22	*Clarendon Press*	797	2.566
23	Manchester University Press	698	2.520	23	Manchester University Press	698	2.520
24	*Alfred A Knopf*	670	2.913	24	Columbia University Press	834	2.518
25	*SAGE Publications*	647	1.982	25	*Berg Publishers*	470	2.438
26	*Belknap Press*	633	6.750	26	*Blackwell Publishing*	1030	2.428
27	Indiana University Press	631	2.418	27	Indiana University Press	631	2.418
28	*Penguin Group*	630	3.166	28	University of Michigan Press	475	2.412
29	University of Minnesota Press	592	2.390	29	University of Minnesota Press	592	2.390
30	*Verso Books*	541	3.699	30	*Pearson Longman*	440	2.349
31	University of Toronto Press	494	2.160	31	*Viking Press*	416	2.339
32	University of Michigan Press	475	2.412	32	University of Illinois Press	367	2.218
33	*Berg Publishers*	470	2.438	33	University of Toronto Press	494	2.160
34	*Pearson Longman*	440	2.349	34	*Routledge*	3975	2.152
35	*Polity Press*	432	2.719	35	*Palgrave Macmillan*	2512	2.097
36	*CH Beck Verlag*	425	1.981	36	New York University Press	406	2.095
37	*IB Tauris Publishers*	422	1.985	37	*Vandenhoeck & Ruprecht Unipress*	383	2.077
38	*Basic Books*	418	3.357	38	*Boydell Press*	322	2.073
39	*Viking Press*	416	2.339	39	University of Texas Press	339	2.059
40	New York University Press	406	2.095	40	McGill-Queen's University Press	339	2.048
41	*Peter Lang*	404	1.280	41	Rutgers University Press	313	2.006
42	*Rowman & Littlefield Publishers Inc*	391	1.527	42	University of Nebraska Press	333	1.994
43	*Vandenhoeck & Ruprecht Unipress*	383	2.077	43	*IB Tauris Publishers*	422	1.985
44	*HarperCollins Publishers*	368	1.643	44	*SAGE Publications*	647	1.982
45	University of Illinois Press	367	2.218	45	*CH Beck Verlag*	425	1.981
46	University of Texas Press	339	2.059	46	*Brill Academic Publishers*	1311	1.908
47	McGill-Queen's University Press	339	2.048	47	*Ashgate Publishing Company*	1097	1.789
48	University of Nebraska Press	333	1.994	48	*HarperCollins Publishers*	368	1.643
49	*Boydell Press*	322	2.073	49	*Rowman & Littlefield Publishers Inc*	391	1.527
50	Rutgers University Press	313	2.006	50	*Peter Lang*	404	1.280

When we finalized all citation counts linked to standardised publisher names, we assigned the top 500 with the most citations to the following categories: a) *university* (includes institutes, centres, schools, departments); b) *commercial*, and c) *other* (includes libraries, associations, museums,

foundations etc.). Figure 1, above, shows the distribution of these 500 publishing houses within their respective categories. Note that the *university* presses possess the strongest range of general 'citedness', with two prominent outliers: *Oxford* and *Cambridge*. In the *commercial* category there is a larger cluster of outliers; thus indicating that a very specific group tends to produce the most highly cited books (i.e., by History scholars). Overall, both university presses and commercial publishers display skewed citation counts.

Table 2, above, lists the top 50 publishers based on raw citation frequencies and mean citation per book (CPB). As expected, both Oxford University Press and Cambridge University Press are ranked at the top in terms of raw counts, and note that almost half from the list (n = 24) are commercial publishers. When we examine this ranking based on the CPB for each press/publisher, *Cambridge* and *Oxford* move to 16th and 17th place and the new highest ranked publishers then become *Belknap*, *Princeton University Press*, *Harvard University Press*, and *University of North Carolina Press*. Note that due to our data collection methods the CPB does not take uncited books by a particular press into account.

Directed journal-to-publisher network

Our directed network approach to evaluating publisher prestige involved using two mapping tools: 1) VOSViewer (van Eck & Waltman, 2010) and 2) Pajek (de Noy et al, 2005). Both tools enabled us to further explore the relationship between the Scopus *History* journal set and the various presses/publishing houses cited by individual papers throughout 2007 to 2011. The network arcs are directed and include a selection of the top 501 strongest citation links. Altogether the map includes a total of 354 international journals and 147 of the most frequently cited international book publishers. *Cambridge* and *Oxford University Presses* are recipients of the most in-links from this 354 journal set; however the bottom threshold for link strength was set at $n = 10$, so that many smaller publisher houses would be featured.

Figure 2. Scopus History *Journals linked to Publishers (16 clusters).*

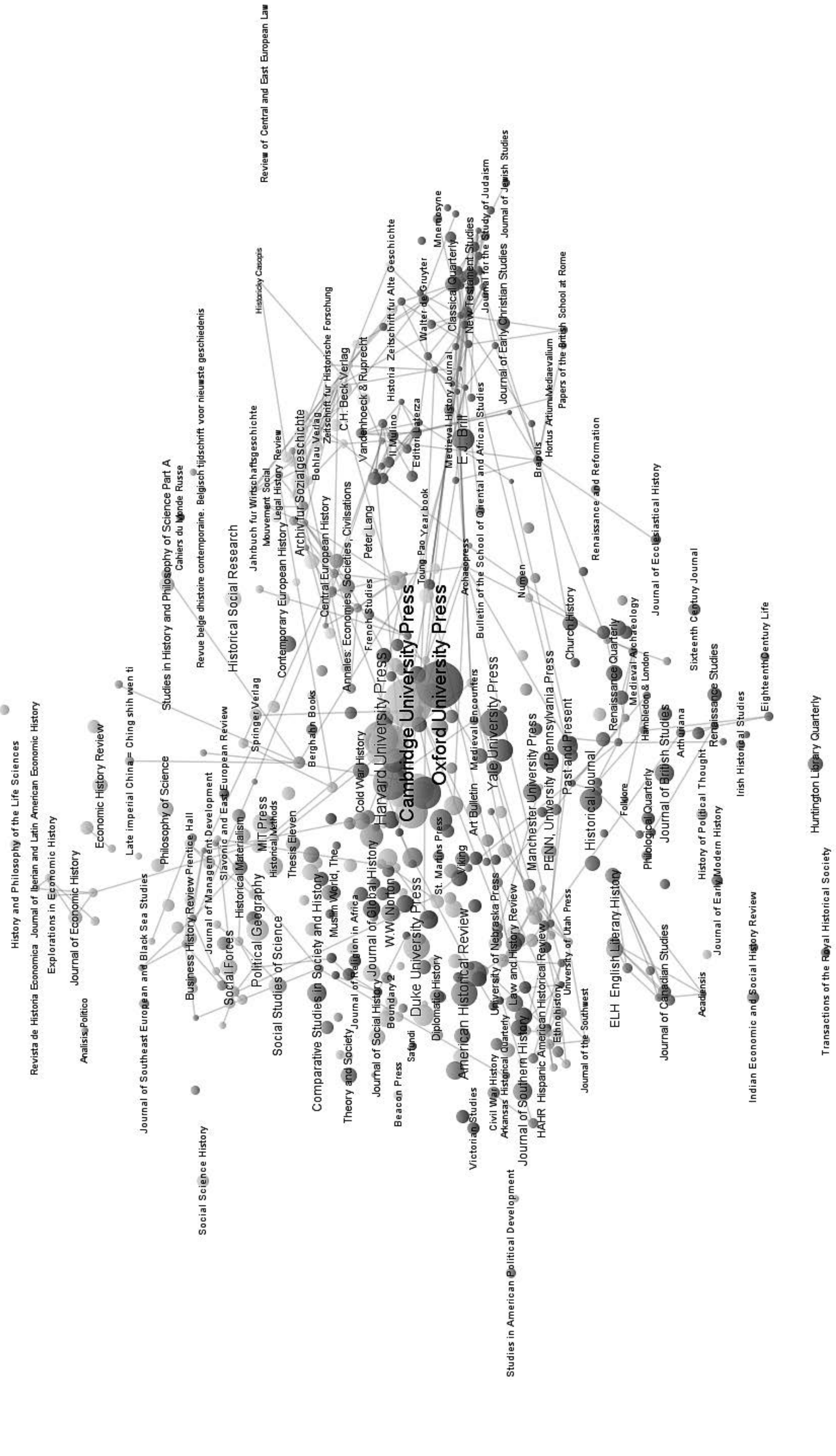
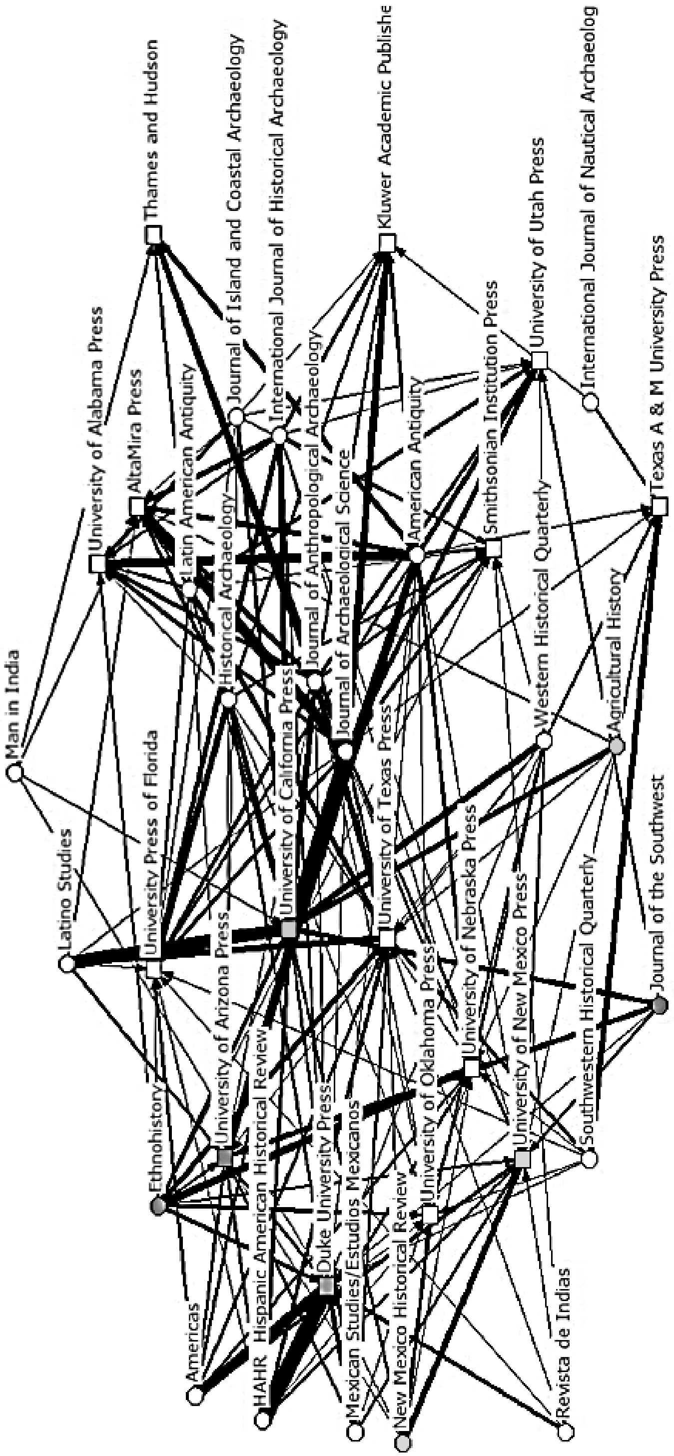


Figure 2 can be interpreted on the basis of sub-themes and/or geographic regions. At the bottom, there is an emphasis on publishers and journals dedicated to period studies (e.g., the renaissance period; sixteenth century; eighteenth century and medieval times). There are two somewhat isolated clusters to the right that emphasize the history of religion (e.g., early Christian studies, the study of Judaism; new testament studies) as well as Italian and Roman studies. The top portion of the map highlights the history and philosophy of science, and towards the left we see journals and publishers that focus on law, politics, diplomatic history, social history and the civil war in the United States. *Cambridge University Press* is clustered in the subfield of economic history, while *Oxford University Press* is aligned predominantly with general British and Irish history, from earlier times as well as the present.

If we 'zoom' closer into the map (Figure 2) and investigate distinct clusters or partitions, we learn more about underlying journal-to-publisher relationships. Figure 3 below indicates that certain presses that publish both journals and books are directly linked. For instance, books published by the *University of New Mexico Press* tend to be cited often in research papers appearing in this parent press' own journal: *New Mexico Historical Review*. Note also that books published by the *University of California Press* are often cited in research papers from the journal *Agricultural History* (i.e., same 'parent' publisher). What this means is that when we rank presses, or use citations to determine press 'prestige', we need to be careful with how we view some of the most economically powerful publishers. Historians do not necessarily have to publish with a high-ranking press, like *Cambridge* or *Oxford* to gain recognition. The choice of publisher often depends on the type of academic audience one would like to reach. Publishing with the likes of *Cambridge University Press* instead of the *University of New Mexico Press* will not necessarily guarantee a greater scholarly advantage, if the desired impact is in a very specific research area.

Figure 3. Specialty Journals and University Presses in Latin American History and Archaeology (n = 35 node cluster from Figure 2).



Continued Research

Further work is needed to compare publishers that were cited during an earlier period (e.g., Scopus *History* journals, 1996 to 2000), to see if this may or may not yield radically different ranking results. We also intend to study books/publishers that were cited in journals from other research fields, like *Literary Theory & Criticism* and *Philosophy*. An additional way of examining the prestige of a publisher might be to compare Scopus journal citations or Thomson Reuter's Book Index (BCKI) citations with WorldCat® international library holding counts. With multiple data-type comparisons we can focus on the perceived cultural benefit of book titles from specific publishers versus their actual use or 'citedness' in the scholarly literature. The normalisation process is indeed yet another step that we would like to take in this methodological foray into ranking. University presses and academic publishing houses are not equal with respect to size, yearly output and sales revenues; hence there may be a natural bias in rankings towards larger, more powerful publishers, rather than smaller regional publishing units. Procedures for selecting books for a book citation index can be informed to some extent by citation-based rankings, but this should not be done at the expense of publishing houses that that do not normally produce texts in English, or focus almost exclusively on manuscripts from a small area of a broader research field.

References

Cronin, B., & La Barre, K. (2004). Mickey Mouse and Milton: Book publishing in the humanities. *Learned Publishing*, 17, 85–98.

de Nooy, W., Mrvar, A., & V. Batagelj, V. E. (2005). Exploratory Social Network Analysis with Pajek, Structural Analysis in the Social Sciences 27, Cambridge University Press, 2005.

Donovan, C., & Butler, L. (2007). Testing novel quantitative indicators of research "quality," esteem and "user engagement:" An economics pilot study. *Research Evaluation*, 16, 231–242.

Garand, J.C. & Giles, M.W. (2011). Ranking scholarly publishers in political science: an alternative approach.

Goodson, L.P. (1999). Ranking the presses: Political scientists' evaluations of publisher quality. *Political Science and Politics*, 32(2), 257–262.

Gump, S.E. (2006). Prestige and the university press. *Journal of Scholarly Publishing*, 37(2), 69–85.

Frankfurt Special. (2010). The worlds biggest publishing groups. *Livres Hebdo*. Retrieved February 25, 2013 from http://www.publishersweekly.com/binary data/ARTICLE_ATTACHMENT/file/000/000/127–1.pdf.

Lewis, J.S. (2000). An assessment of publisher quality by political science librarians. *College and Research Libraries*, 61, 313–323.

Metz, P.,& Stemmer, J. (1996). A reputational study of academic publishers. *College and Research Libraries*, 57, 234–247.

Pasco, A.H. (2002). Basic advice for novice authors. *Journal of Scholarly Publishing*, 33(2), 75–89.

Pratt, D.J. (1993). Why publish with a university press? *Scholarly Publishing*, 24(2), 118–121.

Publishers Weekly. (2012, June 25). *The global 50: The world's largest book publishers*, 2012. Retrieved March 1st, 2013 from http://www.publishersweekly.com/pw/by-topic/industry-news/financial-reporting/article/52677-the-world-s-54-largest-book-publishers-2012.html.

Rowson, R.C. (1995). The scholar and the art of publishing. In A. L. DeNeef and C.D. Goodwin (Eds.), *The Academics Handbook*, 2nd ed. (273–285). Durham, NC: Duke University Press.

SENSE-Research School for Socio-Economic and Natural Sciences of the Environment. (2009). *SENSE ranking of academic publishers*. Retrieved February 25, 2013 from http://www.sense.nl/qualityassessment.

Torres-Salinas, D., & Moed, H.F. (2009). Library catalog analysis as a tool in studies of social sciences and humanities: An exploratory study of published book titles in economics. *Journal of Informetrics*, 3, 9–26.

Van Eck, N.J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538.

White, H., Boell, S.K, Yu, H., Davis, M., Wilson, C.S., & Cole, F.T.H. 2 (2009). Libcitations: A measure for comparative assessment of book publications in the humanities and social sciences. *Journal of the American Society for Information Science and Technology*, 60(6), 1083–1096.

Williams, P., Stevenson, I., Nicholas, D., Watkinson, A., & Rowlands, I. (2009). The role and future of the monograph in arts and humanities research. ASLIB Proceedings: *New Information Perspectives*, 61(1), 67–82.

Zuccala, A. & Guns, R. (2013). Comparing book citations in Humanities journals to library holdings: scholarly use versus perceived cultural benefit. Paper presented at the 14th International Society of Scientometrics and Informetrics Conference, July 15–19, Vienna, Austria.

II. Posters

Measuring “teaming for excellence” in the FP7

Vladimír Albrecht albrecht@tc.cz / Technology Centre of the Academy of Sciences of the Czech Republic, Ve Struhách 27, Prague 6, 160 00 (Czech Republic)

Introduction

This article proposes a simple indicator, which for given country characterizes the collaboration of its teams with the teams of “TOP10 institutions” in the Seventh Framework Programme (FP7). The TOP10 group consists of ten excellent European research institutions, which receive the highest support from the FP7 budget. We call this index country “Collaborative Excellence”(CE). Whereas the CE was estimated in our previous work (Albrecht, Frank, Vavříková 2011) from a short beginning period of the FP7 the current analysis pertains to the period 2007–2011. Recently a group of premier European research institutions launched a white paper “Teaming for excellence”, which is aimed at “building high quality research across Europe through partnership”. Since the FP7 TOP10 belong to the premier European research institutions we consider the CE as a measure of the teaming for excellence process in the FP7.

Selection of FP7 TOP10 institutions

This study is based on data extracted from the European Commission database E-CORDA (release date June 2012), which contains a rich set of characteristics all relevant data on project proposals

Table 1. FP7 TOP10 institutions

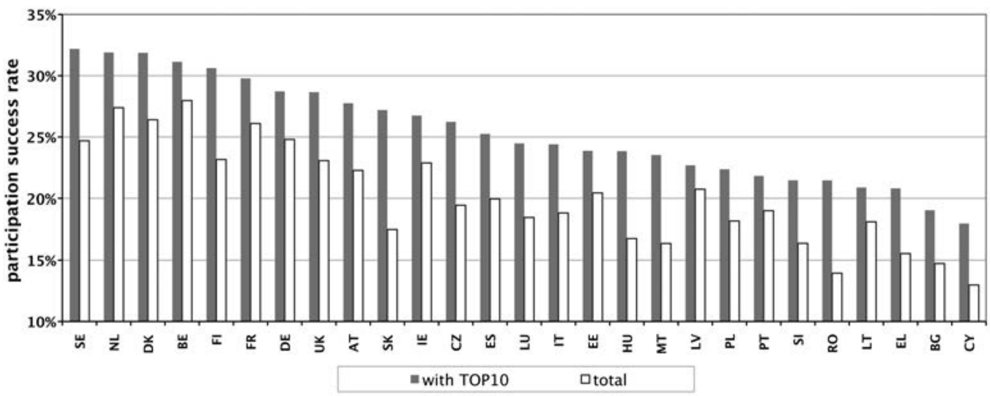
FP7 TOP10 Institutions	Participations
Centre National De La Recherche Scientifique	1020
Fraunhofer-Gesellschaft Zur Förderung Der Angewandten Forschung	708
The University Of Cambridge	486
Max Planck Gesellschaft Zur Förderung Der Wissenschaften E.V.	466
Commissariat a L Energie Atomique Et Aux Energies Alternatives	475
The University Of Oxford	432
Eidgenössische Technische Hochschule Zürich	376
Ecole Polytechnique Federale De Lausanne	350
University College London	334
Imperial College Of Sci., Technology And Medicine	420

and contracted projects. The analysis is restricted to the FP7 specific programme "Cooperation", the projects of which are mainly solved by consortia composed of teams from different countries. All institutions participating in projects of the Cooperation programme were ranged according summary contributions they request from the European Commission for solving the projects. Ten institutions requesting the highest contribution for solving the FP7 projects are listed in tab. 1, together with the number of their participations in the Cooperation programme. Note that the TOP10 group comprises two research universities from Switzerland, i.e. selection of TOP10 institutions is not restricted to the EU member states.

Preparing project proposals with TOP10 increases the participation success rate

Prior estimating the CE for the EU27 member states we want to show that preparing the project proposals in collaboration with the TOP10 teams considerably increases the country participation success rate, see Fig 2. The framed white columns show the total country participation success rate in the FP7 and the full coloured columns indicate the participation success rate of project proposals prepared jointly with TOP10 teams.

Figure 1. Comparison of total participation success rate with participation success rate (framed columns) of proposals prepared in collaboration with TOP10 teams (black columns).



Whereas the total participation success rate ranges from 13% to 28%, the participation success rate in proposals prepared in collaboration with at least one TOP10 team starts at 18% and reaches 32%. The average increase of the participation success rate due to preparing proposals with the TOP10 teams is approximately 5,5%.

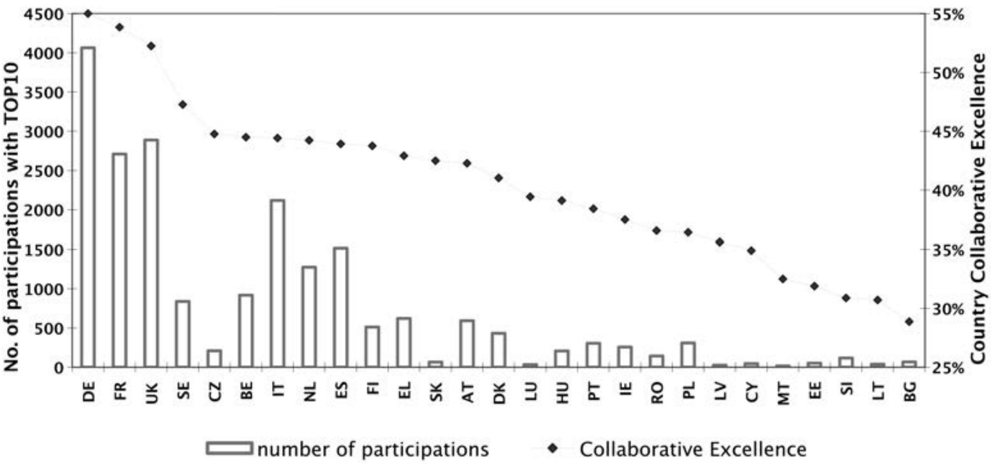
Index of "collaborative excellence"

Let for given country Σ(EC) denotes the sum of the eligible costs that its teams invest to participate in the Cooperation programme projects. Further, let in Σ(ECTOP10) are summed up only the eligible costs invested to participate in consortia comprising some TOP10 team. Then the index of the country Collaborative Excellence is defined by the ratio

CE = Σ(ECTOP10)/Σ(EC).

In Fig. 2 the EU27 member states are ranked according to their CE (the point graph with scale on the right side) and the columns indicate (scale on the left side) indicate the number of country teams collaborating with TOP10 institutions.

Figure 2. The CEs for the EU27 countries (point graph – right side scale) and numbers of country teams collaborating with TOP10 in the Cooperation programme.



The CE ranges from 29% to 55%. The CE is greater than 50% in the three big countries DE, FR, UK, not only because eight of the TOP10 institutions come from them. Namely, note that the column graph indicates that the numbers of DE, FR and UK teams collaborating with TOP10 are considerably higher than the respective numbers of TOP10 teams presented in tab.1. The correlation between the CE and the number of teams collaborating with TOP10 teams is high (corr = 0,82) and could still increase when countries like CZ, SK, HU and PL increase the number of their teams collaborating with TOP10.

Since the Cooperation programme comprises ten different thematic priorities the interpretation of the CE value might be disputable similarly as the interpretation of the total participation

success rate (Rietschel at all 2009).Thus more relevant results can be obtained when the TOP10 institutions are detected separately within every thematic priority and the ten "thematic CEs" are evaluated. The thematic CEs for the whole EU27 and for the Czech Republic are given in Tab.2.

Note that all CEs of the EU27 are always higher than 50%, i.e. the EU27 institutions spend in any thematic priority more than 50% of their participation costs to collaboration with the respective TOP10 group. Note further, that in six priorities the EU27 institutions invest more than 2/3 of their total participation costs to collaborate with the respective partial TOP10 group.

Table 2. The CEs of EU27 and CZ (the Czech Republic) evaluated separately in every thematic priority of the Cooperation programme.

Thematic priority of the Cooperation programme	EU27	CZ
Health	70,4%	70,2%
Food, Agriculture and Fisheries, Biotechnology	82,7%	69,4%
Information & Communication Technologies (ICT)	59,6%	34,3%
Nanosciences, nanotechnologies, materials & new production technologies	69,2%	58,0%
Environment	67,5%	59,4%
Energy	50,8%	51,6%
Transport (including aeronautics)	60,8%	63,9%
Space	73,2%	72,6%
Security	71,0%	73,4%
Socio-economic Sciences and the Humanities (SSH)	58,9%	25,5%

The Czech Republic has a considerable decline from the EU27 CEs particularly in the ICT and SSH. Improving CZ participation in these priorities might be attained when increasing participation of CZ teams in consortia with TOP10 teams.

Conclusions

The FP7 creates an environment supportive to teaming for excellence. The EU member states allocate high portion of their participation costs to collaborate with "thematic TOP10 institutions". Thus in each priority teaming with the respective TOP10 group is decisive for attaining targets of the Cooperation programme.

References

Albrecht, V., Frank, D., Vavříková, L. (2011). Rating collaborative excellence, relevance, significance and competitiveness of the country to assess its participation in the EU Framework Programme, poster, ENID conference, Rome, http://www.enid-europe.org/conference/programme%20CONFERENZA_DEFINITIVO_4%20agosto.pdf.

E-CORDA, MS Access database, (2012). European Commission, Release date: 2012-06-29.

Rietschel, E., Th. et al. (2009): Evaluation of the Sixth Framework Programmes for Research and Technological Development 2002–2006. European Commission, February 2009.

Teaming for excellence: Building high quality research across Europe through partnership, from: http://www.mpg.de/6877962/WP_Teaming_for_excellence.pdf

Citations of social sciences and humanities publications in the Flemish Academic Bibliographic Database

Lucy Amez lucy.amez@vub.ac.be / Vrije Universiteit Brussel, R&D department and ECOOM, Pleinlaan 2, 1050 Brussel (Belgium)

Introduction

The institutional allocation of research funding provided by the special Research Fund of the Flemish Government (BOF) is based on a set of performance driven indicators (Debackere and Glänzel 2004). The majority of the data are collected and monitored by ECOOM, the Centre for R&D Monitoring. Especially bibliometric data are at the core of the performance and monitoring system. Introduced in 2003, the bibliometric component consisted initially of Web of Science (WOS) indexed items: articles, reviews, letters, notes and proceedings papers. Responding to criticism coming mainly from the social sciences and humanities, who challenged the lack of coverage for their respective disciplines, the bibliometric data are since 2008 enriched with publications indexed in the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (VABB-SHW).

The VABB-SHW is a retrospective bibliographic database, assembling publications by authors attached to social sciences and humanities disciplines at a Flemish university. Publication types include, apart from journal/proceedings articles, author/editor of book and book chapters. The BOF regulations stipulate a number of eligibility criteria for the inclusion of an item, conditions which are both formal (publicly accessible, identification by ISBN/ISSN), as well as content related: the publication must have been submitted to a prior peer review process by experts in the discipline and it has to contribute to the development of new insights or applications. In practice, publications obeying the formal criteria are exported from the university's institutional repositories. Based on the incoming references, an Authoritative Panel of 18 professors affiliated with Flemish universities and university colleges, composes a list of journals and a list of publishers which, in their view, obey the content related conditions. The 2012 approved journal list contained about 2000 non-WOS indexed titles. Those lists trigger the final inclusion of an institution's item into the VABB-SHW database.

A number of studies by ECOOM/University of Antwerp, manager/coordinator of the database, profiled the included publications (Engels, Ossenblok & Spruyt 2012). A study of Ossenblok, Engels & Sivertsen (2012) compares the VABB-SHW with the Norwegian database Cristin, which it was inspired by, in terms of Web of Science coverage and language use. WOS coverage of SSH articles can vary substantially depending on discipline, with an average of 30–40%. However, contrary to the situation in Norway, this percentage is increasing for Flemish universities, poten-

tially pointing to incentive effects caused by the absence of valorization of non-WOS publications in the period before 2008.

Apart from the number of WOS publications, also the number of WOS citations and impact factors are gathered by ECOOM and considered in the calculation of the BOF-key. It allows administrations to perform comparative citation analysis' relating performances at institutional or at Flemish level to global averages. However, little is known as to the citation impact of non-WOS publications accepted for the VABB-SHW. The threshold level being academic peer reviewed, the list of authorized journals contains journals with both national and international scope, in English or other languages. Therefore, the esteem of authorized journals might vary, raising questions on how to measure the influence of the items included. This paper studies the amount of WOS citations obtained by non-WOS journal articles accepted for the VABB-SHW database and compares results with statistics on the average number of citations per publication for matching WOS-subfields.

Method

A set of 289 VABB-SHW approved articles was considered, authored by scientists of the Vrije Universiteit Brussel (VUB) and publication year between 2002 and 2008. The citations were collected through a cited reference search in the online Web of Science. This study therefore considers citations of non-WOS, social sciences and humanities publications, by Web of Science included items. The methodology is in accordance with the extended citation analysis to non-sourced data proposed by Butler and Visser (2006). All citations are counted from the year of publication till 2013. The citation searches were performed mainly based on the publication title. This implies there might be citation variants undiscovered, making the presented data minimal estimations. Self citations, on the other hand, are included.

The publications were divided into 17 social sciences and humanities disciplines based on the affiliated team. When co-authored, the publication was fully accredited to both disciplines. In order to compare the citations per publication (CPP) with the average citation of WOS publications, teams were matched to JCR/SS edition subcategories. Sometimes multiple subcategories are attached to one team and different teams are joined together into a discipline. For example, the VUB team Applied Economics is attached to the JCR subfields economics as well as to management and the team Accountancy, Auditing & Business Finance is connected to the JCR subfield Business & Finance. The VABB-SHW publications and citations of both teams are then aggregated into the discipline "Business & Economics". The world benchmark is a weighted average of all WOS citations per publications (exclusive proceedings database) for the JCR subfields that were connected to that discipline, for the same time span 2002–2008, with citation window to 2011. For the sketched example, the WOS-CPP consists of the weighted average of all WOS citations/publications for the JCR subfields Economics, Management, Business & Finance, the weights equaling the importance of the team in the sample. Generally, there is no one to one mapping between VUB teams and JCR subfields.

Results are presented in table 1. Disciplines have been aggregated into two groups, Social Sciences and Humanities.

Table 1. % Cited and CPP for VABB-SHW

Social Sciences**	**304 obs**		
	Min	**Max**	**Avg**
% cited VABB-SHW	16	34	23.6
CPP-VABB	0.47	5.20	1.70
CPP-WOS*	1.5	5.29	3.17
Humanities**	**49 obs**		
	Min	**Max**	**Avg**
% cited VABB-SHW	13	28	20
CPP-VABB	0.31	0.55	0.47
CPP-WOS*	0.48	1.78	0.98

Obs: type article in VABB-SHW (Vrije Universiteit Brussel)/non-WOS pubs
VABB-CPP: VABB-SHW number of citations per publication obtained with the online Web of Science/cited reference search *CPP-WOS: Web of Science number of citations per publication: data sourced by Thomson Reuters Web of Knowledge (formerly referred to as ISI web of science)/licence of ECOOM/There is no one to one concordance with JCR disciplines. ** Only disciplines with >10 obs.

Table 1. shows that for the humanities disciplines about 20% of the publications receives at least one citation. For the social sciences this value is slightly higher equaling 23%. The highest percentage of items cited is observed for Communication Science (34%) in the Social Sciences and for Philosophy and Ethics (28%) in the Humanities. The citation per publication is about half the WOS-CPP for the Humanities and slightly above half for the Social Sciences. However, large differences are observed when comparing disciplines. In the Social Sciences 3/10 considered disciplines perform better than the WOS average, Ergonomics and Sports Sciences (5.2 versus 3.9), Political Science (1.81 versus 1.50) and law (1.84 versus 1.6), the others score substantially lower. In the humanities, Languages & Linguistics as well as History are close to the WOS-average. The performance of Philosophy and Ethics is lower (0.55 versus 1.78) where this discipline scored well in terms of percentage cited.

Conclusion

This paper analyses the citation impact of social sciences and humanities publications accepted for the Flemish Academic Bibliographic Database for Social Sciences and Humanities. As work in progress, data were considered for the Vrije Universiteit of Brussel. The set of observations is small and elaboration is needed to publications of other Flemish universities to obtain a more complete picture. Current results could reflect institutional effects and the comparison with the WOS-CPP

is indicative rather than absolute because there isn't a one to one mapping between the VUB teams and the JCR subcategories.

Keeping in mind those reservations, some preliminary conclusions can be drawn. Although the VABB-SHW list of approved journals is heterogeneous, composed out of journals with both national and international scope, a substantial percentage of the approved articles is referred to in international literature. The CPP is on average lower than the WOS average, but that isn't so for all disciplines. The results therefore seem to complement the study of Ossenblok et al. (2012) in the sense that not only the VABB-SHW completes the WOS publications in terms of coverage, but also in terms of citation impact.

References

Butler, L. & Vissser M.S. (2006). Extending citation analysis to non-source items. *Scientometrics*, 66, 327–343.

Debackere, K. & Glänzel, W. (2004). Using a bibliometric approach to support research policy making: The case of the Flemish BOF-key. *Scientometrics*, 59, 253–276.

Engels, T.C.E, Ossenblok T.L.B. & Spruyt E.H.J. (2012). Changing publication patterns in the Social Sciences and Humanities, 2000–2009. *Scientometrics*, 93, 373–390.

Ossenblok, T. , Engels, T. & Sivertsen, G. (2012). The representation of the social sciences and humanities in the Web of Science: a comparison of publication patterns and incentive structures in Flanders and Norway (2005–2009). *Research evaluation*, 21:4, 280–290.

Co-publication analysis of German and Chinese institutions in the life sciences[1]

Martin Barth m.barth@fz-juelich.de / Project Management Jülich, Bonn (Germany)

Stefanie Haustein stefanie.haustein@umontreal.ca / École de bibliothéconomie et des sciences de l'information, Université de Montréal, Montréal (Canada) and Science-Metrix, 1335 Avenue Mont-Royal Est, Montréal, H2J 1Y6 (Canada)

Barbara Scheidt b.scheidt@fz-juelich.de / Forschungszentrum Jülich, Central Library, 52425 Jülich (Germany)

Introduction

Given China's exceptional growth, the German Federal Minstry of Education and Research (BMBF) is interested in intensifying scientific collaboration with China. To monitor the development of collaborations the BMBF has commissioned several bibliometric studies (e.g. Frietsch and Meng, 2010; Kroll, 2010; Scheidt et al., 2011). This study examines scientific cooperation in the life sciences on the basis of co-publications of German and Chinese institutions as an indicator of formal cooperation between them.

Methods

The dataset for this analysis is based on articles, proceedings papers and reviews published during the 2007 to 2011 period and covered by the Web of Science (WoS) in 97 WoS subject categories relevant to the life sciences. International output was analysed on country level. Co-publication relationships were extracted from the addresses of the authors (Glänzel & Schubert, 2007). Institutional addresses were cleaned and aggregated, collecting all synonyms and address variants for Chinese-German co-publications. To obtain more meaningful information in terms of collaboration structures and identify important sites, institutions that were part of a society or umbrella organization such as Max Planck Society (MPG) or the Chinese Academy of Sciences (CAS) were treated as independent institutions. Network and cluster analysis based on all institutions with at least five co-publications was carried out using VOSviewer software (van Eck & Waltman, 2010).

1 The study was conducted between March and September 2012, when SH was working in the bibliometrics team of the Central Library of Forschungszentrum Jülich.

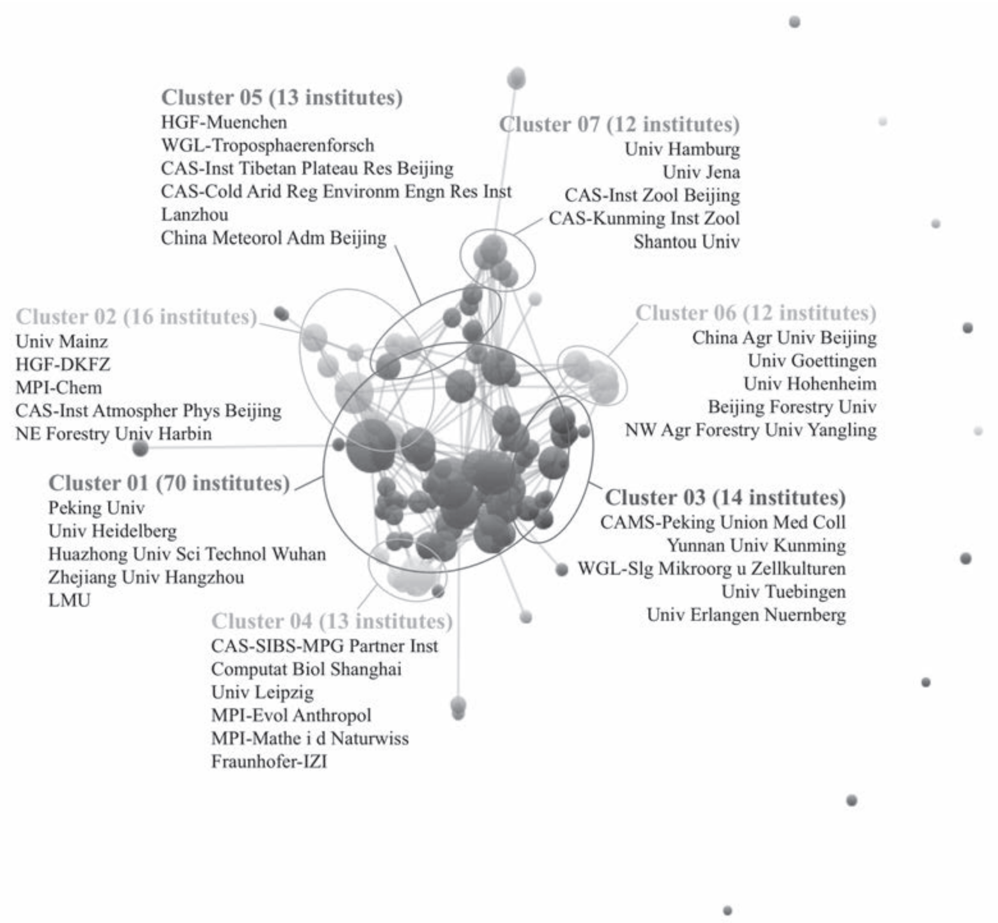
Results and Discussion

During the five-year period 3.35 million life sciences papers were published of which almost one third (32.3%) are produced by US authors. The USA are followed by the UK (8.2%) and Germany (7.4%). China ranks fourth in terms of output during the 2007 to 2011 period, although it has passed Japan, Germany and the UK to become the second largest producer of life sciences publications in 2011. Salton's measure of international collaboration (Salton & McGill, 1986) reveals an exceptional strength of Chinese-US collaboration and shows that it is still strongly increasing above general publication. In 2011, Germany ranked eighth strongest partner of China in the life sciences showing that there is room to extend collaborations.

Based on the 4,009 documents in the life sciences jointly published by authors from Germany and China between 2007 and 2011, a total of 531 German and 700 Chinese co-publishing institutions were identified. *Heidelberg University* (283 co-publications with Chinese institutions) and *Peking University* (300 co-publications with German institutions) are the most frequently publishing institutions from both countries. The strongest German-Chinese cooperation pairs are *Heidelberg University* / *Capital Medical University* in Beijing (82 co-publications), *Heinrich Heine University Düsseldorf* / *Peking University* (44), and the *Leibniz Institute for Tropospheric Research* (WGL-Troposphaerenforsch) in Leipzig / *Peking University* (44).

Figure 1 gives and overview of the co-publication landscape of Chinese and German institutions and shows the seven largest of a total of 26 clusters. With 70 German and Chinese institutions, cluster 1 is by far the largest and most heterogeneous in the network, containing a number of large multidisciplinary universities. The 16 institutions in cluster 2 conduct research on interdisciplinary subject areas. *Mainz University*, one of the key players in this cluster, also receives public funding for interdisciplinary projects in the life sciences.

Figure 1. Co-publication network (VOSviewer) of German and Chinese institutes cooperating in the life sciences. For the seven largest clusters the institutes with the largest number of publications within the network are listed.



While cluster 3 deals with medical issues divided into two thematic subfields, cluster 4 is largely dominated by the *CAS-MPG Partner Institute for Computational Biology*, which is a jointly established institute of MPG and CAS, the two main natural science organizations in Germany and China.

Cluster 5 represents bilateral projects from the geosciences and atmospheric sciences by both university and non-university institutions. Cluster 6 represents agricultural sciences and its related fields – a key element of German-Chinese cooperation, and one that is also strongly supported by public funding. Cluster 7 focuses on a classic core area of the life sciences – zoology.

Conclusions

The present study has shown that China is an important cooperation partner in the life sciences particularly for the USA. Germany is in danger of falling behind in the global competition for the top Chinese cooperation partners. Existing partnerships between German and Chinese institutions in the life sciences have been identified and visualized to inform funding agencies and researchers about the structural conditions for German-Chinese cooperation in the field of life sciences.

References

Frietsch, R. & Meng, Y. (2010). Indicator-based reporting on the Chinese innovation system 2010 – Life sciences in China. *Fraunhofer ISI Discussion Papers. Innovation Systems and Policy Analysis*, 26.

Glänzel, W. & Schubert, A. (2007). Analyzing scientific networks through co-authorship. In H. F. Moed, W. Glänzel & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems*, (257–276). Dordrecht: Kluwer Academic Publishers.

Kroll, H. (2010). Indicator-based reporting on the Chinese innovation system 2010 – The regional dimension of science and innovation in China. *Fraunhofer ISI Discussion Papers. Innovation Systems and Policy Analysis*, 25.

Salton, G. & McGill, M. J. (1986). *Introduction to Modern Information Retrieval.* New York: McGraw-Hill Inc.

Scheidt, B., Haustein, S., Holzke, C, & Tunger, D. (2011). Bibliometric Analysis Asia-Pacific Research Area 2010. Comissioned by the International Bureau of the BMBF. Jülich: Forschungszentrum Jülich. http://www.inter nationales-buero.de/_media/Bibliometric_Analy sis_APRA_2010.pdf. Accessed 25 February 2013.

Van Eck, N.J. & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523–528.

Mobility of researchers in the European Union[1]

Thomas Blondiau — Thomas.blondiau@ideaconsult.be / IDEA Consult, Kunstlaan 1–2 box 16, Brussels, 1210 (Belgium)

Miriam Van Hoed — Miriam.Vanhoed@ideaconsult.be / IDEA Consult, Kunstlaan 1–2 box 16, Brussels, 1210 (Belgium)

Arnold Verbeek — Arnold.Verbeek@ideaconsult.be / IDEA Consult, Kunstlaan 1–2 box 16, Brussels, 1210 (Belgium)

Annelies Wastyn — Annelies.Wastyn@ideaconsult.be / IDEA Consult, Kunstlaan 1–2 box 16, Brussels, 1210 (Belgium)

Introduction

The EU2020 Strategy builds on Europe's strongest (knowledge) asset, human capital and more precisely Europe's researcher population. Creation of a European Research Area (ERA) and the Innovation Union initiative in particular, is one of the cornerstones of the EU 2020 strategy (EC, 2010) and future economic competitiveness of Europe. In 2007 the ERA concept was put high on the European policy agenda through the publication of the 'ERA Green Paper' (EC, 2007) and the launch of various related policy initiatives.

Mobility of researchers is in this context considered an important factor to increase cooperation, experience, skills and research excellence. Several instruments to stimulate mobility are put in place at regional, national and transnational level. At the EU level, the European Commission has taken the lead by introducing new, and adapting existing R&D support schemes, such as: the Framework Programmes and the Marie Curie Actions, the adoption and implementation of the European Charter for Researchers, and the Code of Conduct for the Recruitment of Researchers, the 'scientific visa' package, and the integrated European Researcher Partnership.

Literature Review & Research questions

Research has shown however that not all researchers are equally inclined to become mobile. Both professional characteristics like career stage, discipline and working conditions in current employment, and personal characteristics like gender, age, family status and country of origin, tend to affect one's decision to become mobile.

1 This work was supported by the European Commission – Directorate General for Research & Innovation through the MORE2-project: http://www.more-2.eu

According to Carr et al. (2005), decisions to go abroad are often motivated by opportunities for improvement of the lives of family members, particularly educational opportunities for their children. Factors to decide to stay or to return to their home country were related to family concerns such as having parents at home and attitudes of spouses or partners.

Most research on gender differences related to mobility still point in the direction of female researchers being less inclined to long term mobility. For example, Avveduto et al. (2004) noted that women are underrepresented in international mobility.

In addition to these personal characteristics, professional factors play a role as well. Cañibano et al. (2011) suggested that differences in culture of mobility occur across fields of science. Mahroum (2000) also found large inter-disciplinary differences as to what motivates people to move overseas.

Even though there is substantial academic interest, the exact relation between the researcher characteristics and the different forms of mobility and collaboration have not before been estimated and described. In order to support policy and help orienting further research, we use the extensive and statistically representative (at country level) database of the MORE2 to shed light on this issue. We focus on the following research questions: How does a researcher's propensity to be long-term & short-term mobile depend on gender, career stage, age, field of science and family situation? And how does a researcher's motivation to be long-term mobile differ with respect to gender and career stage?

Methodology & some descriptive statistics

We use the dataset of the MORE2-project (http://www.more-2.eu) for answering our research questions. This dataset consists of a large survey among the population of EU researchers. These are researchers who are active at a higher education institute in the EU27 countries, in one of the associated countries (Iceland, Norway, and Switzerland) or in one of the candidate countries (Croatia, FYROM/Macedonia, and Turkey). The population is not restricted to EU nationals and can thus consist of non-EU nationals as well.

The methodology we used for collecting the sample of researchers is stratified random sampling with first stage clustering of Higher Education Institutions. The individual researchers are stratified by field of science. The rationale behind this is that the field of science closely affects most variables in the analysis, such as mobility profile, collaboration profile, etc. Further, the sample size was chosen to obtain a sample which is representative at country level and at field of science. This means that the maximum sampling error is below 5% for all the countries and all the field of sciences in our survey. The same level of representativeness could not be ensured at the level of field of science per country. Maximum sampling error at this level was typically around 7%. The total sample size is equal to 10,547 units.

For explaining the mobility decisions, we estimate two logit models with a dummy variable as dependent variable that indicates whether the researcher has (recently) been long-term and short-term mobile. Among the explanatory variables we include researcher characteristics related to gender, field of science (agricultural sciences, engineering, humanities, medical sciences, natural sciences, social sciences), age, career stage (post PhD, established researcher, leading researcher), family status (in couple or single, with or without children) and collaboration indicators. Among the collaboration indicators we include whether the researcher collaborates with other universities or with non-academic sector. We also distinguish between domestic, EU and outside-EU collaboration. For analyzing how the motivation to be mobile differs with researchers' characteristics, we estimate a set of logistic models with gender (logit model) and career stage (ordered logit model) as dependent variables[i]. The dependent variables in this regression are dummy variables indicating the main motivation for each mobility move (self-reported by researchers in the survey).

Results

We display the regression results of the logit model on long-term mobility and short-term mobility in Table 1. We find that males are significantly more mobile than females in EU both for long-term and short-term mobility, controlling for other effects such as age, career stage and field of science. Moreover, we also find that both types of mobility decreases with age, but increases with career stage. With respect to field of science, we do not find wide differences in mobility profiles. For long-term mobility, the natural sciences seem to have a tendency to be more mobile, but the effect is not significant at 5% level. For short-term mobility, agricultural sciences seem to contain significantly more mobile researchers than the other fields. With respect to family profile, having children significantly decreases long-term mobility, whereas being in couple does not lead to any significant effect. There does seem to be some tendency to substitute long-term for short-term mobility experiences though. Finally, with respect to collaboration profile, we find that most coefficients are in line with what we anticipated, except for the observation that cooperation with non-academic sector within the EU does not lead to increased mobility but, in contrast to significantly lower long term mobility.

*Table 1. Regression results for long-term and short-term mobility profiles (t-values between brackets, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)*

	(1) LTmobility	(2) STmobility
Base = Female		
Male	0.475*** (5.25)	0.176** (2.26)
Base = Agricultural sciences		
Engineering	-0.0177 (-0.07)	-0.425** (-2.31)
Humanities	0.0545 (0.21)	-0.303 (-1.53)
Medical Sciences	-0.0696 (-0.29)	-0.768*** (-4.25)
Natural Sciences	0.293 (1.24)	-0.410** (-2.31)
Social Sciences	0.0438 (0.19)	-0.352** (-2.02)
Base = < 35		
Agegroup 35–45	-0.433*** (-3.14)	-0.227* (-1.77)
Agegroup 45–55	-1.375*** (-8.80)	-0.403*** (-2.87)
Agegroup 55–65	-1.575*** (-8.99)	-0.318** (-2.07)
Agegroup 65+	-1.709*** (-6.09)	-0.414* (-1.91)
Base = careerstage Post-PhD		
Careerstage established researcher	0.403*** (3.55)	0.323*** (3.27)
Careerstage leading researcher	0.813*** (6.25)	0.531*** (4.72)
Base = single		
In couple	-0.111 (-0.62)	0.253 (1.50)
Base = Without children		
With children	-0.352*** (-2.96)	-0.190* (-1.79)
Collaboration with		
Universities in your country	-0.501*** (-3.87)	-0.0628 (-0.52)
Non-academic, in your country	-0.451*** (-4.80)	-0.302*** (-3.77)

… Continuation Table 1

	(1) LTmobility	(2) STmobility
Universities, other EU countries	0.403*** (3.72)	0.349*** (3.81)
Universities, non-EU countries	0.603*** (6.32)	0.428*** (5.26)
Private, other EU countries	-0.327** (-2.48)	0.00566 (0.05)
Private, non-EU countries	0.324** (2.05)	0.131 (0.93)
Constant	-0.408 (-1.33)	-0.475* (-1.85)
Observations	8210	8210

In a second step, we analyse the motivations for mobility and compare these along gender and career stage dimensions. We find that males are mostly motivated by 'working with leading experts', 'cultures and/or language', 'remuneration' and 'working conditions in general, whereas women are more often motivated by 'career progression', 'facilities and equipment', 'quality of training and education' and 'personal/family reasons'. Along the career stage dimension, we find that 'research autonomy', 'remuneration' and 'working conditions' become more important at higher career stages. In contrast, 'finding a suitable position', 'quality of training and education', 'personal/family reasons' and 'career progression' are more important at lower career stages.

References

Ackers, L. (2010). Internationalisation and equality: The contribution of short stay mobility to progression in science careers. *Recherches sociologiques et anthropologiques*, 1, 83–103.

Ackers, L. (2008). Internationalisation, mobility and metrics: A new form of indirect discrimination? *Minerva*, 46, 411–435.

Avveduto, S. (2004). *International Mobility of PhDs. National Research Council.* Rome: Institute for Studies on Scientific Research and Documentation.

Cañibano, C., Otamendi J. F. & Solís, F. (2011). International temporary mobility of researchers: cross-discipline study. *Scientometrics*, 89, 653–675.

Carr, S. C., K. Inkson, et al. (2005). From global careers to talent flow: Reinterpreting 'brain drain'. *Journal of World Business*, 40(4), 386–398.

European Commission (2010). *Europe 2020 Flagship Initiative – Innovation Union*, SEC(2010) 1161 final, Brussels, 6 October 2010.

European Commission (2007), *The European Research Area: New Perspectives – Green Paper*, 04.04.2007, Luxembourg.

Mahroum, S. (2000). Highly skilled globetrotters: Mapping the international migration of human capital. *R&D Management*, 30(1), 23–31.

Inzelt, A. (2011). Analysis of Researchers' Mobility in the Context of the European Research Area, Retrieved June 16, 2012 from: http://ec.europa.eu/research/evaluations/pdf/archive/fp7-evidence-base/experts_analysis/a.%20inzelt_-_researchers%27_mobility.pdf

i Methodologically, this boils down to a condense representation of a comparative correlation analysis between gender and mobility motivation, and career stage and mobility motivation. The robustness of the results that we present here has been checked in a more comprehensive logistic regression model with main motivation as dependent variable and simultaneous inclusion of gender, career stage, age, field of science and family situation as dependent variables to allow for ceteris paribus interpretation.

Technology decisions: The relevance of indicators and the influence of social relations

Nuno Boavida nuno.boavida@partner.kit.edu / Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology, Karlstr 11, 76131 (Germany) and CESNOVA, Universidade Nova de Lisboa, Av. de Berna 26 C, Lisboa, 1069-061 (Portugal)

António Moniz antonio.moniz@kit.edu / Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology, Karlstr 11, 76131 (Germany)

Manuel Laranja mlaranja@iseg.utl.pt / Instituto Superior de Ecnomia e Gestão, Universidade Técnica de Lisboa, Rua do Quelhas, n.º 6, Lisboa, 1200-781 (Portugal)

Introduction

The aim of this paper is to shed light on the actual importance of indicators and social influences in knowledge-intensive groups when making technology decisions. The work assumes that a focus on those who have access to knowledge under the least possible restrictions will reveal that they use indicators:

H1) to make technology decisions
H2) according to the type of technology decision
H3) according to the level of complexity of the decision
H4) more intensively to decide than they are influenced by social relations.

The focus was on three knowledge-intensive innovation groups:

(a) Researchers, academics and health R&D personnel
(b) Business R&D and Innovation related personnel
(c) Policy makers

This work deals with "indicators" as quantified objective tools available to help people decide about technology, such as costs, technical characteristics, market share, R&D expenditures, etc. The term "technology decision" is used in broad terms to define a decision related to technology, such as to:

(1) acquire an equipment or a specific technology,
(2) develop a product or a specific technology,
(3) buy property rights,
(4) design technology policies (programs, measures, actions, projects, etc.).

1 This work was supported by the Portuguese Fundação para a Ciência e Tecnologia

The term “social relations” is used broadly to define all social activities that influenced a technology decision (e.g. networking activities, hierarchical or peer pressures, marketing activities, values and norms, etc).

Theoretical background

The influence of social activities over a technology decision is dependent on where actors are socially situated and integrated (Perri 6 2002). Perri 6 claimed that decision makers use only a certain amount of knowledge when making a decision, and their judgments are rather dependent on where actors are socially situated and integrated. The author supports the view that forces of social regulation and social integration exist and shape individuals’ decisions, and can be used to explain how several social actors use information, behave and judge. Therefore, social relations acts upon individuals, both consciously and unconsciously, constraining and guiding them throughout their decision making.

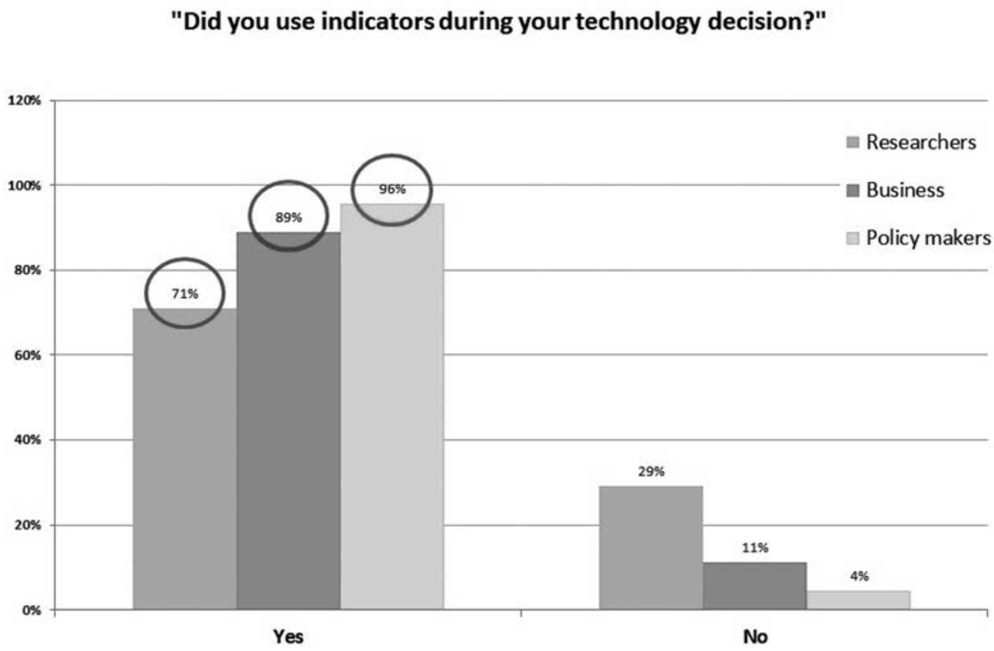
However, the influence of these social activities can vary according to the type of social group involved. Individuals should be more bounded to use indicators to make a decision in groups where knowledge exists in abundance and is part of their work activities. To Anthony Giddens (1994) society has “expert systems”, which are expected to act as instruments of security and social thinking. To Feller-Länzlinger et al. (2010) “indicator systems” can guide society and experts in vast and complex issues to monitor their actions – with increasing self-reflection and self-discipline – striving to change and improve themselves, both at the state and individual level.

Methods

The initial work combined literature research, analysis of official documents and exploratory interviews. The exploratory interviews were designed to prepare and calibrate the questionnaires and subsequent interviews. Eleven privileged informants were interviewed in Germany, Switzerland and Portugal, between October 2011 and January 2012. Afterwards, the work included the design and deployment of three closed online questionnaires, complemented with seventeen semi-structured interviews. The questionnaires addressed Portuguese representatives of the three knowledge-intensive communities under analysis, collecting fifty one answers from February 2012 to June 2013. Response rates varied from 63% in the business group, 40% in the researchers community and 39% for policy makers.

Findings

H1) Innovation groups **use indicators** to make technology decisions



H2) Innovation groups tend to use indicators to make the **same type of decisions**

	Researchers	Business R&D&I	Policy makers
Buying equipment/technology	+++	+	+
Development of products/technology	+	+++	+
Purchase of intellectual property rights	+	+	-
Policy design	-	-	+++

Note: Intensity of use of indicators was classified as High (+++), Moderate (++), Low (+) and No answers (-)

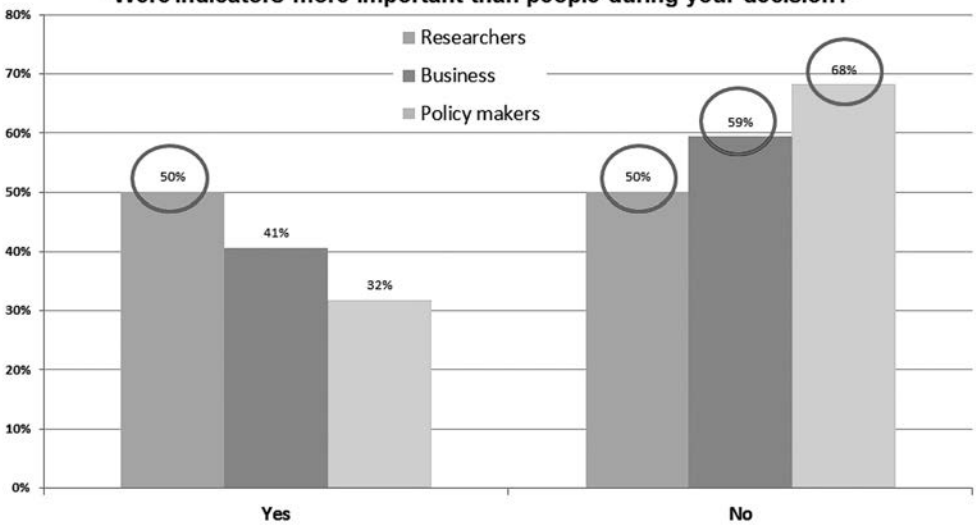
H3) Innovation groups **use indicators in different ways** to decide

	Researchers	Business R&D&I	Policy makers
Use given to indicators *before* the decision	+++	+++	++
Use given to indicators *after* the decision	+	+	++

Note: Intensity of use of indicators was classified as High (+++), Moderate (++) and Low (+)

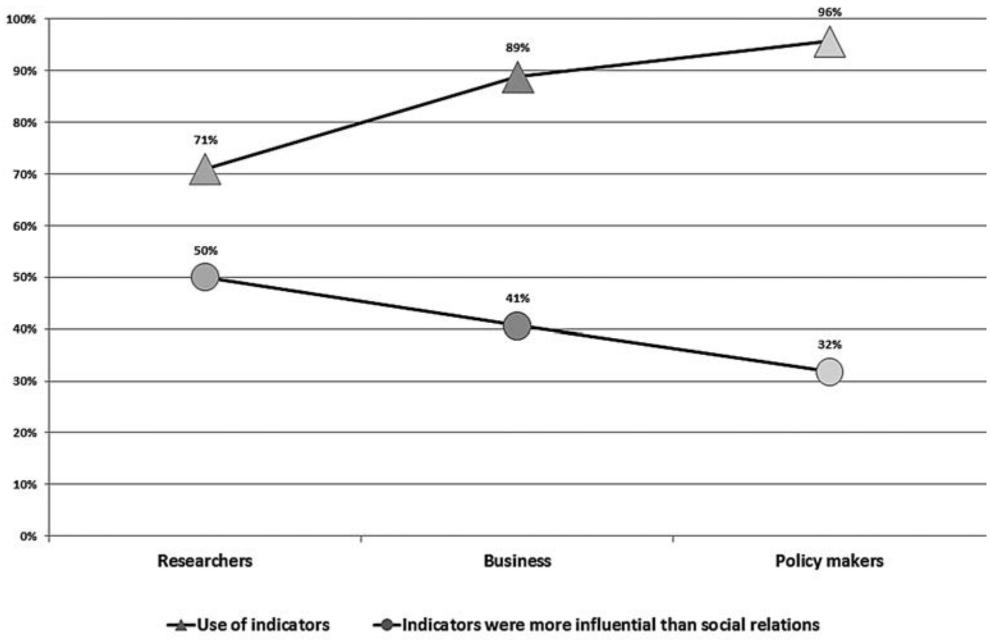
H4) Innovation groups **do not** use indicators more intensively to decide than they are influenced by social relations: Researchers use indicators **as intensively**; policy makers and business groups are **more influenced by social relations**

"Were indicators more important than people during your decision?"



Discussion

Ranking of most influential groups	Researchers	Business R&D&I	Policy makers	Total
Liable politicians / Management / Managers		1	1	1
Technology users	1		2	2
Researchers / Academics	2		3	3
Experts		2		3
Colleagues	3	3		3



Conclusions

The findings confirmed the hypotheses that knowledge-intensive innovation groups use indicators *H1)* to make technology decisions; *H2)* according to the type of technology decision and *H3)* according to the level of complexity of the decision. However, the last hypothesis *H4)* was not confirmed. In fact, researchers use indicators as intensively as they are influenced by social relations (50%); and policy makers and business groups are more influenced by social relations than they are by indicators in their technology related decisions.

Furthermore, researchers revealed stronger influences from technology users and their peers. Business related R&D personnel revealed that managers, experts and colleagues were the most influential groups in their technology decisions. Last, policy makers indicated that liable politicians, technology users and researchers were the most influential groups in shaping their decisions.

In addition, findings also pointed to the idea that among knowledge-intensive groups those working closer to social activities tend to be less influenced by indicators. Those were work is less socially prone tend to be more influence by indicators.

In conclusion, indicators are one among the tools available to decide. Their influence varies according to the social ties each knowledge-intensive group has in their activities.

References

Feller-Länzlinger, R., Haefeli, U., Rieder, S., Biebricher, M., & Weber, K. (2010). *All sized up - Counting, calculating and controlling in the knowledge-based society. TA-SWISS* (26–40). Bern: TA-SWISS. Retrieved from http://www.ta-swiss.ch/en/projects/social-and-cultural-ta/indicators/

Giddens, A. (1994). Living in a Post-Traditional Scoiety. In U. Beck, A. Giddens, & S. Lash (Eds.), *Reflexive modernization: politics, tradition and aesthetics in the modern social order* (56–109). Polity Press.

Perri 6. (2002). Can Policy Making be Evidence-Based ? *MCC: Building knowledge for integrated care*, 10(1), 3–8.

The characteristics of highly cited non-source items in political science

Pei-Shan Chi chi@forschungsinfo.de / Institute for Research Information and Quality Assurance, Schützenstraße 6a, Berlin, 10117 (Germany) / Berlin School of Library and Information Science, Humboldt-Universität zu Berlin, Unter den Linden 6, Berlin, 10099 (Germany)

Background

The citation distribution is extremely right-skewed and affected strongly by outliers (Seglen, 1992; Bornmann et al., 2008; Redner, 2004; Wallace, Larivière & Gingras, 2009). A few extremely highly cited papers contribute to the most citations. Because of their significant effect on statistical properties of the whole distribution, the highly cited papers catch the attentions of researchers to explore their characteristics (Aksnes, 2003a; Aversa, 1985; Oppenheim & Renn, 1978) and are often used as evaluation indicators in bibliometric studies (Tijssen el al., 2002; Plomp, 1994). The highly cited papers in natural sciences shows several characteristics which were discovered, such as, they are authored by many researchers, they are resulted largely by international collaboration, they are mainly published in high-impact journals, they are mainly cited by foreign scientists, the share of self-citations is low, they age less rapidly than other articles (Aksnes, 2003a; Glänzel, Rinia & Brocken, 1995; Aversa, 1985).

However, most studies exploring highly cited papers are applied in natural sciences, and focus only on papers which are indexed by WOS. Those non-source items, which are not indexed in WOS, are missing in these inquiries, especially in the fields of social sciences and humanities. Therefore, this study will probe the characteristics of highly cited items in the non-source items of political science in the light of different document types, to answer the following questions.

- Are highly cited non-source items authored more researchers than other items?
- Do highly cited non-source items involve more international collaboration?
- Are highly cited non-source items more cited by foreign researchers?
- Do highly cited non-source items have lower share of self-citations than other items?
- Do highly cited non-source items age more slowly than other items?
- What are differences of highly cited items among different document types?

Methods

This study concentrates on the five year research outputs in political science. Four-year fixed citation window (Glänzel, 1997; 2008) is applied in this study. The 1,015 publications of 33 professors of two top-ranking German institutions, Department of Political Science at Mannheim

University and Institute of Political Science at University of Muenster (CHE, 2010; Hix, 2004) from 2003 to 2007 were collected from researchers' official websites, institutional repositories, and German Social Science Literature Information System (SOLIS). After data collection, all publications were sent to the professors for verification. References and citations of these items were obtained in March till December 2012 from the WOS in-house database of the Competence Centre for Bibliometrics for the German Science System (Kompetenzzentrum Bibliometrie).

The citations of all publications are matched according to different rules set for different document types. For example, each *Journal Article* was searched the combination of first author's lastname, pubyear, volume and first page in three rounds (I: pubyear & volume & firstpage, II: pubyear & volume & lastname & sourcetitle, III: lastname & sourcetitle) in the references of articles in our WOS in-house database. Besides exactly matching the lastname and sourcetitle, a reference should also match at least two items among pubyear, volume, and firstpage, to be identified as a citation of samples. For *Book Chapter*, we search for the first word of title and the first author's lastname in the WOS references in the first round, then check the results and include those items published in ±1 year and try to get the different abbreviations of journal titles and author names from primary salvaged items. The second round is to search the references by the first page and the improved first author's lastnames. In the end, the combinations from these two rounds will be duplicated the duplicates. While checking the data, the reference with exact firstpage data is allowed to extent to ±1 year in pubyear data, but the one without firstpage data should match the exactly pubyear. The firstpage data is allowed to extent to ±2 when other conditions are matched.

According to Borgman and Furner (2002, p. 16), *author self-citation* occurs when at least one of the authors of a cited document is the same person as one of the authors of the citing document. This study takes the synchronously way to identify the above definition of self-citations (Aksnes, 2003b). We check for all items to see if any one of the authors' name appears in the references of the same item. The cited references with the same author name should be matched to the author's publication list to make sure they are the same author. For highly cited items, this study takes the relative way to set the threshold of the definition of 'highly cited', the top 5% items of each document type.

Table 1. shows the primary data we already gained after searched and matched in our database. The overall picture is: the WOS coverage is around 7% since 70 of 1,015 publications are indexed in WOS; highly cited items gain around 6 citations in average; generally highly cited items are authored slightly more researchers than other non-highly cited items, except for *Journal Articles* and *Books*; highly cited items are not cited by more foreign researchers, but have lower self-citation rate than other items. Besides, the general self-citation rate is about 20%; in most of document types, these 5% highly cited items contribute nearly 50% (or more) citations, except for the 16% of *ISI Journal Article*.

Table 1. Statistic data of publications of two German political science institutions.

	No. of publ.		Ave. cit. Rate		Ave. no. Author		Ave. no. countries in cit.		% of self-cit.	
	All	**HCI**	**All**	**HCI**	**All**	**HCI**	**All**	**HCI**	**All**	**HCI**
ISI JA	70	3	4.3	16.0	1.8	1.3	1.2	1.3	24.0	18.8
Non-ISI JA	151	7	0.6	7.3	1.8	1.7	1.1	1.2	12.4	7.8
Book	45	2	1.2	12.5	1.9	1.0	1.1	1.2	19.6	4.0
BK(E)	76	3	1.8	23.7	2.6	3.3	1.3	1.2	18.8	26.8
BK Chap.	396	24	0.3	3.3	1.5	1.7	1.1	1.1	19.6	21.5
Conf. P	151	5	0.1	2.0	1.7	1.8	1.5	1.2	15.0	10.0
Others	126	6	0.2	3.0	1.7	2.2	0.9	0.7	33.3	27.8
Total	1015	50	0.7	6.0	1.7	1.8	1.2	1.2	20.7	18.5

References

Aksnes, D. W. (2003a). Characteristics of highly cited papers. *Research Evaluation*, 12(3), 159–170.

Aksnes, D. W. (2003b). A micro study of self-citation. *Scientometrics*, 56(2), 235–246.

Aversa, E. S. (1985). Citation patterns of highly cited papers and their relationship to literature aging: A study of the working literature. *Scientometrics*, 7(3–6), 383–389.

Borgman, C. L., Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, 36, 3–72.

Bornmann, L., Mutz, R., Neuhaus, C., Daniel, H.-D. (2008). Citation counts for research evaluation: Standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics*, 8, 93–102.

CHE University Ranking 2010/11-Political science (2010). Retrieved 10. 08. 2010, 2010, from http://ranking.zeit.de/che2010/en/rankingkompakt?esb=28&hstyp=1

Glänzel, W. (1997). On the possibility and reliability of predictions based on stochastic citation processes. *Scientometrics*, 40(3), 481–492.

Glänzel, W. (2008). *Seven Myths in Bibliometrics. About facts and fiction in quantitative science studies*. Paper presented at the Proceedings of Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting, Berlin.

Hix, S. (2004). A global ranking of political science departments. *Political Studies Review*, 2(3), 293–313.

Oppenheim, C., Renn, S. P. (1978). Highly cited old papers and the reasons why they continue to be cited. *Journal of the American Society for Information Science*, 29(5), 225–231.

Plomp, R. (1994). The highly cited papers of professors as an indicator of a research group's scientific performance. *Scientometrics*, 29(3), 377–393.

Redner, S. (2004). Citation Statistics From More Than a Century of Physical Review. *Physical Review*, physics/04, 17.

Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, 43(9), 628–638.

Tijssen R. J. W., Visser M. S., van Leeuwen T. N. (2002). Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics*, 54(3), 381–397.

Brain Drain or Brain Circulation? The Extent of German Scientists to Work Abroad

Sonia Conchi Sonia.Conchi@isi.fraunhofer.de / Fraunhofer Institute for Systems and Innovation Research ISI, Breslauer Strasse 48, 76139 Karlsruhe (Germany)

Carolin Michels Carolin.Michels@isi.fraunhofer.de / Fraunhofer Institute for Systems and Innovation Research ISI, Breslauer Strasse 48, 76139 Karlsruhe (Germany)

Introduction

A global international exchange between scientists is essential to drive progress. Thus, scientists form international networks to promote the exchange of information in their research field.

The term "Brain drain" was used to refer to scientists who migrate from one country to another with no intention of returning (Grubel, 1994) – an action that has international, economic and political impacts, especially in developing countries (see, e.g. Lowell 2002). According to Sjastaad (1962) migrations are seen as individual investments in human capital which can be divided into costs and benefits. It follows that those countries with a higher income and a lower cost of moving are more attractive for emigrants (Ette, 2010). Also, the more years are left on the labor market, the more likely a person is to move to another country (Ette, 2010). Emigrants build relationships across national boundaries and create a permanent connection between home and foreign country that enables further exchange.

Bibliometric data can be used to track scientists over time. Roberge and Campbell (2012) conducted an analysis of Canadian researcher migration on Scopus data, which reveals a net migration flow on a very low level. Recently, Moed et al. (2013) also showed that it is possible to trace scientists and their mobility using Scopus data. According to their study, countries to which German scientists migrated most frequently were the USA, UK, Switzerland, France, Austria and the Netherlands.

The aim of this study is to assess scientific migration in Germany. A data set of a cohort of German scientists and their movements for a period of 10 years was created based on Scopus. This paper provides a descriptive analysis of the observed mobility. In particular, the following hypotheses were tested:

H1: The share of scientists publishing abroad is increasing over time.
H2: Most scientists return after two years abroad.
H3: The most popular countries are English speaking countries.

Methodology

The method of data compilation is comparable to that of Laudel (2003). Similar to Moed et al. we created a diachroneous approach, in which a group of authors was followed over 10 years. The basic data set contains all the scientists who published in 2000 and also in the previous two years 1999 and/or 1998 with a German affiliation. We use the condition of publications in the previous years to minimize the proportion of immigrants staying temporarily in Germany. As a further restriction at least one publication in 2010 was required. To reduce the merge author error (cf. Moed et al., 2013), the number of publications was limited: First, all authors with less than 5 publications were ruled out, then the upper quartile was excluded. In that way, the maximum publication number was 80 publications in 10 years. The publication activity of these 15,798 German researchers was traced for 10 years using the Scopus' author ID.

Results

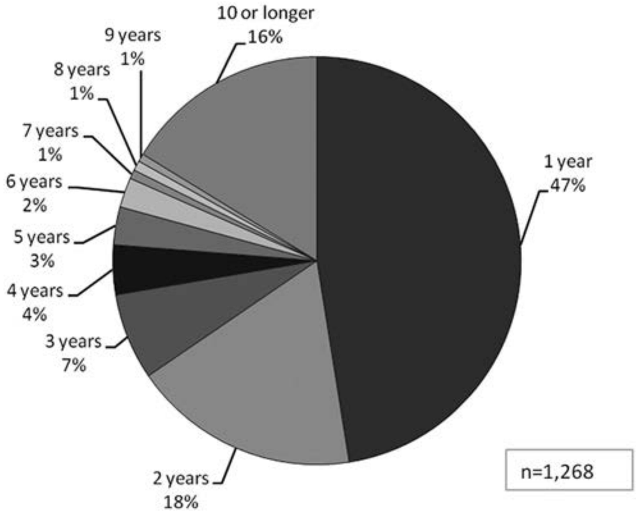
H1: The share of scientists publishing abroad is increasing over time.

Apart from scientists with only a German or foreign address, there were some with at least one German and one foreign affiliation in one single year. The number of scientists publishing with a foreign address increased during the time period up to 13% in 2010. In 2004, it even exceeds the number of scientists with a German and a foreign address – so called co-affiliations. It can be concluded that the probability of giving up the German address is increasing with time.

H2: Most scientists returned after two years abroad.

Those scientists who went abroad were categorized by the maximum number of years they published consecutively with foreign addresses in 10 years. All the researchers who were still living abroad at the end of the observation period were excluded except for those that had been abroad since 2001 and thus were unlikely to return. Figure 1 shows the distribution of the period spent abroad for this set of researchers.

Figure 1. Duration of stay abroad in years

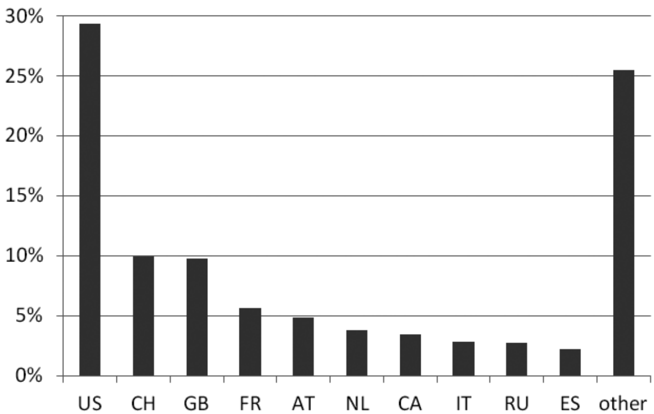


A relatively high proportion of those who moved in the observation period returned after one (47%) or two years (18%). 16% of them did not return to Germany within 10 years and are thus unlikely to do so in the future. The hypothesis can be confirmed, as most of the researchers have a longest stay abroad of less than three years.

H3: The most popular countries of migration are English speaking countries.

Figure 2 shows the top 10 most popular migration countries for German scientists in the period between 2000 and 2010.

Figure 2. Top 10 migration countries.



The US are the most popular target with 29% of movements. The next highest ranked migration countries are the UK and Switzerland (10% each). France, Austria, the Netherlands, Canada, Italy, Russia and Spain are on approximately the same level (between 2% and 6% of movements) with the ranks 4 to 10. Apparently, the language as well as the proximity plays a role in the migration country selection.

Discussion and Conclusions

On average 18% of German scientists take part on international mobility according to the Scopus data. To detect whether a brain drain is taking place, it was necessary to look at those scientists who give up their German addresses and never come back. This holds for 8% of all migrations between 2000 and 2010. However, only a "finite" data set could allow deductions for a permanent foreign residency. In this data set, German scientists were defined as those who published 1998 to 2000 in Germany, but this could also apply to foreign scientists who move back to their home countries afterwards. Therefore, the next steps include an online survey based on the email addresses provided in Scopus. The survey's aim is to examine the motivations of scientists, especially those involved when selecting the targets for staying or moving abroad. Differences in research fields might be detectable. At the same time, the Scopus data can be validated on a larger scale using the answers in the survey.

Furthermore, a regression analysis of publication and citation data could provide evidence of a causal relationship between migration and research success based on bibliometric indicators.

References

Ette, A., Sauer, L. (2010): Auswanderung aus Deutschland. Daten und Analysen zur internationalen Migration deutscher Staatsbürger. VS Verlag für Sozialwissenschaften. 24ff.

Grubel, H.G. (1994): "Brain Drain, Economics of", in T. Husen and T. Neville Postlethwaite (eds.), *The International Encyclopedia of Education*, 1, 554–561.

Laudel, G. (2003): Studying the brain drain: Can bibliometric methods help? *Scientometrics*, 57 (2). pp 215–237.

Lowell, B.L. (2002): Skilled Labour Migration from Developing Countries: Annotated bibliography. *International Migration Papers*, 56.

Moed, H.F., Aisati, M., Plume, A. (2013): Studying scientific migration in Scopus. *Scientometrics*, 94 (3), 929–942.

Roberge, G., Campbell, D. (2012): Canadian Researchers Migration Analysis based on Scopus Author IDs. STI 2012 – 17th International Conference on Science and Technology Indicators. Montréal. 884–885. http://sticonference.org/Proceedings/vol2/Roberge_Canadian_884.pdf.

Sjaastad, L.A. (1962): The costs and returns of human migration. *Journal of Political Economy*. 70(5 Pt 2), 80–93.

Academic careers in the Netherlands: mobility of academic staff

Marije E. E. de Goede — m.degoede@rathenau.nl / Science System Assessment, Rathenau Institute, PO box 95366, 2509 CJ, The Hague (The Netherlands)

Rosalie Belder — r.belder@rathenau.nl / Science System Assessment, Rathenau Institute, PO box 95366, 2509 CJ, The Hague (The Netherlands)

Jos de Jonge — j.dejonge@rathenau.nl / Science System Assessment, Rathenau Institute, PO box 95366, 2509 CJ, The Hague (The Netherlands)

Introduction

Traditionally, the academic career system is viewed as a static system in which a large group of doctoral students and postdoctoral researchers try to move up on the academic career ladder. There is however only room for a small number of researchers at the top of the ladder; the rest will either reach a position further down the ladder, or leave academia. A career outside academia is still considered an 'atypical' career for doctorate holders. There is also a strong assumption that once a researcher moves out of academia, it is very difficult to transition back. Surprisingly, there is little information in the Netherlands to support or reject this assumption.

Available data sources do not provide information on the career steps of researchers throughout their careers, and therefore are no good indicators for understanding career mobility. More insight into the movements of academic staff into and out of academia could serve policymaking for two reasons. First, it could guide young researchers (in) building their future careers. Second, it could be an indication of the transfer of knowledge and ideas between academia and other sectors of labour, since mobility between sectors appears to be an efficient driver towards innovation (ESF, 2012)

Method

We used three different sources of data to give an estimate of the 'average' mobility from, to and between the Dutch universities. First, we reanalyzed micro-data on academic staff at Dutch universities. Based on these data we were able to map the internal mobility within one university for each of the academic positions (spanning the years 2003–2011). However, with these data it is not possible to track where departing staff is going or where incoming staff originate from. To address this question we used reported data on the mobility from, to and between universities in the period 2003–2006 and 2006–2008. Based on the mobility figures we estimated the 'average' mobility per academic position from, to and between Dutch universities. Finally, we used data on the productivity of doctoral students, in terms of time needed to finish the doctorate degree, to

complete the mobility picture. All in all, combining these three data sources gave us a new and unique indication of the mobility of Dutch academic staff and the dynamics of the Dutch academic labour market.

Results

The academic job market in the Netherlands appears dynamic on all levels. This is not a closed system; it is decisively open, be it with a strongly selective character. Each year, on average, 32 per cent of the postdoctoral researchers and teachers change positions (registered as other academic staff; OAS), around 15 per cent of assistant professors and associate professors do so, and 10 per cent of professors. For OAS, most mobility takes place between universities and the job market elsewhere; 68 per cent of the mobile OAS leaves academia, and simultaneously, an equal number of the vacant positions is taken by 'outsiders'. Associate professors are the least mobile: 28 per cent that change function leave university, and only 18 per cent of the new associate professor positions is taken by people from outside academia. Surprisingly, the position of professor is decidedly less closed. 50 per cent of the mobile professors (5 per cent of the total) leaves for a different function outside of academia, and over a third of the newly appointed professors come from outside the university.

The average age of academics in a specific position is an indication of the duration of an academic career. PhD students are on average 29.5 years when they obtain their PhD and continue their scientific career to a position within OAS. Those taking the next step, to assistant professor, are on average 37 years old. This means that the average duration of an OAS position is 7.5 years. The average assistant professor position takes 5 years. The step to professor on average takes place at the age of 49 years. All in all, it takes an average of 16.5 years to climb from PhD to professorship.

Figure 1. Origin of new academic staff (average for the period 2004–2011)

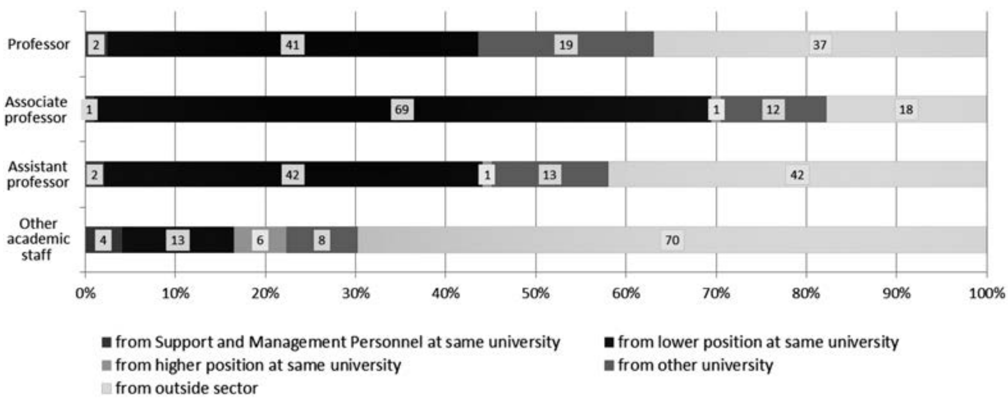
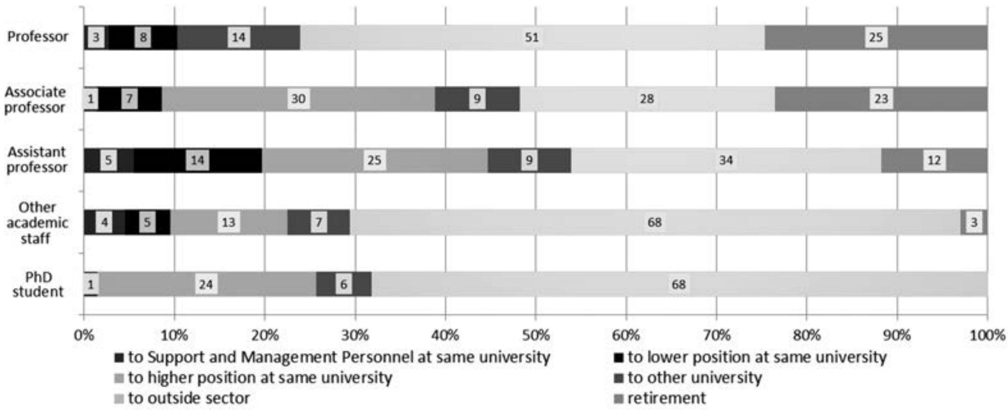


Figure 2. Destination of departing academic staff (average for the period 2003–2010)



Conclusions

Only a small part of doctoral holders continues in academia; many leave the (Dutch) universities. The primary reflex in the Netherlands remains that PhD students are trained for academia; those who do not immediately succeed are considered dropouts. This study shows there is a lively exchange with other parts of the labour market and even at the highest level there is a dynamic of mobile professors. This study gives a stimulus to consider a different agenda for academic career policy. Three elements are important. First, universities should take into account that their PhD students are also trained for other jobs in the private and public sectors. The preparation for this purpose should be strengthened. Second, the postdoc position is the most vulnerable part of our system; mobility is large and positions are temporary. Finally, mobility is not just a Dutch topic; Dutch and non-Dutch academic staff is very mobile between the Dutch academic labour market and other sectors. A more international way of thinking about mobility is necessary.

Reference

ESF (2012) How to track researchers' careers – Joint ESF-FNR Workshop Report, Luxembourg/Strasbourg 2012, http://www.esf.org/fileadmin/links/cssd/mo_fora/careers/careertrackingdraftreport.pdf

Factors related to Gender Differences in Science: A co-word analysis

Tahereh Dehdarirad tdehdari@yahoo.com / University of Barcelona, Melcior de Palau, 140, Barcelona, 08014 (Spain)

Anna Villarroya annavillarroya@ub.edu / Dept. of Public Economy, Political Economy and Spanish Economy, University of Barcelona, Melcior de Palau, 140, Barcelona, 08014 (Spain)

Maite Barrios mbarrios@ub.edu / Dept. of Methodology of Behavioral Sciences, University of Barcelona, Melcior de Palau, 140, Barcelona, 08014 (Spain)

Introduction

The issues of gender mainstreaming, the role of gender in academic appointments and evaluation, and the participation of women in science as indicators of social and economic progress have attracted substantial attention from a broad array of researchers (European Commission, [EC], 2008; Larivière, Vignola-Gagné, Villeneuve, Gélinas, & Gingras, 2011). A variety of initiatives and strategies have been designed to analyse the presence of women in science and to promote gender equality (Women in Industrial Research [WIR], 2003; Enlarge Women in Science to East [ENWISE], 2004; EC, 2013). However, systematic imbalances and differences between the sexes in the number, representation, seniority and influence of women and men in scientific studies and professions were observed (EC, 2013).

Authors have sought to explain these discrepancies in various areas of science and academia by incorporating family-related factors, personal and institutional (structural) factors, professional factors, demographic and individual issues and factors related to disciplinary fields (Fox, 2005; Fox, Fonseca, & Bao, 2011. Hunter & Leahy, 2010, Sax, Hagedorn, Arredondo, & Dicrisi III, 2002).

Nevertheless, these studies do not provide the kind of systematic and comprehensive overview of factors related to gender differences that would help to guide future research and practices in the field. Therefore, the aim of this study is to provide an analysis of the related literature using co-word analysis. This kind of analysis helps to visualize the division of a field into several subfields and the relationships between them, providing insights into the evolution of the main topics discussed in the field over the years. Using co-word analysis, the present study aims to determine the structure of the knowledge network between the terms, in order to describe the current state of the literature on the factors that influence gender differences in science.

Method

The data set consists of a corpus of 651 articles and reviews dealing with factors related to gender differences in science, published between 1991 and 2012. The data were extracted from the ISI Web of Science in February 2013, using a search that combined the principal terms related to the subject.

To carry out the co-word analysis, four sequential steps are required: extraction and standardization of the keywords, construction of the co-occurrence matrix, clustering, and visual presentation of keyword groups. Author-provided keywords were extracted from papers. The keyword plus were also used in cases that there were no author-provided keywords.

Keywords and phrases were standardized manually and finally a total of 170 keywords were selected. In order to show the development of the observed scientific field the results were divided into three periods i.e. 1991–2001 (n=164, 25.19%), 2002–2007 (n=147, 22.58%), and 2008–2012 (n=340, 52.23%).

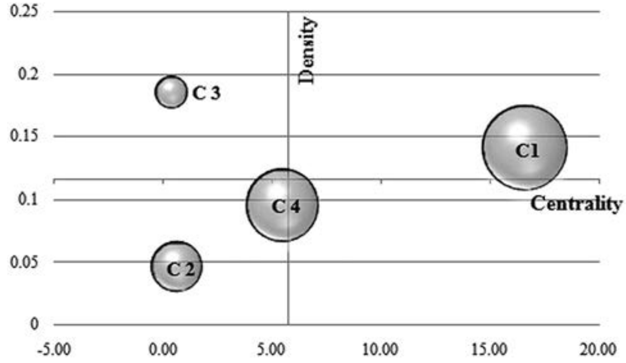
The word-document occurrence matrix for each period was automatically built via SPSS v. 20. The resulting matrices were then exported to Ucinet v. 6. In Ucinet the word-document matrix was transformed into a word co-occurrence matrix; the similarities between items were also calculated using the Jaccard similarity index. Hierarchical clustering analysis was then conducted via SPSS v. 20 using Ward's method and applied the Squared Euclidean distance as the distance measure.

Based on the dendrogram generated by the clustering algorithm, clusters of keywords were derived. The clusters were then transformed into networks in Ucinet v.6. Finally, after calculating the density and centrality for each cluster, the keywords networks were displayed in a strategic diagram using Excel. It should be considered that, in each strategic diagram, the volume of the spheres is proportional to the number of documents corresponding to each cluster.

Results

Based on the hierarchical clustering results four clusters of keywords were identified in the first period (1991–2001). For the second period (2002–2007) a ten cluster solution was obtained and finally for the third period (2008–2012) a sixteen cluster of keywords were identified. Strategic diagrams depicting the relative positions of each cluster were produced for each period in order to assist interpretation (Figures 1, 2 and 3).

Figure 1. Strategic diagram. First period (1990–2000)



Only five motor-themes appeared in the upper-right quadrant of the diagrams and regarded as mature and well-developed themes. These themes in each period were "*Gender discrimination in labor markets and universities*" (C1) in the first period, "*Career satisfaction in medicine*" (C1) and "*Academic career in sociology*" (C9) in the second period and finally "*Advancement in academic medicine*" (C9) and "*Climate and staff composition in academia*" (C11) in the third period.

Figure 2. Strategic diagram. Second period (2001–2007)

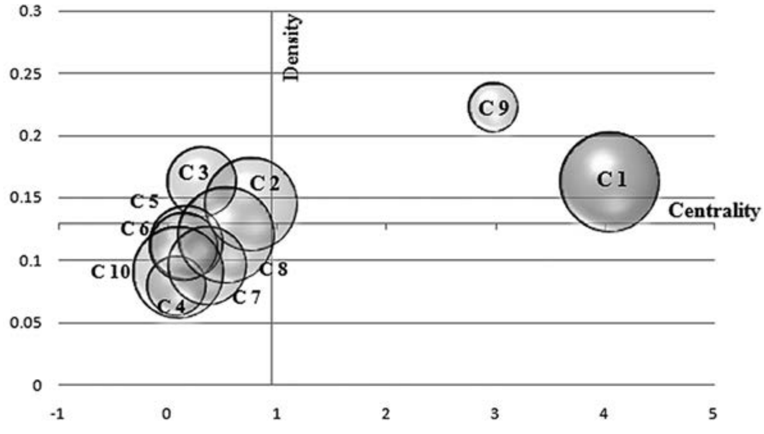
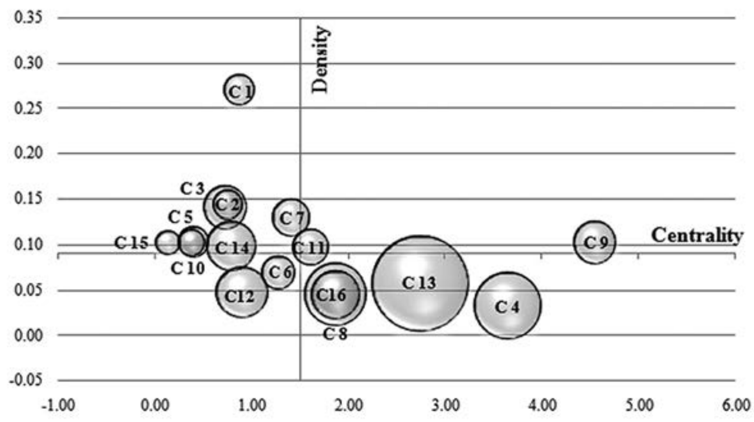


Figure 3. Strategic diagram. Third period (2008–2012)



Moreover, only one theme in the diagrams was present in all three periods: "*Mobility of women academics*". Some themes emerged and were maintained in the subsequent periods: "*Work-life balance in academia*", "*Racial discrimination at universities*" and "*Advancement in academic medicine*" appeared in both second and third periods. "*Sex differences in promotion*" appeared in both first and second periods.

Conclusion

The present results provide interesting insights into the evolution of the literature examining factors related to gender differences in science. The number of themes has experienced a significant increase over the years, ranging from four in the first period to ten in the second and to sixteen in the third period. The evolution trends of themes in strategic diagrams reveal that many themes are still immature in the studied field.

References

Enwise. (2003). *EUR 20955 – Waste of talents: turning private struggles into a public issue. Women and Science in the Enwise countries*. Luxembourg: Office for Official Publications of the European Communities.

European Commission (2008). Benchmarking Policy Measures for Gender Equality in Science. Brussels: Director-General for Research.

European Commission (2013). *She Figures 2012 – Gender in Research an innovation. Statistics and Indicators*. Luxembourg: European Union.

Fox, M. F. (2005). Gender, Family Characteristics, and Publication Productivity among Scientists. *Social Studies of Science*, 35, 131–150.

Fox, M. F., Fonseca, C., & Bao, J. (2011). Work and family conflict in academic science: Patterns and predictors among women and men in research universities. *Social Studies of Science*, 41(5), 715–735.

Hunter, L. A., & Leahey, E. (2010). Parenting and research productivity: New evidence and methods. *Social Studies of Science*, 40(3), 433–451.

Larivière, V., Vignola-Gagné, E., Villeneuve, C., Gélinas, P., & Gingras, Y. (2011). Sex differences in research funding, productivity and impact: an analysis of Québec university professors. *Scientometrics*, 87(3), 483–498.

Sax, L. J., Hagedorn, L. S., Arredondo, M., & Dicrisi III, F. A. (2002). Faculty Research Productivity: Exploring the Role of Gender and Family-related Factors. *Research in Higher Education*, 43(4), 423–46.

Citation trends and scientific fields: an empirical analysis

Ugo Finardi

u.finardi@ceris.cnr.it / CNR-CERIS, National Research Council of Italy Institute for Economic Research on Firm and Growth, via Real Collegio, 30 – I-10024 Moncalieri (TO) (Italy)

Introduction

Aim of this work is contributing to the research stream studying timing of received citations (see among others Glänzel, 1992; Egghe, 2000; Bornmann & Daniel, 2010; Egghe et al., 2011; and Bouabid, 2011; Pollmann, 2000; and Yu & Li, 2010). With respect to this literature it performs an empirical data analysis to describe the time evolution (with respect to publication year) of average citations received by journal articles of very different fields.

It does so computing data on a sample of journals from two ISI subject categories: "Chemistry, multidisciplinary" (ISI Science Edition) and "Management" (ISI Social Science edition). Time span of cited articles goes from 1999 to 2010, and that of the relative received citations goes up to 2011.

To obtain a better insight a further parallel analysis has been performed on data of ISI "Cited half life", calculating mean and median of all the values for each of the two categories for each year.

Methodology

The sample has been extensively described in Finardi (2013). It contains journals extracted with a systematic sampling from two categories chosen as being not too specific in their scope: 14 journals from "Chemistry, multidisciplinary" and 7 from "Management". The sampled journals are just above 10 % of the two categories. Data mining was performed on Thomson Reuters – ISI Web of Knowledge® in May 2012.

The methodology adapts for received citations that described in Finardi (2011). Considering articles published in year a and cited in year b, with b ≥ a, we calculate:

$$MEAN_{a,b} = \frac{CITATIONS_b}{ARTICLES_a}$$

Then, keeping constant in each summation the value of (b-a), we calculate:

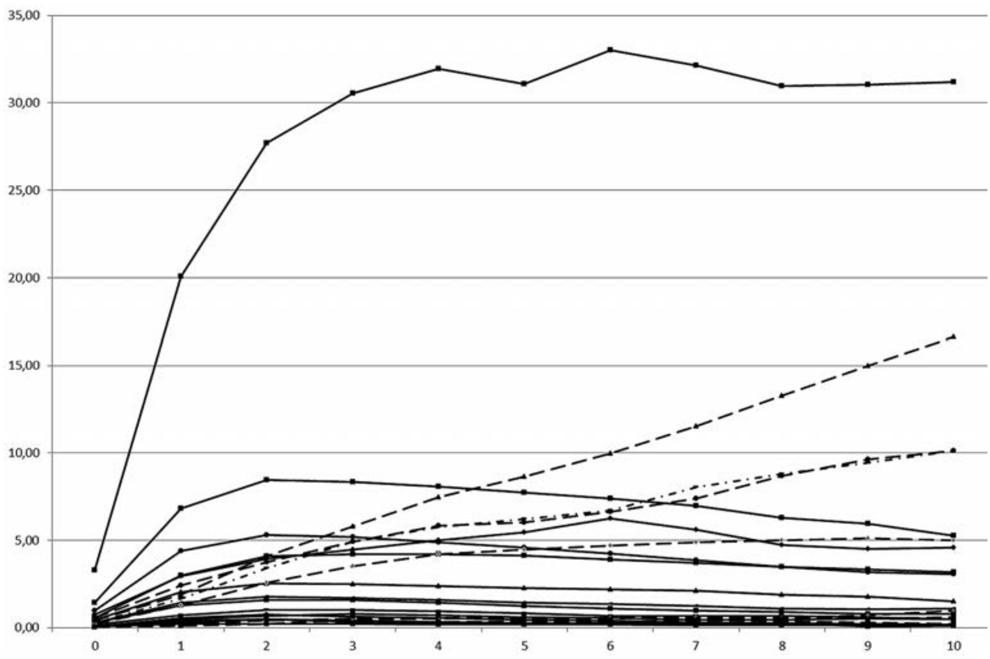
$$SUM\_MEAN_{(b-a)} = \frac{\sum_{a=1}^{n} MEAN_{a,b}}{n}$$

where n is the number of $MEAN_{a,b}$ values summed. n depends on the time distance between a and 2012, and thus goes from 2 to 12 according to the cited-citing distance in years. The final result is the plot of $SUM\_MEAN_{(b-a)}$ vs. *(b-a)*, that is, the distance in years from publication and received citation. Figure 1 presents the plots; in figure 2 the plot of the journal "Chemical Reviews" is withdrawn.

Results and discussion

The evolution of average received citations *within* categories is similar in most cases. Results *between* the two categories present instead strong differences. Most "Chemistry" journals present a peak of average received citations (11 at 2 years after publication, and one at 3 years) and then decrease steadily. Roughly considered, the higher the average received citations, the least pronounced the slope. The journal "Chemical Reviews" instead grows in the first 3–4 years, and then stabilizes at a very high value. The last journal, "Journal of Computational Chemistry", presents a (not pronounced) peak at 6 years, and then slightly decreases. "Chemical Reviews" publishes only long review articles on specific topics, and is feasibly cited as a reference for a longer time than other, experimental journals.

Figure 1. Plots of SUMb-a vs. b-a Chemistry: cont. lines; Management: dashed lines

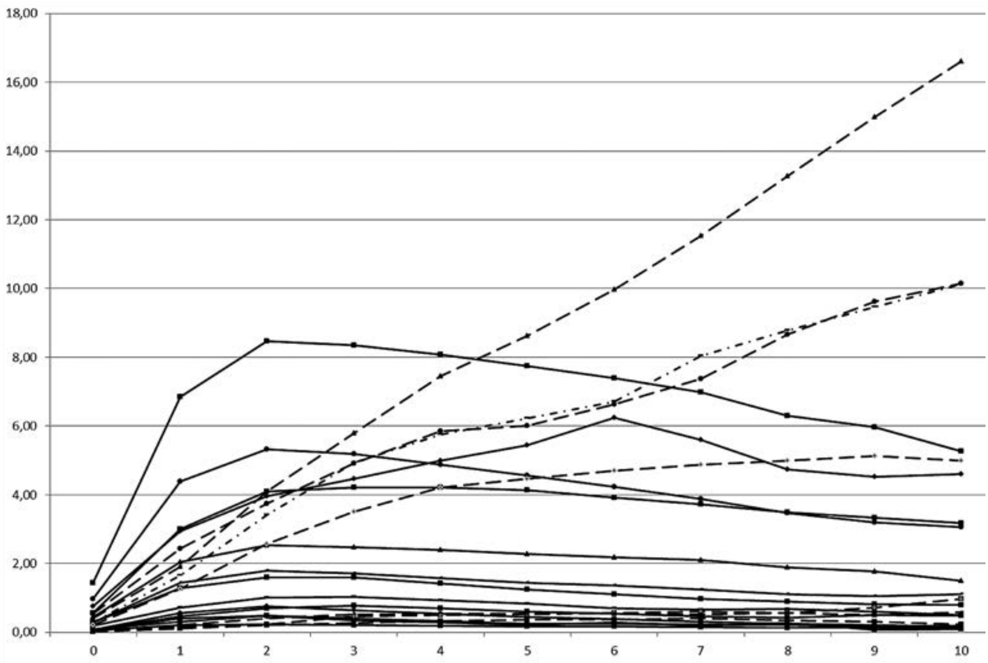


5 Management journals instead present a continuous growth of average received citations from year 0 to year 10. The last two instead stabilize around a value after some years.

Thus the different trends for the two categories suggest that citedness strongly differs between (very) different fields.

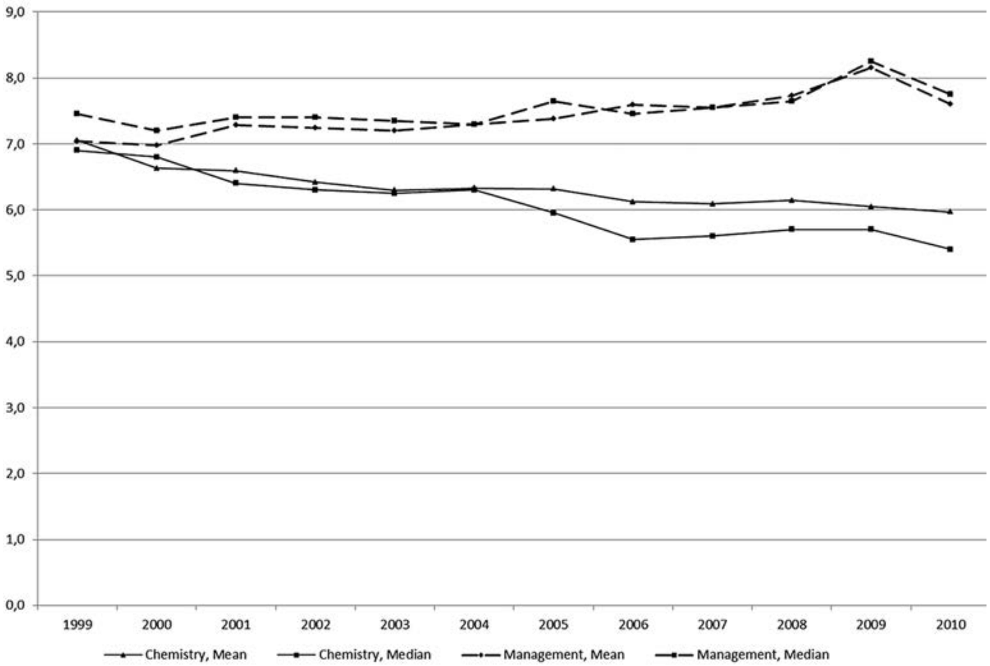
Results may in part account for the (strong) difference between categories in the average values of Journal Impact Factors. Figures show that for most cited "Management" journals the maximum value of average received citations (at 10 years) is higher than the peak of top Chemistry journals (apart from "Chemical Reviews"). Nevertheless these citations are not captured neither in JIF nor in 5 years-JIF. Thus JIF for those "Management" journals is much lower than in the other case.

Figure 2. Plots of SUMb–a vs. b–a (Chemical Reviews withdrawn from graph) Chemistry: cont. lines; Management: dashed lines



Also "Cited Half Life" analysis (figure 3) shows different trends between the two categories, while inter-category values for Mean and Median present rather consistent values. Values for "Chemistry" (continuous lines) decrease steadily, while those for "Management" (dashed lines) slightly increase, starting around similar values. This suggests that "Chemistry" journals tend to cite more recent articles as time goes by, while the opposite is happening in "Management".

Figure 3. Time evolution of median and mean of Cited Half lives for the two categories



A peak at two years has been yet described for hard sciences (though peculiar cases such as "Chemical Reviews" may deserve further exploration). Instead the trends of "Management" journals are rather un-described.

Finally, present results suggest that citations trends for a specific field should not be generalized and used to describe the general behavior of other or larger fields. Though the topic deserves further exploration, this work shows that there is not common behavior in citedness among different scientific fields.

References

Bornmann, L., Daniel, H.-L. (2010), Citation speed as a measure to predict the attention an article receives: An investigation of the validity of editorial decisions at *Angewandte Chemie International Edition, Journal of Informetrics*, 4 (1), 83–88

Bouabid H. (2011). Revisitin citation aging: a model for citation distribution and life-cycle prediction. *Scientometrics*, 88 (1), 199–211

Egghe, L (2000), A heuristic study of the first-citation distribution, *Scientometrics*, 48 (3), 345–359

Egghe, L., Bornmann, L., Guns, R. (2011), A proposal for a First-Citation-Speed-Index, *Journal of Informetrics*, 5 (1), 181–186

Finardi, U. (2011), Time relations between scientific production and patenting of knowledge: The case of nanotechnologies, *Scientometrics*, 89 (1), 37–50

Finardi, U. (2013), Correlation between Journal Impact Factor and Citation Performance: an experimental study, *Journal of Informetrics*, 7 (2), 357–370

Glänzel,W (1992). On some stopping times of citation processes. From theory to indicators. *Information Processing & Management*, 28 (1), 53–60

Pollmann T. (2000). Forgetting and the ageing of scientific publications. *Scientometrics*, 47 (1), 43–54

Yu G. and Li Y.-J. (2010). Identification of referencing and citation processes of scientific journals based on the citation distribution model, *Scientometrics*, 82 (2), 249–261

A longitudinal Perspective on Research Career Intentions during PhD Candidacy

Kalle Hauss kalle.hauss@uni-potsdam.de / Institute for Research Information and Quality Assurance, Schuetzenstr. 6a, Berlin, 10117 (Germany)

Jakob Tesch tesch@forschungsinfo.de / Institute for Research Information and Quality Assurance, Schuetzenstr. 6a, Berlin, 10117 (Germany)

Introduction

What young researchers seek after graduating from doctoral training has become an increasingly important issue to policymakers: Doctoral candidates and doctoral holders are considered key players in the creation of new knowledge (European Commission, 2005). Thus, the system of PhD training is expected to not only attract the brightest minds but also to produce scientists contributing to scientific knowledge and societal development.

The literature on career intentions of PhD students reveals two aspects: First, pursuing an academic career seems not to be the most attractive career path for a considerable proportion of doctoral candidates in different countries.

Second, there seems to be evidence that in the course of the doctoral training candidates veer off from their initial plans to enter academia and career intentions shift towards jobs in the private sector and in industry. The reasons behind this seem to be strongly related with better career prospects and higher salaries in labour markets outside academia (Ostriker, Kuh, & Voytuk, 2011; Sauermann, Roach, & Nunes Amaral, 2012).

For Germany only limited information on the career intentions of young researchers is available. Most studies show that a career in academia is a risky endeavour and that most young researchers leave academia after the PhD (Enders, 2005; Flöther & Höhle, 2013). Estimates based on official data show that only about 10 per cent of a cohort of doctorate holders in Germany will eventually climb the career ladder towards a professorship (Hauss et al., 2012, p. 38) with a weaker participation of women at every career stage (Lind, 2004). What determines a changing research intention of PhD students during PhD candidacy?

Data and methods

Using data from the German Doctoral Candidates and Doctoral Holders Panel Study ProFile we take a longitudinal perspective and analyze the dynamics of the doctoral student's career intentions over the course of the doctorate. Our sample consists of 274 PhD holders that are

observed twice during PhD candidacy and as PhD holders over a total period of three years. For comparative purposes we also provide information on the career intentions of 1,417 first year PhD candidates.

Method

We assume that the change in the intention to enter a research career is effected by a vector of variables (x) as well as an idiosyncratic error term that vary over time and individuals. Fixed-effects (FE) regression is used to estimate the effects of childbirth separately for men and women, the number of supervisors, formal agreements with the supervisor, and the frequency of exchange between the PhD student and the supervisor. The resulting model is as follows:

$$(y_{it} - \bar{y}_i) = b_1(x_{it} - \bar{x}_i) + \varepsilon_{it} - \bar{\varepsilon}_i$$

Time-constant exogenous variables (i.e. gender, field of research, ability) whether observed or unobserved are cancelled out. Hence, unobserved heterogeneity is not a problem in our model.

Sample for fixed effects regression

We base our estimates on three time points during PhD candidacy: The first and second observation mark the last two years of PhD candidacy. The third observation marks the time of graduation one year later. Thus, we base our estimates on those changes that are observable within the last two years of the PhD candidacy. While changes are measured only during PhD candidacy, the dependant variable is observed once during PhD candidacy (whenever the candidates enter the panel) and at graduation.

Variables

Our dependent variable is surveyed using the following question: "*How closely would you like your future career to be connected with the following areas*?" The respondents rated seven different vocational fields on a five-point-scale of which we choose "research" for the analysis (Figure 1)

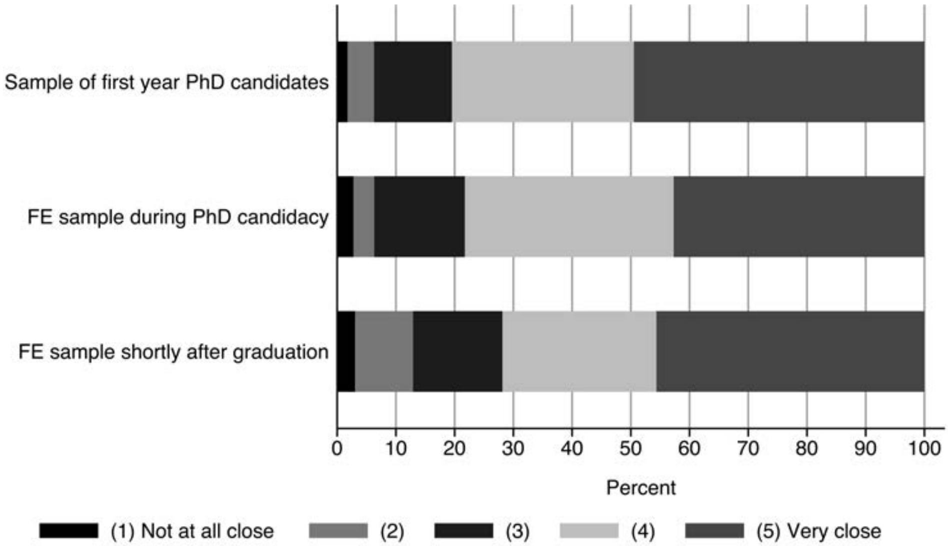
Results

Descriptive Results

The intention to pursue a research career is widespread among first year PhD candidates, over the course of PhD candidacy as well as after graduation (Figure 1). At every stage almost 50% of the respondents indicate a very strong intention to pursue a research career.

The intention for a research career remains relatively stable both over the course of the PhD as well as between candidacy and after graduation. 47% of the respondents report the same level of intention during candidacy and at graduation. A one point increase is observed for 17% of the sample and one point decrease for 19%.

Figure 1. Intention to pursue a research career



Source: ProFile 2012

Results of fixed effects model

The intention to enter a research career is independent of changes in supervision (Table 1). Women show a stronger intention towards a research career when they have not yet given birth. Giving birth to a child in the last years of PhD candidacy causes a significant drop in the intention towards a research career. The effect is significant for women but not for men.

Table 1. Coefficients of the fixed effects linear regression model for Δintention towards research career

	ΔCareer intention towards research
Mother	-0.90*
Father	-0.32
No. of supervisors	-0.19
Written agreements	0.15

... Continuation Table 1

	ΔCareer intention towards research
Exchange frequency	0.01
No. of observations	506
No. of persons	274
Rho	0.56

* p-value<0.05, own calculations
Source: ProFile 2012

Discussion

Our results show that changes in the supervision during last two years of PhD candidacy do not have an effect on career intentions for research of PhDs. The small changes which are observable can be explained for women becoming mothers during the final years of their PhD candidacy. One side result of our analysis is that those dropping out of PhD training most likely are not those candidates with a low intention for research in the first place. Career intention of PhD dropouts appear to be quite similar to those who finish their PhD training successfully.

References

Enders, J. (2005). Border crossings: Research training, knowledge dissemination and the transformation of academic work. *Higher Education*, 49(1–2), 119–133. doi:10.1007/s10734-004-2917-3.

European Commission. (2005). *Commission Recommendation on the European Charter for Researchers and on a Code of Conduct for the Recruitment of Researchers*. C 576. Brussels: European Commission.

Flöther, C., & Höhle, E. A. (2013). B1.3.2 Karrierewege Promovierter innerhalb und außerhalb der Wissenschaft – Ergebnisse der KOAB-Absolventenstudien. In *Bundesbericht Wissenschaftlicher Nachwuchs 2013. Statistische Daten und Forschungsbefunde zu Promovierenden und Promovierten in Deutschland* (1st ed., 288–295). Bielefeld: W. Bertelsmann Verlag.

Hauss, K., Kaulisch, M., Zinnbauer, M., Tesch, J., Fräßdorf, A., Hinze, S., & Hornbostel, S. 2. (2012). *Promovierende im Profil:: Wege, Strukturen und Rahmenbedingungen von Promotionen in Deutschland. Ergebnisse aus dem ProFile-Promovierendenpanel* (iFQ-Working Paper No. 13). Berlin, Germany. Retrieved from http://www.forschungsinfo.de/Publikationen/Download/working_paper_13_2012.pdf.

Lind, I. (2004). *Aufstieg oder Ausstieg? Karrierewege von Wissenschaftlerinnen. Ein Forschungsüberblick. cews.Beiträge: Vol. 2*. Bielefeld: Kleine Verlag.

Ostriker, J. P., Kuh, C. V., & Voytuk, J. A. (Eds.). (2011). *A Data-Based Assessment of Research-Doctorate Programs in the United States: Committee on an Assessment of Research Doctorate Programs*. Washington, D.C.: National Academies Press.

Sauermann, H., Roach, M., & Nunes Amaral, L. A. (2012). Science PhD Career Preferences: Levels, Changes, and Advisor Encouragement. , 7(5), e36307. doi:10.1371/journal.pone.0036307.

Governance by indicators in German higher education: A look at internationality

Michael Jaeger — m.jaeger@his.de / Department "Steering, Funding, Evaluation", HIS-Institute for Research on Higher Education, Goseriede 9, 30159 Hannover (Germany)

Susanne In der Smitten — smitten@his.de / Department "Steering, Funding, Evaluation", HIS-Institute for Research on Higher Education, Goseriede 9, 30159 Hannover (Germany)

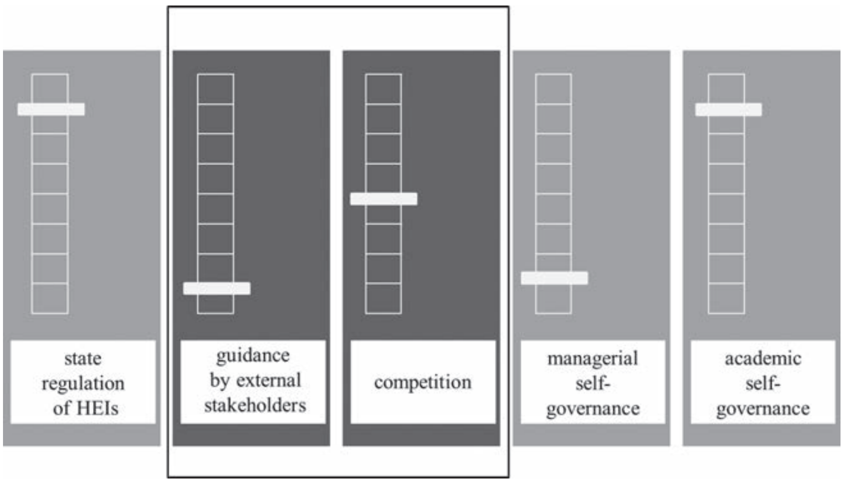
Johannes Wespel — wespel@his.de / Department "Steering, Funding, Evaluation", HIS-Institute for Research on Higher Education, Goseriede 9, 30159 Hannover (Germany)

Introduction

This contribution investigates the use of indicators of internationality in German higher education (HE) governance. It focuses on the use of such indicators in target and performance agreements and performance-based funding formulas.

HE governance can be conceptualised using the "equalizer model" advanced by De Boer/Enders/Schimank (2007). A HE governance system is then defined by the specific constellation of five equalizer "controls", i.e. dimensions of governance:

Figure 1. The "Governance Equalizer"



Performance indicators can be applied in all of these dimensions, but are particularly common in guidance by external stakeholders and competition. These two dimensions will therefore be looked at in more detail below.

Instruments of governance in guidance by external stakeholders and competition

Different instruments of steering and communication are applied in each dimension of governance. Two such instruments will be investigated here:

- target and performance agreements (TPAs) as a means of guidance by external stakeholders;
- performance-based funding formulas (PBFs) as a means of both guidance by external stakeholders and competition.

TPAs are used to negotiate and define performance goals. Indicators are applied in order to evaluate the degree to which goals have been achieved. In PBFs, indicators are applied to measure and compare past performance of several HE institutions. Funds are allocated based on the measurement results.

The case of internationality

Internationality in performance measurement

Internationality is a cross-sectional area of performance embedded in the core processes of HE institutions (teaching and learning and research). This means that there are two ways in which aspects of internationality can be captured by indicators:

- Internationality is subsumed under more general indicators, such as number of students or amounts of third party funds.
- Dedicated indicators measuring internationality-related performance are defined, such as number of foreign students or amounts of international third-party funds.

Methodology

The findings presented below are based on content analyses which were conducted using official descriptions of TPAs between Länder governments and HE institutions, and state models of PBF. The work is related to projects in which state funding models were evaluated (e.g. Jaeger/In der Smitten 2010) or in which governance structures were examined from a theoretical point of view (In der Smitten/Jaeger 2012).

- TPAs from all 16 Länder were content-analysed in 2012/2013.
- PBFs were analysed in 2010; out of 13 Länder that used such models, nine Länder used dedicated indicators of internationality.

Table 1 below gives an overview of the findings concerning indicators with a special focus on internationality.

Table 1. Indicators of internationality in target agreements between Länder governments and universities (state: 2013), and in state-level performance-based funding formulas (state: 2010)

	Indicator	Applied in target and performance agreements		Applied in state funding formula models (number of models in brackets, N=9)	
		Full / Technical Universities	Universities of Applied Sciences	Full / Technical Universities	Universities of Applied Sciences
Teaching & learning	Number/share of incoming students	✓	✓	✓ (6)	✓ (6)
	Number/share of outgoing students	✓	✓	✓ (2)	✓ (2)
	Number/share of foreign graduates	✓	✓	✓ (3)	✓ (2)
	Number/share of students in international study programmes	✓	✓	✗	✗
	Number of study programmes with international components	✓	✓	✗	✗
	Number of scholarship holders of the German Academic Exchange Service	✓	✗	✗	✗
Research	Number of internationally cooperative projects / graduate schools	✓	✓	✗	✓ (1)
	Volume of international third party funds	✓	✓	✗	✗
	Number of outgoing professors	✗	✓	✗	✗
	Number of guest professors	✗	✓	✗	✗
	Number of foreign PhD students	✓	✗	✗	✗
	Number of scholarships for international researchers	✓	✗	✓ (3)	✗
Other	Number of foreigners in staff	✓	✓	✗	✗
	Number of outgoing management staff	✗	✓	✗	✗
	Budget/Third party funds for activities of internationalisation	✓	✓	✗	✗

Findings

- Counts of incoming students (and, to a lesser extent, graduates and outgoing students) are the most common indicators across instruments and models.
- With one exception, PBFs only count individuals (and their achievements), whereas TPAs also count structures (programmes, projects etc.).
- Indicators of internationality in research are much less common in PBF than TPAs.
- Further analyses show: TPAs include a multitude of qualitative criteria besides measurable indicators to promote internationality, such as "involvement in int. organisations" or "improvement of support for foreign students".

Discussion

- Indicators of internationality are much more diverse in TPAs than in PBFs. This might be due to the nature of the predominant dimensions of governance associated with them: PBFs are based on competition, and competition needs comparability and thus standardisation of targets. TPAs emerge from a direct bilateral relation. They are more suitable to promote institutional differentiation; thus they can make use of a much wider range of indicators than PBFs.
- Despite being widespread, the financial significance of indicators of internationality is rather low. This might be due to the special nature of internationality as a cross-sectional area of performance which is indirectly covered by other, more general indicators as well.

Conclusion and outlook

The analyses provide a fertile ground for research into the interrelations between TPAs and PBFs. Further investigations into the complementary use of different indicators included in the two steering instruments appear worthwhile: Do PBFs actually promote international experience of individuals at all institutions, while TPAs encourage the development of distinct support structures and (internationality) profiles of the HE institutions?

References

In der Smitten, S. & Jaeger, M. (2012). *Ziel- und Leistungsvereinbarungen als Instrument der Hochschulfinanzierung* (HIS:Forum Hochschule Nr. F16/2012). Hannover: HIS GmbH.

Jaeger, M. & In der Smitten, S. (2010). *Evaluation der leistungsbezogenen Mittelvergabe an die Hochschulen in Mecklenburg-Vorpommern* (HIS:Forum Hochschule Nr. F04/2010). Hannover: HIS GmbH.

De Boer, H., Enders, J. & Schimank, U. (2007). On the Way towards New Public Management? In D. Jansen (Ed.), *New Forms of Governance in Research Organizations* (137–152). Dordrecht: Springer.

Granular and Exclusive Clustering of Scientific Journals with Constraints by Essential Science Indicators[1]

Hirotaka Kawashima kawashima.h.ac@m.titech.ac.jp / Department of Engineering Fundamentals and Strategic Planning, Tokyo Tech, Ookayama 2-12-1, Meguro, Tokyo, 1528550 (Japan)

Masashi Shirabe shirabe.m.aa@m.titech.ac.jp / Department of Engineering Fundamentals and Strategic Planning, Tokyo Tech, Ookayama 2-12-1, Meguro, Tokyo, 1528550 (Japan)

Introduction

A new classification of scientific journals is proposed. When we want to tally scientific activities from journal publication database, we usually do it with respect to each research domain. There are two types of research domain classifications, non-exclusive ones and exclusive ones. The former is useful to analyze a specific research area or keywords. And the latter is useful to analyze portfolios. One of the most widely-used exclusive classifications is Thomson Reuter's Essential Science Indicators (ESI). ESI is a set of 22 labels of scientific domains on qualified journals in Web of Science (WoS). But in many cases, "22" is too few (e.g., during analysis of portfolios of universities for institution management or for public funding). Some authors solved this problem by creating completely new classifications (Rosvall, 2008, Raflos, 2009 and a series of Leydesdorff's, etc.) But for practical purposes, due to a lack of compatibility with the existing classification, we cannot refer and connect our data to previous scientometric or bibliometric studies or public reports that use ESI classification. Here, we suggest systematic clustering procedures to make subdomains of existing ESI. At clustering citation network, we have "coarse-grained" except for focused ESI. It can add information from an outside network to the clustering inside the focused ESI.

Method

Data

All data in this work is sourced from an offline version of WoS. A citation network of scientific papers recorded in 2011 is used. Available papers are limited by document type: {'Article', 'Letter', 'Note', 'Proceedings Paper', and 'Review'}; journal publication type {'Journal', 'Book in series'}; and ESI {labeled one of 21 ESI, except 'multidisciplinary' at 2011}. Citations to papers published between 2007 and 2011 are included. Journal self-citations are omitted.

1 This work was supported by JST/RISTEX research-funding program "Science of Science, Technology and Innovation Policy".

Preprocessing

At first, for journal aggregation, we identified journals with corresponding ISSNs, journal titles, or abbreviation titles. The journals in the bottom 20% of summation of the number of "cites" and "cited" were then removed. That left 3,339 available journals.

Clustering

To ascertain journal community constraints within each ESI, we processed a 3,339 by 3,339 adjacency matrix. First, one ESI was selected as a focus (e.g., Neuroscience & Behavior). Second, elements of the matrix not belonging to the focused ESI are agglomerated by each ESI (e.g., the matrix size changes to ESI "21–1" + "218" journals belonging to "Neuroscience & Behavior"; see Table 1). The matrix is then normalized as the summation along raw is "1". Finally, network clustering is applied to this normalized matrix. These procedures are applied to each ESI (21 adjacency matrices).

While clustering, we adopt the Louvain method. The method is based on modularity-optimization (Blondel, 2008). For implementation, we use Gephi (Bastian, 2009).

Result

Table 1. Cluster Results

ESI	#journal, #comms	Max, Min
Agricultural Sciences	[194, 7]	{64, 6}
Biology & Biochemistry	[373, 8]	{91, 16}
Chemistry	[439, 8]	{96, 1}
Clinical Medicine	[1726, 13]	{331, 1}
Computer Science	[223, 7]	{58, 14}
Economics & Business	[357, 9]	{122, 1}
Engineering	[653, 10]	{111, 12}
Environment/Ecology	[272, 6]	{84, 14}
Geosciences	[292, 8]	{75, 6}
Immunology	[87, 5]	{41, 6}
Materials Science	[229, 8]	{57, 12}
Mathematics	[362, 7]	{85, 5}
Microbiology	[113, 5]	{45, 3}
Molecular Biology & Genetics	[243, 7]	{79, 1}
Neuroscience & Behavior	[218, 7]	{63, 5}
Pharmacology & Toxicology	[189, 8]	{66, 1}
Physics	[265, 9]	{50, 5}
Plant & Animal Science	[603, 9]	{136, 1}
Psychiatry/Psychology	[434, 13]	{86, 1}

… Continuation Table 1

ESI	#journal, #comms	Max, Min
Social Sciences, general	[866, 10]	{263, 12}
Space Science	[46, 5]	{33, 2}

Table 1 shows the number of [journals, clustered communities] and the Maximum and Minimum number of journals of communities.

Discussion

The right column of Table 1 shows that our clustering did not cause diffusion and revealed the preferable size of communities. By visual confirmation, the member of each community reflects the subdomain in the ESI. We also tried to evaluate this result with using another WoS category information "Subject Category" (that is non-exclusive). We counted whether the journals that belong to the same community have the same Subject Category (hit rate, in a manner) and the journals that belong to the different community have the different Subject Category (correct rejection rate, in a manner) (Klavans, 2006). In this brief test, we prepared another clustering result that applied to each of 21 ESI with bibliographic coupling (and of course this network doesn't need to be "coarse-grained") for comparison. In consequence, our suggested clustering result is superior to the result of bibliographic coupling network like previous study, evaluate bibliographic coupling (Shibata, 2009).

For future study, with more precise verification, once the subdomain is fixed, we can re-cluster each journal without the ESI label or each paper recorded in journals belonging to the "multidisciplinary" ESI or multidisciplinary subdomains. The portfolio at the country- or institution-level activities will be analyzed for policy implications.

Reference

Bastian, M., Heymann, S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of community hierarchies in large networks. *Journal of Statistical Mechanics*, P1008

Klavans, R. & Boyack, K.W. (2006). Identifying a Better Measure of Relatedness for Mapping Science. *Journal of the American Society for Information Science and Technology*, vol. 57, iss. 2, 251–263

Raflos, I. & Leydesdorff, L. (2009). Content-Based and Algorithmic Classifications of Journals: Perspectives on the Dynamics of Scientific Communication and Indexer Effects. *Journal of the American Society for Information Science and Technology*, vol. 60, iss. 9, 1823–1835

Rosvall, M. & Bergstrom C.T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 4, 1118–1123

Shibata, N., Kajikawa, Y., Takeda, Y. & Matsushima, K. (2009). Comparative study on Methods of Detecting Research Fronts Using Different Types of Citation. *Journal of the American Society for Information Science and Technology*, vol. 60, iss. 3, 571–580

Bibliometrics Analysis of Publication Trends in Young and Fast Developing Universities

K. A. Khor mkakhor@ntu.edu.sg / Research Support Office and Bibliometrics Analysis, Nanyang Technological University, #B4-01, Block N2.1, 76, Nanyang Drive, Singapore 637331 (Singapore)

K. N. Santoso Ketut@ntu.edu.sg / Research Support Office and Bibliometrics Analysis, Nanyang Technological University, #B4-01, Block N2.1, 76, Nanyang Drive, Singapore 637331 (Singapore)

L.-G. Yu mlgyu@ntu.edu.sg / Research Support Office and Bibliometrics Analysis, Nanyang Technological University, #B4-01, Block N2.1, 76, Nanyang Drive, Singapore 637331 (Singapore)

Introduction

In the last two decades, Singapore has successfully moved from the process of building a nation to becoming a global city truly embracing the concept of knowledge based economy (Amaldas, 2009). Nanyang Technological University (NTU), as a relatively young university (established in 1991), has a unique role and position in Singapore's academic ecosystem (Andersson, 2010).

As NTU has undergone vast development and evolution since its birth, the S&T output of NTU in terms of scientific publications in international journals, their impact, which comes under citation analysis, in respective fields and their trend are indeed vital factors to be identified in order to guide future planning.

The purpose of this study is to present the status of Nanyang Technological University in the science and technology world in terms of scientific productivity and its influence over the science and technology community, and to provide the trends of its output and impact in comparison to other similar fast developing universities around the world, in finding possible crucial integrants for further research. This study also presents the collaboration trends of Nanyang Technological University with local and international institutions over the last 20 years.

Methodology

Thomson Reuters' Web of Science database is the main source for extracting publication data of Nanyang Technological University for the period of 1991 to 2012. Over 43,000 publication records were downloaded for analysis. The publications of some institutions established in the contemporary of NTU, like Hong Kong University of S&T (HKUST, est. 1991), Pohang University of S&T (POSTECH, est. 1986), Universidad Carlos III de Madrid (UC3M, est. 1989), Pompeu Fabra Univ. (UPF, est. 1990), City University of Hong Kong (CityU, est. 1984), Queensland University of Technology (QUT, est. 1990) and University of Plymouth (UOP, est. 1992) and other institutions like University of Ulm (ULM), Korea Advanced Institute of Science and Tech-

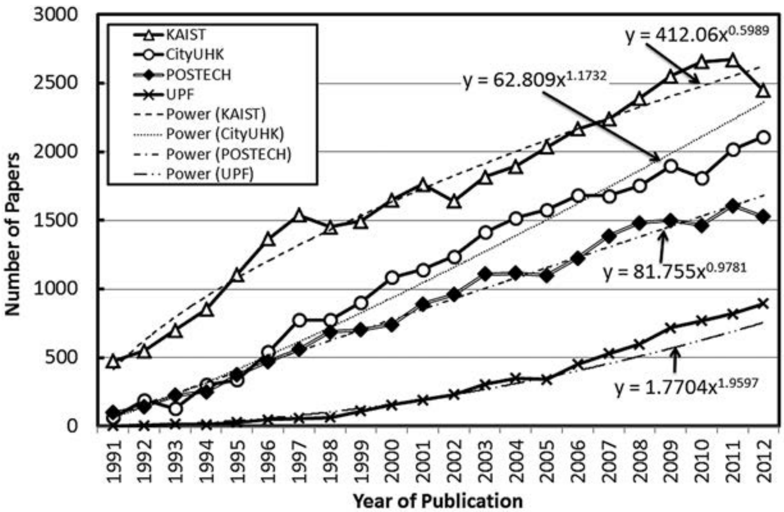
nology (KAIST) were also downloaded for output and impact comparison. During the local or international collaboration analysis, every publication was assigned to all the countries or institutions involved in the publication. Hence, such publications were replicated in the analysis. MS Excel was used for data analysis in obtaining the bibliometric indicators and plotting the graphs for the study.

Results and Analysis

Scientific output and Impact of NTU and Selected Institutions

Figure 1 shows the scientific output of NTU, KAIST, CityU, POSTECH and UPF and their respective regressions by power law for the period 1991–2012. It can be seen that NTU's publication output over the years can be roughly regressed to a power law with the power slightly above 1 (~1.2), and only UPF has a higher power (~1.96) than NTU. This is because UPF has a much lower number of paper in 1991 (4 papers) compared to that of NTU (125 papers in NTU). From 2000 to 2012, NTU's publication output has an annual increment ratio of ~10% ($\sqrt[13]{\frac{4034}{1181}} - 1 = 9.91\%$).

Figure 1. Publication trends of NTU, KAIST, CityU, POSTECH and UPF: 1991–2012



A comparison of the yearly numbers of NTU's publications with those of HKUST, POSTECH, UC3M, UPF, KAIST, ULM, CityU, QUT and UOP shows that the No. of papers published by NTU is much higher than the other institutions, especially after 1999.

Yet, the citations-per-paper (CPP) analysis shows that NTU only has slightly higher citations-per-paper values than NC3M from 2002 onwards, and are much smaller than that of HKUST, POSTECH and UPF, KAIST, ULM, CityU, QUT and UOP. This indicates that there is room for improvement for NTU in respect of its publications' impact.

Collaboration Analysis for NTU

Figure 2 shows the numbers of papers solely authored by NTU staff and the numbers of NTU's collaborated papers in the last 22 Years. The percentages of NTU collaborated papers in the last 22 Years are also shown in Figure 2. It can be found that the numbers of NTU's collaborated papers and the percentages of NTU collaborated papers have all steadily increased in the last 22 Years, from 15.20% in 1991 to 62.15% in 2012, with a crossing point (50%-50%) around 2009.

Figure 2. NTU's Collaboration trends over the last 22 years

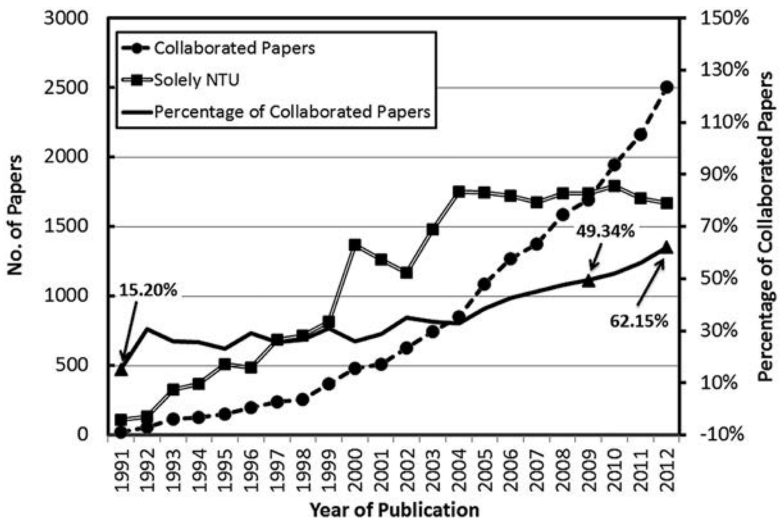
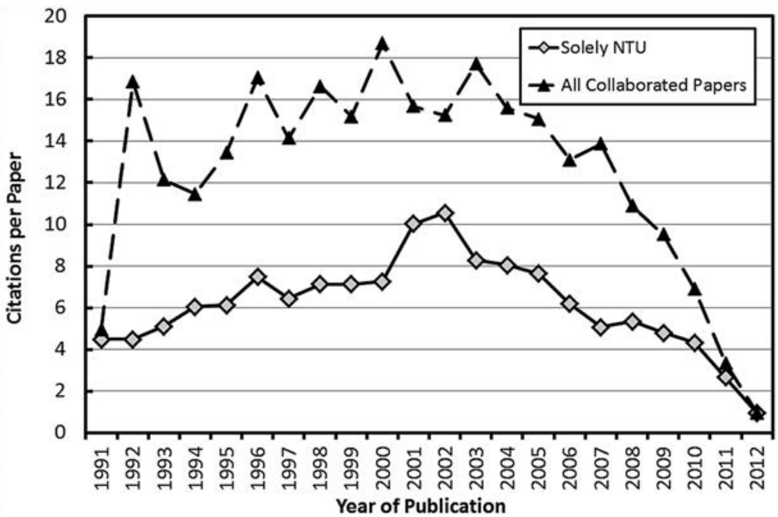


Figure 3 provides the CPP for solely NTU papers compared to that of all collaborated papers. One can find that the CPP of solely NTU papers are much lower than that of collaborated papers in all years, indicating that collaboration can greatly increase the impact of a paper to the research society.

Figure 3. Citations per paper for Solely NTU papers compared to that of collaborated papers.



The investigation of NTU's local collaboration shows that A*Star institutions have the largest collaboration papers with NTU, followed by NUS. Other institutions and industry companies has the lest portion of number of collaboration papers with NTU.

For international collaboration, China and Hong Kong institutions together contributed the largest portion of NTU's international collaboration papers, followed by European institutions, and institution in Asia (Excluding China & Hong Kong), USA and Canada, and those in Australia and New Zealand.

The analysis of NTU's collaboration with institutions in Europe, Asia (excluding China & Hong Kong) and Australia in last 5 years reveals that, University of London, Imperial College of London and University of Cambridge are the top collaborating institutions in Europe; and Indian Institute of Technology (IIT), National Taiwan University and Bilkent University (Turkey) are the top collaborating institutions in Asia (excluding China and Hong Kong). University of New South Wales, University of Sydney and Curtain University of Technology are the top collaborating institutions in Australia.

Conclusion Remarks

In this paper, bibliometric analysis is applied to depict the publication and citation evolution trends of Nanyang Technological University (NTU) through the last two decades. The results are compared to some similar young and fast developing institutions. It is found that NTU's number

of publications increased in a power law with the power slightly above 1 (~1.2), which is lower than that of Pompeu Fabra University (~1.96), but higher than those of the other institutions surveyed in this research. Although NTU has more publications and citations than the other young institutions in this study, its citations per paper values are markedly lower than those of the selected institutions. Collaboration analysis reveals that NTU has a steadily increasing collaboration ratio from 15.20% in 1991 to 62.15% in 2012.

References

Amaldas, M. (2009) The Management of Globalization in Singapore: Twentieth Century Lessons for the Early Decades of the New Century. *Journal of Alternative Perspectives in the Social Sciences* (2009) Vol 1, No 3, 982–1002.

Andersson, B. and Mayer, T. (2010). Singapore and the Nanyang Technological University – A young country with a young university on the move. Biointerphases, 5(3), FA9-FA14.

Research institutes' locally anchored collaboration and international co-authorship networks

Katarina Larsen katarina.larsen@abe.kth.se / Division of History of Science, Technology and Environment, KTH Royal Institute of Technology, Teknikringen 72, Stockholm, 100 44 (Sweden) / NIFU – Nordic Institute for Studies in Innovation, Research and Education (Norway)

Peter Sjögårde sjogarde@kth.se / Unit for Publication Infrastructure, KTH Royal Institute of Technology, Osquars backe 25, Stockholm, 100 44 (Sweden)

Background and Purpose

The aim of this study is to examine publication patterns over time and also analyze international collaboration in the broader context of technical-industrial institutes in Nordic countries. The wider motivation for the study is to get a better understanding of the collaboration patterns of technical-industrial institutes with local and national universities and other higher education institutions. A second motivation is that the analysis of organizations' co-publication networks is relevant to better discern international collaboration patterns.

Furthermore, the analysis of collaboration networks is relevant in understanding institutes' activities and future roles (EC 2007). Previous bibliometric analysis of collaboration patterns in the Swedish regional context, demonstrate concentration of R&D activities to the urban areas (Danell and Persson 2003) but has not focused on research institutes.

There are numerous studies of inter-organizational collaboration in biotechnology networks (Powell et al. 1996). It is warranted with further analysis of scientific collaborative undertakings in other technology areas and analysis of technical-industrial institutes provides a broad range of technology areas. Previous analysis of institutes (Larsen and Sjögårde 2012) raised the methodological question about analysis of organizations' networks and the level of organization and the level of the field of publication channels. The focus of the current study is on collaboration patterns of institutes in terms of local and international co-authorship.

The study here represents an analysis of a group of institutes included in RISE (Research Institutes of Sweden), which are to a large extent concentrated near university locations. These institutes are organized in the four groups including Innventia (representing forestry), Swedish ICT (software, fibre optics and hardware), SWEREA (materials, process, product and production technology) and SP (development and evaluation of technologies, material, products, and processes).

Method and data

The bibliometric analysis was conducted by searching Web of Science for publications containing the addresses of the RISE institutes. The records were downloaded and processed in Bibexcel, Excel and Gephi.

Frequencies of the publication output were calculated for the institutes within the RISE group for the time period 2002 to 2011. Articles, Proceedings Papers, Reviews and Letters were included in the publication count. Totally 2542 publications were analyzed.

To analyze the collaboration network of the RISE group frequencies of organizations and countries were calculated and co-occurrence visualizations conducted. Co-occurrence analysis of affiliated organizations was conducted using the method described in Persson et al. (2009). This was combined with a review of a sample of the articles to understand what types of research that result in co-publications.

Results

The results show an increasing publication frequency by the institutes, from 235 publications in 2002 to 289 publications in 2011. However, if using fractionalized counts, the publication frequency has been rather stable during the analyzed time period. Hence, the number of co-authored publications has grown. Figure 1 shows an increase of the number of publications co-authored by five authors or more.

Also the international collaboration, measured by the share of internationally co-authored publications, has been augmenting, from about 31 percent in 2002 to about 43 percent in 2011. Figure 2 shows the share of international co-authored publications for the four groups of institutes.

Figure 1. Number of RISE publications and number of co-authors per publication, 2002–2011.

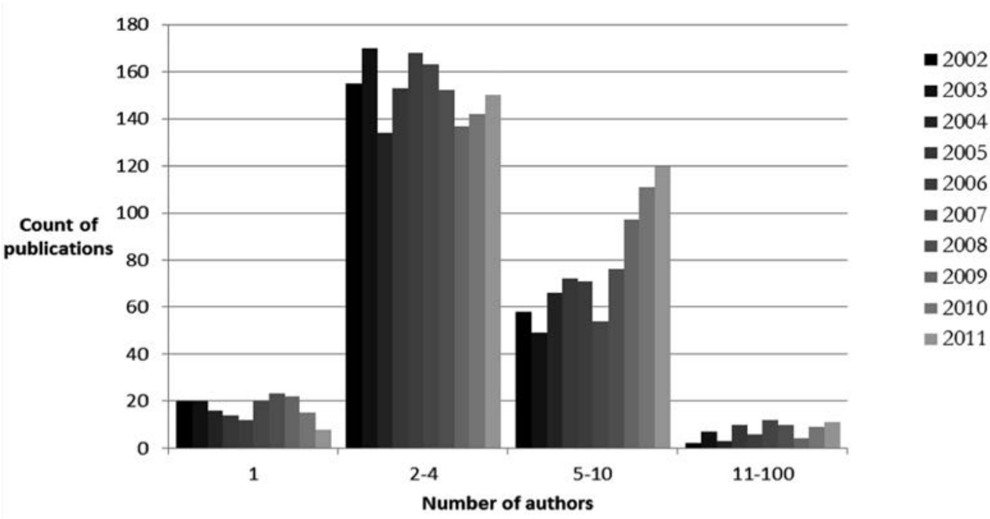
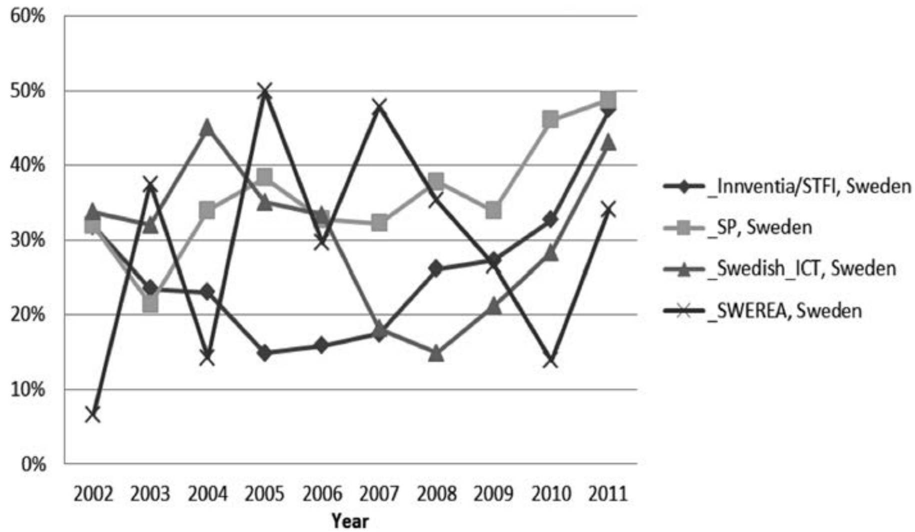


Figure 2. Share of internationally co-authored publications



The analysis shows that Swedish universities are most common collaborators for the institutes, of which KTH Royal institute of technology is the most frequent occurring university in the address field (521) publications.

Results from the review show examples of collaborative research between several institutes, e.g. combining sensor technology and measurement of polymer surface. Other examples include institute-university collaboration between KTH and Innventia (lab analysis of pulp- and paper properties) and between Swerea and national universities on CO2-reduction in steelmaking processes (with Luleå Technical University). Internationally co-authored articles covered e.g. research about aircraft simulations (with Imperial College London).

Discussion

This study examines co-authorship links between research institutes and universities, focusing on technical research institutes. The collaboration patterns show a substantial collaboration resulting in increasing numbers of co-authored scientific papers. Even though the number of internationally co-authored papers has been increasing the main collaboration partners are still local technical universities in Sweden.

Moreover the results about local ties between institutes and universities prompts further studies of knowledge transfer and mobility between these sectors and the relation between strong local research environments and international research collaboration and recognition.

Co-authorship with the private sector is not as frequent as with the university sector. However, further analysis of institute co-authorship profiles can be of interest in examination of research institutes' publication practices with the private sector. Studies of the areas of research where these institutes are active can also contribute to further understanding of knowledge transfer between sectors and how converging technology and emerging fields of technology are represented in collaborations between institutes and across traditional sector boundaries.

References

Danell, R. & Persson, O. (2003). Regional R&D activities and interactions in the Swedish Triple Helix. *Scientometrics*, 58(2), 205–218.

EC (2007) European Commission Working paper: The future of key research actors in the European research area. Directorate General for Research Cooperation, EU 22962EN.

Larsen, K. & Sjögårde, P. (2012) Swedish and Nordic institute collaboration networks over time – between institute ego networks and fields. Paper presented at Nordic workshop 17th Nordic Workshop on Bibliometrics and Research Policy, Helsinki 2012.

Persson, O., Danell, R. & Wiborg Schneider, J. (2009). How to use Bibexcel for various types of bibliometric analysis. In: Celebrating scholarly communication studies. Leuven, Belgium: International Society for Scientometrics and Informetrics, 9–24. Available at: http://umu.diva- portal.org/smash/record.jsf?pid=diva2:232838.

Powell, W.W., Koput, K.W. & Smith-Doerr, L. (1996) Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology", *Administrative Science Quarterly*, 41(1), 116–145.

Scientific publishing in West Africa: a comparison with BRICS

Eustache Mêgnigbêto eustachem@gmail.com / Bureau d'études et de recherches en science de l'information, 09 BP 477 Saint Michel, Cotonou (Benin) / University of Antwerp, Information and Library Sciences, Venusstraat 35, B-2000 Antwerp (Belgium)

Introduction

West Africa counts 15 countries[i], all members of the Economic Community of West African States (ECOWAS), a regional economic integration organisation. In early 2012, the region adopted the ECOWAS Policy on Science and Technology (ECOPOST) that recognized the role of science, technology and innovation in regional integration and life conditions improvement. Our objective is to give a view of the West African research landscape on the eve of the ECOPOST adoption. The research question is: How does West Africa compare to emergent or new industrialized countries like Brazil, India, China and South Africa with respect to scientific publishing?

Methods and data

In April 2012, we retrieved from Web of Science all the publications co-authored by at least one scientist from any West African or BRICS country. The databases searched were Science *Citation Index Expanded (SCI-EXPANDED)*, *Conference Proceedings Citation Index- Science (CPCI-S)*, *Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH)* and *Index Chemicus (IC)*. All languages and all document types were selected and timespan restricted to 2001–2010. *Analyse Results* function of Web of Science was used to analyse these data.

Results

Output per country (Table 1) reveals that, on its own, Nigeria produced more than half the total output of the region, far followed by Ghana (11.29%), Senegal (8.96%), Burkina Faso (6.29%) and Cote d'Ivoire (5.88%). The remaining 10 countries output each less than 5% of the region's total scientific papers. The West African publications are written in English mainly (95.5%). H-index equals 100.

The region produced mainly in *Medical and health sciences* (49%), followed by *Natural sciences* (32.81%), *Agricultural sciences* and *Engineering and technology* have approximately the same score (around 13%). Nearly half the West African total output has at least one foreign address. The region's collaborators count 151 countries among them 38 African, 27 American, 40 Asian, 41

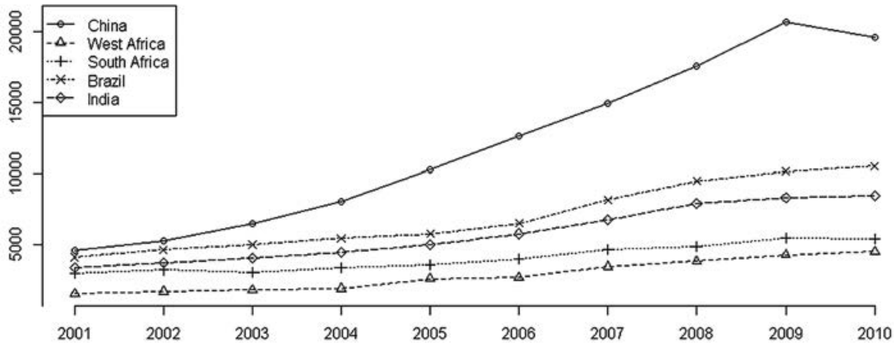
European and 5 Oceanian. France is ranked first (12.59%); just behind, comes the USA (12.54%), followed by the United Kingdom (10.09%). The fourth partner is Germany, far behind with 4.39%. South Africa is ranked fifth and the first African country with 3%.

Table 1. Scientific output of West African countries (2001–2010 – decreasing order)

Rank	Country	Publications share	
		Papers	Percentage
1	Nigeria	15,569	54.86
2	Ghana	3,203	11.29
3	Senegal	2,544	8.96
4	Burkina Faso	1,785	6.29
5	Cote d'Ivoire	1,669	5.88
6	Benin	1,335	4.70
7	Mali	1,204	4.24
8	The Gambia	986	3.47
9	Niger	586	2.06
10	Togo	433	1.53
11	Guinea	241	0.85
12	Guinea Bissau	225	0.79
13	Sierra Leone	117	0.41
14	Cape Verde	52	0.18
15	Liberia	49	0.17

The whole region performs less than Brazil, India, China and South Africa, the leader in science producing on the African continent. Indeed, over the same period of times, Brazil produced tenfold the West African volume of papers, India twelvefold, China about fiftyfold, and South Africa over twofold (Figure 1 and Table 2). Furthermore, the West African h-index is lower than that of each of the four countries; therefore, the quality of these countries' research measured by the h-index is much higher. Approximately, 43% of South African papers have at least one foreign address; India, China, and Brazil's is lower than 30; hence, West Africa depends more on international collaboration than the four countries. USA, Germany and UK are the common partners of West Africa and BRICS countries. France, the first West African partner didn't appear in the Chinese top-5 list.

Figure 1. Comparison of the West African scientific output to that of BRICS



Note: For the clearness and the readability of the figure, the BRICS' outputs were divided by 3.75 (Brazil), 6 (India), 10 (China) and 1.55 (South Africa).

Table 2. Comparing West Africa to China, India, Brazil and South Africa with selected indicators

	Output	Citable	Coll.	H	Partners	
WA	28,380	94.00	49.66	100	FR US UK DE ZA	12.59 12.54 10.09 4.39 2.99
BR	261,876	98.28	25.12	285	US FR DE UK ES	10.00 3.18 2.77 2.60 1.82
IN	346,992	97.67	17.56	281	US DE UK JP FR	6.06 2.41 1.76 1.66 1.40
CN	1,199,239	99.23	16.27	350	US JP DE UK CA	6.62 2.15 1.38 1.35 1.22
ZA	63,087	94.86	43.38	216	US UK DE AU FR	14.52 9.12 5.35 4.17 3.84

Note: We computed the West African h-index from the data we collected; BRICS' h-indexes are taken from SCImago (2007) and are related to the period 1996–2011. WA = West Africa – BR = Brazil – IN = India – CN = China – ZA = South Africa – FR = France – US = USA – UK = United Kingdom – DE = Germany – ES = Spain – JP = Japan – CA = Canada – AU = Australia

Discussion and conclusion

The main West African partners both for the whole region and for individual countrie are are France and United Kingdom, the former colonizes and USA, the big global science producer that has partnership with almost all countries in the World (Adams et al., 2010; Boshoff, 2009; Tijssen, 2007; Toivanen & Ponomariov, 2011). West Africa and BRICS countries shared 4 countries as top-5 partners; they are France, USA, UK and Germany. But West Africa depends much on them than BRICS (35% against 30% for South Africa and 10 to 15% for Brazil, India and China.

Table 3 gives some social, economic and research statistic of the region and BRICS countries. Even though China and India are ranked first and second respectively with the total output, they take third and fourth position respectively, behind Brazil (1st) and South Africa (2nd) while output is reported to the population size. West Africa counts the highest poor population share (60%). That is, the region is facing social and economic difficulties that make it could not prioritize research funding. Therefore, the investment in research and development is weak. Boshoff (2009) underlined that Sub-Saharan Africa's countries "are struggling to reach the target of allocating at least 1% of GDP to R & D" as they committed in the Lagos Action Plan (Organization of African Unity, 1980).

Table 3. Socio-economic and research data for West Africa and BRICS

	Output / inh.	HDI	GDP per capita	GERD	% poverty
WA	0.098	0.404*	1190.33*	–	60
BR	1.332	0.748	12,594	1.08	3.8
IN	0.279	0.568	1,489	–	41.6
CH	0.890	0.725	5,445	1.47	15.9
ZA	1.250	0.604	8,070	0.93	17.4

Note: *: average of the region computed by ourselves

References

Adams et al. (2010). Global research report: Africa. Thomson Reuters.

Boshoff, N. (2009). Neo-colonialism and research collaboration in Central Africa. *Scientometrics*, 81(2), 413–434.

Organization of African Unity. (1980). *Lagos plan of action for the economic development of Africa: 1980–2000*. Adis-Ababa: Organization of African Unity.

SCImago. (2007). SJR – SCImago Journal & Country Rank. Retrieved from http://www.scimagojr.com.

Tijssen, R. J. W. (2007). Africa's contribution to the worldwide research literature: new analytical perspectives, trends and performance indicators. *Scientometrics*, 71(2), 303–327.

Toivanen, H., & Ponomariov, B. (2011). African regional innovations systems: bibliometric analysis of research collaboration patterns 2005–2009. *Scientometrics*, 88(2), 471–493.

i In the alphabetic order, they are: Benin, Burkina Faso, Cote d'Ivoire, Cape Verde, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Nigeria, Niger, Senegal, Sierra Leone and Togo

Changing the budgeting model in times of scarcity. Opportunity or additional constraint?

Ana Ramos — ana.falcao@unl.pt / Research Support & Institutional Development Office, Universidade Nova de Lisboa, Campus de Campolide 1099-085 Lisboa (Portugal)

Maria Arménia Carrondo — mcarrondo@unl.ptInstituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras (Portugal)

Abstract

In Portugal, the decreasing state budget for HE has been distributed to universities on the basis of previous budgets and the university managers tend to accommodate the reductions by applying an across-the-board percentage cut. Therefore, budgeting is not aligned with the strategic goals of the institution, and do not take into account the recent performance of organizational subunits. In this work, we describe a case study conducted at a portuguese public university aimed at introducing performance measures in internal budgeting that could help to align funding with strategic objectives. A set of indicators to address teaching, employability, research, knowledge transfer and internationalisation was defined. The proposal was presented to the top managers of the university and its units, which recognised the advantages of progressing to a more rational budgeting model. The indicators and their relative weightings will be further analyzed and discussed; emerging results from this process will be presented at the Conference.

Background and Introduction

In contrast to the end of the nineties, when only a few countries used performance indicators in funding (Jongbloed & Vossensteyn, 2001), recent studies showed that a significant number of European countries have introduced measures of performance in Higher Education (HE) funding (Jongbloed, 2010; OECD, 2010). A percentage of the core funding is affected on the basis of performance and this money can, in general, be freely used by the universities. In some cases, the state funding model is mimicked in the internal budgeting, whereas other universities opt to use a different model of distribution. Performance-based systems differ considerably from country to country in both scope and the percentage of funds distributed, but common indicators include third-party funding, publications, number of students and degree completions in various combinations and weightings. However, the choice of indicators is always a controversial issue, since it is difficult to agree on indicators that adequately capture performance (Jongbloed, 2010).

In Portugal, state funding for HE supports two major expenses: staff and infrastructures. Between 2006 and 2008, this funding was distributed using a formula based on the number of students

and their fields of study, the graduation efficiency and the qualification of the teaching staff. In 2009, the two later indicators were not applied and, since then, the decreasing state budget has been distributed on the basis of previous allocations. University managers tend to accommodate the reductions by "reverse incrementalism" in the process of internal budgeting i.e., by applying an across-the-board percentage cut. This leads to a situation based on "historical budgets" that may not be aligned with the strategic goals of the institution, and thus do not take into account the recent performance of organizational subunits.

In addition to the institutional funding, HE institutions obtain national competitive funding for research (projects and individual grants) following periodical calls by the Portuguese research council, and international funding from other sources, such as the EU. Despite a general trend for an increased share of competitive project funding in the HE budget, the core funding can provide the universities with tools to shape their research agenda or align activities with their mission and strategy. To accomplish this goal, a shift towards rational mechanisms of internal budgeting is desirable: such models should include not only input but also output indicators addressing teaching, research, knowledge transfer and "service to society" (Lepori, Usher & Montauti, 2013).

The present work aims at addressing the following questions:

(1) What relevant indicators can be applied transversally to university Academic Units with very different dimensions and missions?
(2) What is the adequate percentage of the state budget to distribute on a performance basis?
(3) Which dimensions should be considered and what percentage of the money must be allocated to each dimension?
(4) What is the feasibility of introducing performance measurements in internal university budgeting in a country where the distribution of the state money to HEIs is not performance-based?

Methodology and preliminary results

This study was conceived to analyse the feasibility of changing the internal budgeting model in an adverse economic context. For this purpose, a case study was conducted at Universidade Nova de Lisboa, a public university with nine Academic Units with diversified sizes and missions. The proposal consisted in earmarking a small percentage of the state budget to be distributed to the Academic Units based on their relative performance measured by indicators (or proxies) intended to capture the following dimensions: Teaching, Graduate Employability, Research & Knowledge Transfer, Internationalisation.

The set of indicators (Table 1) was chosen because (*i*) they address key-issues for the University management that are not accounted for in the internal allocation of the state funding – which has until now considered only the number of students; (*ii*) they are included in the Strategic Plan; (*iii*)

they can be transversally applied to all Academic Units; (*iv*) the data for their calculation are available from official or independent sources, ensuring accuracy and reliability.

Table 1. Dimensions, indicators and data sources

Dimension	**Indicator**	**Source**
Teaching	Percentage of students that graduate in the number of years expected (N) or in N+1	Ministry Education & Science
Graduate Employability	Percentage of graduates that have a paid job 12 months after obtaining the degree	Observatory of Employability (OBIPNova)
Research & Knowledge Transfer	Revenue for research from international sources/total revenue for research	University Financial Report
	Competitive revenue for research/total revenue for research	University Financial Report
	Non-competitive revenue for research/total revenue for research	University Financial Report
	Normalized impact of Web of Science publications	Centre for Science and Technology Studies, CWTS
Internationalization	Foreign students/total students	Ministry Education & Science
	Foreign academic & research staff/total academic & research staff	Ministry Education & Science

The proposal was first discussed by the top management of the university followed by a presentation to the board of Directors of the Academic Units. The debate focussed on the appropriateness of the indicators rather than on their weightings or the percentage of the budget to allocate. The pertinence and relevance of progressing to a more balanced budgeting model was recognized by all actors, but the absence of stable and transparent rules for distributing the state budget to portuguese HEIs was a major concern. It was also argued that the internal model should be based on the same indicators used by the government once they become clear. Another point of discussion was the inadequate timing for introducing such changes, since in the last years the university had to cope with drastic budget cuts.

In conclusion, although this project could help to attenuate the long term effects of the crisis, the present economic situation limits its implementation and favours immediate action instead of strategic planning.

References

Jongbloed, B. & Vossensteyn, H. (2001). Keeping up performances: an international survey of performance-based funding in higher education. *Journal of Higher Education Policy and Management* 23: 127–145.

Jongbloed, B. (2010). Funding higher education: a view across Europe. Report for the Modern Project: European Platform Higher Education Modernisation, European Centre for Strategic Management of Universities (ESMU).

Lepori, B., Usher, J. & Montauti, M. (2013). Budgetary allocation and organizational characteristics of higher education institutions. A review of existing studies and a framework for future research. *Higher Education* 65: 59–78.

OECD (2010). Performance-based funding for public research in tertiary education institutions: country experiences. DST/STP/RIHR(2010)5.

Evaluation of Yahoo! Placefinder on institutional addresses[1]

Christine Rimmert christine.rimmert@uni-bielefeld.de / Bielefeld University, Interdisciplinary Bibliometric Group, Universitätsstr. 25, 33615 Bielefeld (Germany)

Introduction

Geocoding tools are used in bibliometrics to produce maps and calculate distances between researchers publishing together or distances between cited and citing authors (e.g.: Bornmann & Ozimek, 2012).

Despite the increasing use of these tools, until now there is little knowledge about the quality of their results in the context of institutional affiliation data in bibliometric databases (which are a special case of addresses). A sample of address data has been put into Yahoo! Placefinder (YP) and results have been evaluated according to completeness and correctness at the city level. Furthermore it has been investigated how different kinds of input data can affect the results.

Data

For testing the quality of YP results, institutional address records from papers in the same set of seven journals as used by Bornmann & Ozimek have been studied:

- Journal of the American Society for Information Science and Technology
- Scientometrics
- Information Processing & Management
- Journal of Information Science
- Journal of Documentation
- Information Research
- Journal of Informetrics

From these journals a total of 9,116 distinct address records of 7,085 papers with document type 'article' published between 1989 and 2009 were selected. Address records have been extracted from Web of Science.

1 Certain data included herein is derived from the "Science Citation Index Expanded (SCIE) – Tagged Data" prepared by Thomson Reuters (Scientific) Inc. (TR®), Philadelphia, Pennsylvania, USA:

Methods

The address data were processed via an YP Application Programming Interface (API). YP results include e.g. country, city, latitude, longitude. Different variants of input have been evaluated: input of the full address (as used by Bornmann & Ozimek) vs. input of parts of addresses in structured and unstructured form. Input language parameter has been set to English (en_EN).

Input variants

YP allows input of address data either as a complete string (CS) (location = '…') or as structured elements (SE) (country = '…', city = '…', street = '…'). Both versions have been checked in different ways.

On the one hand address data coming as complete (comma separated) strings from the online version of Web of Science have been used (e.g.: 'Dalhousie Univ, Fac Comp Sci, Halifax, NS B3J 2X4, Canada'). There is no secure solution to extract specific elements like street and city from those strings.

On the other hand, Web of Science raw data files have been used to extract the full address string as well as structured address information with the following data fields:

Table 1. Web of Science address data fields

Code	Content
NF	Full address string
NC	Organization
ND1	Sub-organization 1
ND2	Sub-organization 2
ND3	Sub-organization 3
NN	Street
NY	City
NP	Province/State
NU	Country

Fields may be empty or mixed up (e.g. street name may appear in ND2). ND fields are missing in most cases of publication year 1995 and earlier. Field 'NP' is missing in many cases, so it is ignored here. The important parts remaining for geocoding are 'NN', 'NY' and 'NU'.

Four variants of input into YP have been tested:

(1) CS-NF: NF as *location*
(2) CS-NN/NY/NU: comma-separated NN, NY, NU as *location*
(3) SE-NF: the last (comma-separated) part of NF as *country*, the second last part as *city*, empty string (“) as *street*
(4) SE-NN/NY/NU: NN as *street*, NY as *city*, NU as *country*

The first two variants are processed as complete string (CS) input to YP, the latter two as structured elements (SE). The first and third variant can be extracted from the online version of Web of Science, whereas the other two require access to raw data files.

Results

Number of results (completeness)

Though the fields NN, NY and NU in several cases are incorrectly filled or even missing, YP performs better on these parts of the address than on the full address string, both in ‘no result at all’ cases and in ‘no city result’ cases:

Table 2. Completeness of YP results

YP result	CS-		SE-	
	NF	NF/NY/NU	NF	NF/NY/NU
No result	254	53	118	16
No country	14	14	0	17
No city	269	244	1027	169

To determine the number of found results as a percentage of the number of inputs (9,116), for every variant the sum of the cases ‘no result’ and ‘no city’ has been subtracted from the total (the cases ‘no country’ are contained in the cases ‘no city’). Completeness ratios were calculated for the different versions:

(1) CS-NF: 94%
(2) CS-NN/NY/NU: 97%
(3) SE-NF: 87%
(4) SE-NN/NY/NU: 98%

The structured NN/NY/NU version is clearly better than the unstructured version of NN/NY/NU and even the best at all. Overall the results for NN/NY/NU are better than the results of the NF input. This may be due to YP-overinterpretation of parts of the NF string that do not belong to the postal address.

Correctness of results

The table below shows the differences among the input variants concerning correctness of city results among the results with city level:

Table 3. Correctness of YP results (city level)

Input variant		Correct results	Errors
CS-	NF	8117	476
	NN/NY/NU	8588	231
SE-	NF	7718	253
	NN/NY/NU	8807	124

The structured input SE-NN/NY/NU seems to be also the best variant considering correctness. Ratio of correct results on city level has been calculated as a percentage of the total number of addresses:

(1) CS-NF: 89%
(2) CS-NN/NY/NU: 94%
(3) SE-NF: 85%
(4) SE-NN/NY/NU: 97%

Conclusion

Input format of bibliometric address data makes a difference on the quality of YP results. On city level, the best input variant achieves 97% correct results, while the worst one rates only at 85%. Concerning completeness as well as correctness, the best results can be achieved by the SE-NN/NY/NU variant. However, structured input is not automatically more successful than unstructured versions. For someone with no access to Web of Science raw data it is a better choice to use the full address string as YP input, instead of extracting the last two parts for giving it as structured input.

References

Bornmann, L. & Ozimek, A. (2012). Stata commands for importing bibliometric data and processing author address information. *Journal of Informetrics*, 6, 505–512.

Yahoo! Boss PlaceFinder. [http://developer.yahoo.com/boss/geo/]

An approximation to the measurement of researcher's scientific and technologic careers, using secondary data collections[1]

Sandra Carolina Rivera-Torres crivera@ocyt.org.co / Human Resources Group, OCyT, Carrera 15 Nº 37-59, Bogotá D.C. (Colombia)

César Pallares cpallares@ocyt.org.co / Human Resources Group, OCyT, Carrera 15 Nº 37-59, Bogotá D.C. (Colombia)

Marcela Galvis-Restrepo mgalvis@ocyt.org.co / Human Resources Group, OCyT, Carrera 15 Nº 37-59, Bogotá D.C. (Colombia)

Abstract

In 2011, Colombia submitted a membership request to the Organization of Economic Cooperation and Development – OECD. Joining this club of countries requires, among other things, accurate data and statistics on the performance of the S&T system. In order to support this process, the Colombian Observatory of Science and Technology – OCyT – works in the production of better data by revising and validating new methodologies for S&T measurement like the "Buenos Aires Manual". This initiative has been carried out by the Ibero-American Network of Science and Technology Indicators – RICyT – and is intended to develop a new framework to measure scientist's trajectories. In this Work In Progress report, we comment some of the results already obtained in the process of validation of the Manual and some of the strategies implemented to calculate the proposed indicators. Using secondary data sources, we find that the use of CV datasets can be complemented with other data sources and that age and gender are important determinants of the career path followed by Colombian researchers, since measurements show a change in their production behaviour as they grow old and different patterns according to the sex. Finally the researchers dedicate an important share of their time to R&D activities; this is an important result given the multiple activities they must perform as part of their work in universities.

1 This work was supported by the Colombian Observatory of Science and Technology (OCyT). The authors are thankful to Colciencias by the access granted to ScienTI data.

Cumulative Capabilities in Colombian Universities: an Evaluation Using Scientific Production

Sandra Carolina Rivera-Torres crivera@ocyt.org.co / Human Resources Indicators of Sciences and Technology, Observatorio Colombiano de Ciencia y Tecnología, Carrera 15 Nº 37–59, Bogotá (Colombia) / National University of Colombia

Marcela Galvis-Restrepo mgalvis@ocyt.org.co / Human Resources Indicators of Sciences and Technology, Observatorio Colombiano de Ciencia y Tecnología, Carrera 15 Nº 37–59, Bogotá (Colombia)

Jenny Cárdenas-Osorio jcardenas@ocyt.org.co / Human Resources Indicators of Sciences and Technology, Observatorio Colombiano de Ciencia y Tecnología, Carrera 15 Nº 37–59, Bogotá (Colombia)

Introduction

Research was institutionalized in Colombian universities around fifty years ago with the creation of government institutions to stimulate this activity, and the support of loans from the Inter-American Development Bank – IDB – that were used to start building scientific and technological capabilities (Bucheli et al., 2012; CINDA, 2012; OECD & The World Bank, 2013). The institutional development of the science, technology and innovation system, gives priority to research groups as organizational units for research; a structure largely promoted by Colombian Administrative Department of Science, Technology and Innovation – Colciencias –, the government body in charge of measuring research group's scientific capabilities and assigning resources according to their performance (Villaveces & Forero, 2007; OCyT, 2012).

Research groups as organizational units were adopted by universities and research institutions as well. Some universities have even developed their own financing schemes intended to stimulate their growth and sustainability (Colciencias, 2008; OCyT, 2012). In 2011–2012, the Colombian Observatory of Science and Technology – OCyT – developed a methodology to evaluate the research groups of one of Colombia's leading universities, which can be used for tracking and impact evaluation of R&D policy.

As a result, in this work we focus on scientific capabilities of research groups measured by their researchers' scientific production, as defined by Colciencias' model. In the previous evaluation (OCyT, 2012), we found evidence that the production is highly concentrated in a few researchers following Zipf's power law in an eight year time frame (Sutter & Kochner, 2001). This approach allows for a representation researcher's capabilities, taking into account the accumulated knowledge in the publication trajectory of the population.

Methodology

We looked at publications registered in ScienTi database associated to researchers working in research groups in the incumbent university; afterwards, we applied concurrency algorithms to match those papers, with publications in ISI- Thomson Reuters Web of Science database.

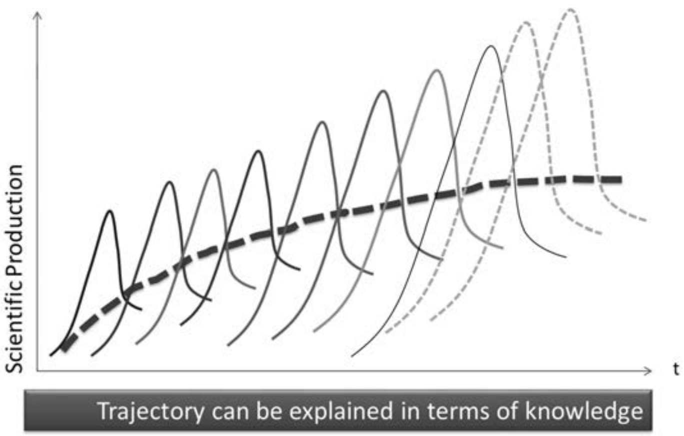
Additionally, since the knowledge codified in publications is an evolutionary variable that can be accumulates over time; in this analysis we use cumulative distributions in order to determine the future trend of scientific publications as means to understand the individual's capabilities (Bozeman & Corley, 2004; Lepori & Barré, 2008).

According to Abe & Rajagopal (2001) the state of the scientific production of a group of researchers fits the Cauchy-Lorentz distribution. This means that the difference in productivity p, between the members (n) of the research group can be modelled with a Lorentz distribution, given by the equation:

$$p(n)=\frac{1}{\pi}\frac{\gamma}{(n-n_c)^2+\gamma^2}, \qquad (1)$$

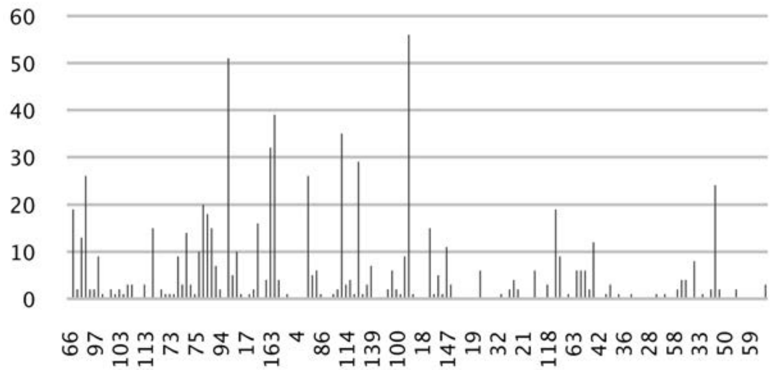
where nc is the arithmetic half of the population and γ is the half-width. As shown in Figure 1, the distribution is useful to evaluate the behaviour of the production of researchers each year and by accumulating it; we can observe the degree of concentration of the knowledge generated.

Figure 1. Capabilities Measurement Using Scientific Production



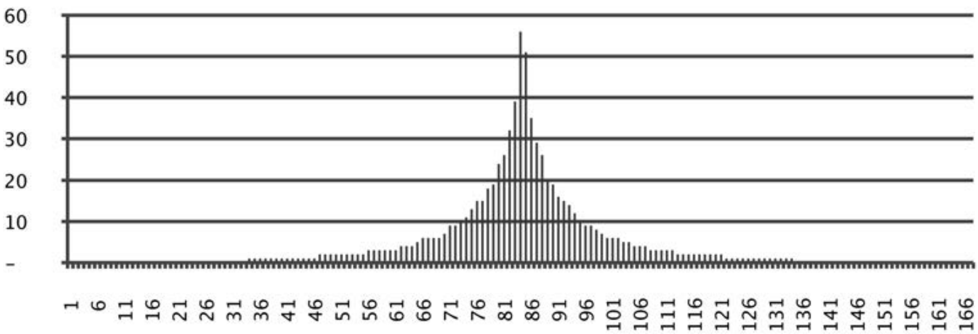
First, we accumulate the production of each researcher in a given scientific field in 2002–2008. In Figure 2 we show an example for a population of 167 researchers located in the horizontal axis in the field of medical sciences of the Universidad de Antioquia (OCyT, 2012); production is represented in the vertical axis.

Figure 2. Accumulated Scientific Production in 2000–2008, Researchers in Medical and Health Sciences



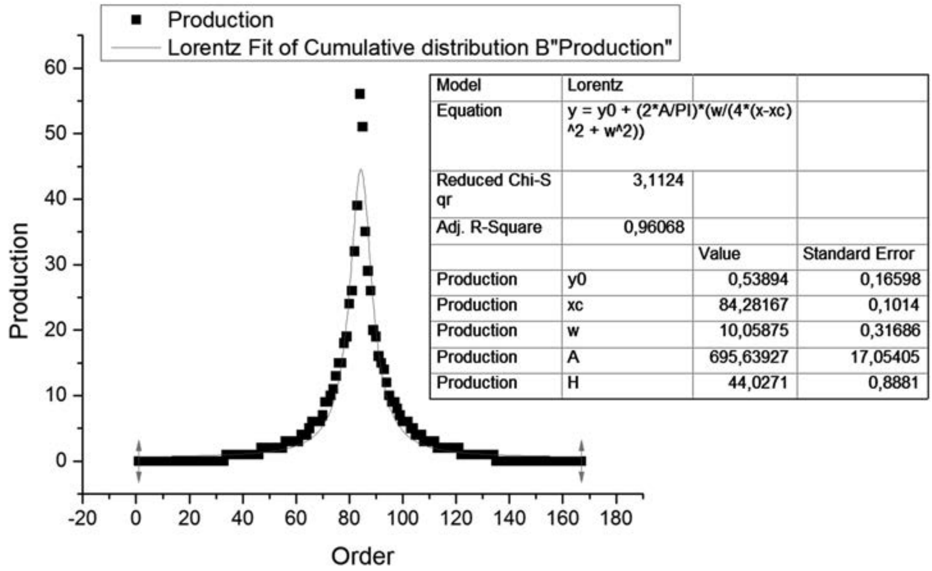
Then we organize the results locating the most productive researcher in the centre; the second and third most productive are located to the right and left and so on. Figure 3 shows the data in Figure 2 organized in a centralized way. These data are shown as an example for a small population of researchers. In the case of Colombian universities, this tool is useful to describe the trajectory of the research groups.

Figure 3. Centralized Cumulated Scientific Production in 2000–2008, Researchers in Medical and Health Sciences



As shown in Figure 4, production is highly concentrated on a few researchers who have the highest share of the scientific production. In this graph we fitted a Lorentz distribution to the centralized data.

Figure 4. Lorentz Fitting of the Centralized Distribution



Using Lorentz distributions, the effect of a researcher in the population can be defined as its percentage contribution given by the relationship:

$$E(n) = \frac{p(n)}{P_T}, \tag{2}$$

Usually the centre value is near N/2, so it would be enough to find the number γ (the half width) to determine the distribution of the production in the population. A property of the Lorentz distribution is:

$$P_T = 2\int_{n_c-\gamma}^{n_c+\gamma} p(n)dn, \tag{3}$$

So the range [n_c − γ, n_c + γ] (with size 2γ) contains half of the population. With this result we can define the percentage of the population concentrating 50% of the production as:

$$N_m = \frac{2\gamma}{N}, \tag{4}$$

This result can be used as a measure for the concentration of production in the population of researchers, and eventually for research groups' comparisons.

Conclusions

The methodology presented here is useful for evaluating variables reflecting the knowledge accumulation observed in R&D activities. An additional application consists in mapping the trajectory of different groups like individuals, research groups or institutions; additionally it is useful to observe the degree of development reached by each individual in the S&T system and use it to target specific populations of highly achieving individuals. Specifically in the Colombian context, the methodology helps to measure capacity while it is understood that individuals have an evolutionary trajectory, and they shape a growing and heterogeneous community.

References

Abe, S., & Rajagopal, A. (2001). Information theoretic approach to statistical properties of multivariate Cauchy-Lorentz distributions. *Journal of Physics A: Mathematical and General*, 34(42), 8727.

Bozeman, B., & Corley, E. (2004). Scientists' collaboration strategies: implications for scientific and technical human capital. *Research Policy*, 33(4), 599–616.

Bucheli, V., Díaz, A., Calderón, J. P., Lemoine, P., Valdivia, J. A., Villaveces, J. L., et al. (2012). Growth of scientific production in Colombian universities: an intellectual capital-based approach. *Scientometrics*, 91(2), 369–382.

CINDA. (2010). Educación superior en Iberoamérica Informe 2010.

Lepori, B., Barré, R., & Filliatreau, G. (2008). New perspectives and challenges for the design and production of S\&T indicators. *Research Evaluation*, 17(1), 33–44.

OCyT. (2012). Indicadores de Ciencia y Tecnología Colombia 2012.

OECD, & Bank, T. W. (2012). *Reviews of National Policies for Education: Tertiary Education in Colombia* 2012. OECD Publishing.

Sutter, M., & Kochner, M. G. (2001). Power laws of research output. Evidence for journals of economics. *Scientometrics*, 51(2), 405–414.

15 years Ranking and Indicator Experience: Needs and Requirements

Isabel Roessler isabel.roessler@che.de / CHE Centre for Higher Education, Verler Strasse 6, D-33332 Gütersloh (Germany)

Three studies – one result

When in 1970 the Carnegie Foundation started its classification of colleges and universities to support its program of research and policy analysis the starting point of comparative analyses of higher education institutions begun. Since then several analyses in the sector of higher education have been designed and implemented. Most of them are related to research and therefore collect indicators related to research. The probably most popular known global rankings are the Academic Ranking of World Universities (Shanghai Ranking), and the rankings compiled by Times Higher Education (THE) and by QS. Recently a consortium of European organisations with CHE and CHEPS as the lead partners started a project funded by the European Commission to implement a multi-dimensional, user-driven international ranking:

U-Multirank

U-Multirank pursues major approaches which have not been in the focus of the existing global rankings and collects data for indicators which have not been analysed systematically before, e.g. data about teaching and learning, knowledge transfer, and information about regional engagement. Especially three approaches have to be stressed out:

- Integration of different sources: disposable data and information, self reported data by the departments and institutions and information given directly by students
- No composite indicators: Each indicator is published
- No league table but ranking groups.

The background of the U-Multirank project encompassed the design and testing of a new transparency tool for higher education and research. Institutions are multi-purposes organisations and different institutions focus on different blends of purposes and associated activities. Such an enhanced understanding of the diversity in the profiles and performances of higher education and research institutions at a national, european and global level requires a new ranking tool. Existing international transparency instruments do not reflect this diversity adequately. The new tool will regard the diversity of institutional profiles

With this strategy the U-Multirank project is constructed in its ideas very similar to the national German CHE Ranking: multidimensionality, integration of different sources, field-based infor-

mation and groups instead of league tables are implemented and are used since 1998, when the first German CHE Ranking was published.

Based on the experiences with both CHE Ranking and U-Multirank our paper, published in August 2013, will bring together results from three studies:

- Survey among users from the national CHE Ranking
- Institutional user survey, from CHE Ranking
- Survey among the participants of U-Multirank (available August 2013).

Needs of the target group

A survey among 2157 individual users of the online CHE Ranking gives first answers on the question which the specific needs of the target groups are and why they use a transparency tool like the ranking.

The CHE ranking is used mainly by prospective students and students (86%). Hence it is proved, that the targeted group of users of the CHE ranking is reached.

Main reasons for the use of this ranking are:

- Information about the performance of the own or chosen university compared to other universities (63%)
- Wish to get informed about the universities which are offering the subject the user are interested in (56%)
- Information about the location of the university (53%)

The most important indicators for the users are:

- Student judgments (84%)
- Information about teaching and learning (e.g. student staff ratio and the like) (83%)
- Information about the equipment of the department (77%)
- Information about practical elements and job related elements of the studies (76%)

The vast majority (75%) indicates that the CHE ranking publishes useful information. 17% of the users would give the best mark (out of six) to the ranking, 65% the second best mark and 14% the third best mark (in total: 96%).

Monitoring the pros and cons of using ranking results

Second, an institutional user-survey was able to monitor the pros and cons a university has got by using the ranking results for its own purposes. A selection of about 170 higher education institutes which are participating in the CHE Ranking was surveyed.

Ranking results are used for internal analyses. It could be realised, that:

- 53 % of them use the CHE Ranking for internal analysis,
- 48% are using the German DFG Förderatlas
- 11% use the Times Higher Education Ranking
- 7% the ARWU (Shanghai Ranking)

In most cases the CHE ranking results are used in:

- Public relation units (58%)
- Presidium (51%)
- Within the departments (48%)
- Marketing (44%).

Only a very small number (19%) uses the CHE ranking for controlling.

Based on this national experience it can be resumed that rankings should offer indicators which are geared to the needs of the (different) target groups and that the results of rankings can be helpful in the daily work of the universities by offering them additional data.

Reasons for participating in a new global Ranking: U-Multirank

By now (mid of June 2013) more than 600 HEI out of more than 60 countries applied for participating in the U-Multirank Ranking. Due to this the goal of 500 HEI by the end of June is overfulfiled.

The implementation process of U-Multirank now offers the chance to check the needs of the universities in a broader community. In July 2013 the participating universities will be asked:

- Why they wanted to participate in U-Multirank
- About the purposes of U-Multirank
- What benefits they (want) will have in their opinion
- About the approaches of U-Multirank

Conclusion

Rankings are used by different target groups and with different intentions. The value is obvious:

- Rankings can be helpful in finding the prospective HEI, in public relation or marketing.
- Especially the judgements of students are relevant for the target group of prospective students.
- Even the implementation of a new ranking does not overexcite the HEI: They are still interested in Rankings and are ready to participate in these information tools.

The influence of performance-based university research funding systems on co-authorship practices

Gunnar Sivertsen gunnar.sivertsen@nifu.no / Nordic Institute for Studies in Innovation, Research and Education (NIFU), Wergelandsveien 7, NO-0167 Oslo (Norway)

Nick Deschacht nick.deschacht@hubrussel.be / Faculty of Economics and Business, Hogeschool-Universiteit Brussel (HUB), Warmoesberg 26, B-1000 Brussel (Belgium)

Henrik Aldberg henrik.aldberg@vr.se / The Swedish Research Council, Box 1035, SE-101 38 Stockholm (Sweden)

Tim C. E. Engels tim.engels@ua.ac.be / Centre for R&D Monitoring (ECOOM), University of Antwerp, Middelheimlaan 1, B-2020 Antwerp (Belgium) / Antwerp Maritime Academy, Noordkasteel Oost 6, B-2030 Antwerp (Belgium)

Introduction

We study the effect of the introduction of performance-based university research funding systems (PRFSs) on co-authorship practices in Flanders and Norway. One of the most rapid changes in scientific publishing patterns during the last decades is the increasing number of authors and addresses per publication (King, 2012). The increased occurrence of multiple authorship is a general phenomenon involving almost all fields of research. The driving forces of this evolution are changes in scientific standards and norms, as well as factors internal and external to the organization and workflows of research. These include the increasing specialization in science, the advantage of sharing resources, the reduction of the costs related to communication over (long) distances, the wish to improve access to funds and to increase productivity.

This poster presents a preliminary analysis of the effect of one specific external driver of changing co-authorship practices, namely the influence of performance-based university research funding systems (Hicks, 2012). In Flanders (Belgium), the regional PRFS involves whole counting of publications in the sense that each university that is involved in a paper receives research funding proportional to the full weight of that paper (Debackere & Glänzel, 2004). In Norway, the national PRFS involves fractionalization of publications per university. Hence each university receives only the weight that corresponds to the author fraction that it has provided (Sivertsen, 2010). The Norwegian model may be perceived as not encouraging co-authorship, while the Flemish model seems to stimulate co-authorship in all cases except intra-university collaboration. These respective policies were introduced in 2003 (Flanders) and 2005 (Norway), and have remained unchanged with regard to fractional and whole counting since. Hence their effects on the evolution of co-authorship can now be studied comparatively. Our research question for the present study is: Has the introduction of PRFSs that use whole counting and fractionalization of publications resulted in different developments in terms of co-authorship practices in Flanders and Norway?

Our null hypothesis is that the PRFSs have not influenced the on-going evolution of co-authorship practices meaningfully. In other words, we expect the increase in co-authorship to continue at the same rate in both Flanders and Norway. We expect this because changes in disciplinary standards and norms are the main drivers of the evolution of co-authorship practices.

Data

The data included in this study are the articles, letters, notes, and reviews published between 1982 and 2011 and indexed in one of the journal publication databases (Science Citation Index Expanded, Social Science Citation Index, Arts & Humanities Citation Index) of the Web of Science with an address in Norway (n = 172.740) and/or from a Flemish university (n = 159.990).

Regression Model

In order to evaluate the effects of the Flemish and Norwegian PRFSs on co-authorship, we estimate an ordinary least square (OLS) piecewise regression model of the number of authors Y_i of an article i. The model allows for a structural break in the linear trend before and after the introduction of a policy at time t^*:

$$\text{In}(Y_i) = \alpha + \beta_1 t_i + \beta_2 X_i + \beta_3 F_i + \beta_4 t_i F_i + \beta_5 X_i F_i + \varepsilon_i$$

where t_i is the year in which the article was published, F_i is a dummy variable for country (1 for Flanders and 0 for Norway) and

$$X_i \equiv \begin{cases} 0 & \text{for } t_i < t^* \\ t_i - t^* & \text{for } t_i \geq t^* \end{cases}$$

We use the natural logarithm as the dependent variable because the number of authors is right-skewed, which would lead the OLS estimates to overemphasize extreme cases. In the analysis, we set $t^* = 2006$.

Results

We observe very similar evolutions in terms of co-authorship patterns in Flanders and Norway. Hence the introduction of the PRFSs using whole/fractional counting can be seen as a natural experiment (the effect of which we test).

Table 1. Regression coefficient estimates, standard errors and level of significance.

	Estimated coefficients
Constant	$\alpha = -49.9540$ (.6023)***
Year	$\beta1 = 0.0256$ (.0003)***
Year since 2006	$\beta2 = -.0056$ (.0015)***
Constant x Flanders	$\beta3 = 11.6022$ (0.8972)***
Year x Flanders	$\beta4 = -.0057$ (.0004)***
Year since 2006 x Flanders	$\beta5 = .0082$ (.0022)***
R^2	.08
N	332,728

***$p < .001$

Robust standard errors, which avoid the assumption of homoscedasticity in the analysis, are quasi identical to the standard errors above.

In Norway the number of authors increased by 2.56 percent on average every year until 2006. After that the Norwegian average growth rate slowed by .56 percentage points (to a rate of 2.00 percent per year). In Flanders the average growth rate prior to 2006 stood at 1.99 percent or .57 percent below that of Norway. However, since 2006 the Flemish rate has been .82 percentage points above the Norwegian rate. This implies an acceleration in Flanders of .26 percentage points (-.56 + .82) since 2006. All these results are significant at a .001 level. The effect sizes are considerable. The difference in the effects of the Flemish and the Norwegian PRFSs (.82 percentage points) represents about one third of the overall growth rates in both countries. From this general result it seems that the Norwegian PRFS has had a dampening effect on co-authorship, whereas the Flemish PRFS has stimulated and thus resulted in a more rapid increase of co-authorship as of 2006. Hence we reject our null hypothesis. However, even though we used the natural logarithm of the number of authors as dependent variable, the overall result may have been influenced by the phenomenon of hyperauthorship (Cronin, 2001). As local PRFSs cannot influence the byline of hyperauthorship papers further analysis in which these papers are singled out is needed.

References

Cronin, B. (2001). Hyperauthroship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology*, 52, 558–569.

Debackere, K. & Glänzel, W. (2004). Using a bibliometric approach to support research policy making: The case of the Flemish BOF-key. *Scientometrics*, 59, 253–276.

Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, 41, 251–261.

King, C. (2012). Multiauthor papers: Onward and upward. *ScienceWatch Newsletter*, July 2012.

Sivertsen, G. (2010). A performance indicator based on complete data for the scientific publication output at research institutions. *ISSI Newsletter*, 6, 22–28.

Integrating Science, Technology and Innovation (STI) Indicators in Policy-making: The Nigerian Experience

W. O. Siyanbola williesiyanbola@yahoo.com / National Center for Technology Management, Obafemi Awolowo University, Ile-Ife (Nigeria)

A. D. Adeyeye adey_ad@yahoo.com / National Center for Technology Management, Obafemi Awolowo University, Ile-Ife (Nigeria)

O. R. Olaopa[i] olawale.olaopa.@gmail.com / National Center for Technology Management, Obafemi Awolowo University, Ile-Ife (Nigeria)

O. M. Hassan reachaoh@yahoo.co.uk / National Center for Technology Management, Obafemi Awolowo University, Ile-Ife (Nigeria)

O. B. Oluwatope reachaoh@yahoo.co.uk / National Center for Technology Management, Obafemi Awolowo University, Ile-Ife (Nigeria)

Introduction

That African is one the poorest region in the World has been widely documented (World Bank, 1989; Bigsten and Durevall, 2008, Ogbuji, 2009, Olaopa and Uzodike, 2009, Assefa, 2010). Scholars as well as development partners in and outside the continent have adduced plethora of reasons concerning the continent's precarious situation. In spite of various policies and priority agenda of development the continent problems remained pervasive. There is need to integrate research outputs into policy making process.

Objectives of the paper

The objective of the paper is majorly to share the Nigerian experience in formulating STI policy using scientific indicators.

Methodology

Qualitative method of data analysis was employed. System framework of analysis was also used to bring into perspective the fundamental role played by STI indicators in providing scientific evidences for designing, formulating and implementing national innovation strategies.

Discussion

Development of STI Indicators Project: The Nigerian Experience

In Nigeria, the importance of S&T in driving and achieving rapid industrialization and socio-economic development has been realised. This has motivated the country to intensify efforts at accessing quality and adequate information for effective STI policy making. This is particularly with reference to information in the area of STI indicators. However, the dearth of an effective mechanism for gathering reliable data useful in STI policy process has been an obstacle to Nigeria's technological advancement. It is in realization of this that a project, which was first of its kind in Nigeria, was conceived in 2005 to 'develop STI indicators for Nigeria as a basis for planning and monitoring national (S&T) activities' (NACETEM, 2011).

STI Policy-making Process in Nigeria: The History and major inadequacy

One of the major reasons why the record of STI policy formulation in Nigeria has been depressing and demotivating as concretely reflected by the frequent occurrences of its failure to resolve some of the country's development challenges is that they are not research-informed. As a result, there is need to generally overhaul, empirically identified and attended to most of the areas where the performance of the government is abysmally low and which impede scientific and technological progress. The identification of these was influenced by the development and integration of STI indicators project undertaken by NACETEM on behalf of the Nigerian government.

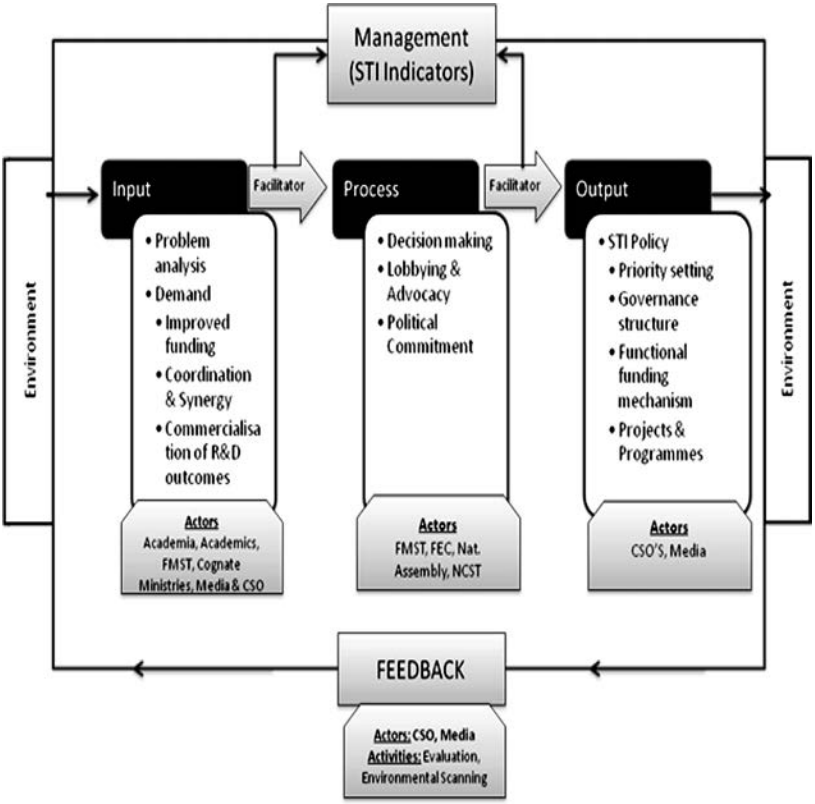
Strategic Location of NACETEM within Nigeria's Innovation System

It is the government agency with the primary mandate of undertaking science, technology and innovation policy research in Nigeria. It is the policy think tank of the FMST, providing unique knowledge support through its mandate of policy research in STI management. In furtherance of this, NACETEM has completed about 20 policy projects since 2005 in different areas of STI. Prominent among these is the development of STI indicators, the Nigerian component of NEPAD ASTII on STI policy review process leveraged on.

Framework for Policy Formulation using STI Indicators

To properly situate the degree of the nexus between effective utilisation of STI indicators in the context of STI policy formulation and development, the theoretical framework for this study is made up of the system theory.

Figure 1. Systems Model Showing Interrelationship between STI Indicators and STI Policy Formulation Process



Source: Adapted from Olaopa, 2011

Integrating STI Indicators in Policy-making: The how; the Lessons

The formulation of a new Nigeria's STI policy provided an opportunity to develop an evidence-based policy with participation from key stakeholders and leveraged on the outcome of the STI indicators survey with the following benefits:

(1) From Science and Technology to Innovation
(2) Shift in Focus: From knowledge Generation to Wealth Creation
(3) Improved Funding
    (a) Creation of Special Funding Agency
    (b) Encouraging Business Sector Funding
(4) Higher Capability for STI

Conclusion

Knowledge is a strategic resource to achieve the business and governance objectives. Its utilisation/application and commercialisation, is the ultimate and the primary driver of organisations' value (Grant, 1996). The exploitation of these requires not only proper monitoring and evaluation but intelligent development, adoption, promotion and integration of some common sets of indicators to benchmark national systems of innovation and developments in science, technology and innovation and policy process.

References

Assefa, T. (2010). Enabling Knowledge Sharing to Promote Innovative Organizations in Africa. A Paper to be Presented at Expert Group Meeting on Harnessing Knowledge to Achieve MDGs, November 4–5, 2010, Addis Ababa, Ethiopia.

Bigsten, A. and Durevall, D. (2008). The African economy and its role in the world economy. Uppsala: Elanders Sverige AB. Retrieved on June 11, 2010 from http://www.oecd.org/dataoecd/23/16/25497010.pdf.

Grant, R.P. (1996). Toward Knowledge based theory of the firm, *Strategic Management Journal*, 17, Winter, Special Issue, 109–21.

NACETEM (2011). Report of 2006/07 R&D Survey (1st edition).

Ogbuji, Kingsley. (2009). Gaddafi and the United States of Africa. *Triumph Newspaper*. Feb. 2.

Olaopa, O.R. and Uzodike, U. (2009). Science, Technology and Innovation (STI) Promotion: An Imperative for Youth Empowerment for Africa's Development. *An Unpublished paper presented at the Joint AISA/SAAPS Symposium on Defining the Contours of Knowledge Production in the 21 Century organised by the Africa Institute of South Africa (AISA) in association with the South African Association of Political Science* (SAAPS) on 15 and 16 October, 2009 at Burgers Park Hotel, Tshwane, South Africa.

World Bank. (1989). *Sub-Saharan Africa: From Crisis to Sustainable Growth–A Long-Term Perspective Study*. Washington, D.C.: The World Bank cited in Mabogunje, Akin L. (2004)."Framing the Fundamental Issues of Sustainable Development in Sub-Saharan Africa." *CID Working Paper No. 104*. Cambridge, MA: Sustainable Development Program, Center for International Development, Harvard University; also published as TWAS Working Paper 1. Trieste, Italy: Third World Academy of Sciences.

---

i O.R. Olaopa, is a doctoral candidate in Public Policy, School of Politics, University of KwazuluNatal, Howard college campus, South Africa, and corresponding and presenting author.

Framing R&D markets: A study of private and public demand for R&D-services[1]

Espen Solberg espen.solberg@nifu.no / Division of Statistics and Indicators, Nordic Institute for Studies in Innovation, Research and Education (NIFU), Wergelandsveien 7, N-0167 Oslo (Norway)

Ole Wiig ole.wiig@nifu.no / Division of Statistics and Indicators, Nordic Institute for Studies in Innovation, Research and Education (NIFU), Wergelandsveien 7, N-0167 Oslo (Norway)

Abstract

Transfers and transactions of R&D are difficult to capture through conventional R&D-indicators. This paper presents work to investigate how such transfers create an R&D market. Although comparable data are scarce, we find significant country differences in the size and composition of national markets for R&D services. Drawing on these findings, we classify national R&D funding systems according to a dichotomy of introvert vs. extrovert funding patterns.

Introduction

Public R&D-funding seems to experience a general shift from direct institutional grants to more project and competition based funding (OECD 2010, Lepori et. al. 2007). At the same time, companies increasingly rely on external knowledge sources in their R&D and innovation activities (Chesbrough 2006). Together, these two trends contribute to a growing market for R&D services, where a variety of actors act as R&D suppliers and compete for public funding and R&D contracts. Yet, the size and functioning of these markets remain poorly measured and understood.

1 This work draws on the report *Markets for Applied Research* (2012), commissioned and financed by the Norwegian Ministry of Education and Research.

Table 1. Stylised typology of R&D-systems

Degree of public project funding →	**Public demand-oriented** (High degree of open public funding but low propensity to use external suppliers in industry)	**Extrovert system** (High degree of open public funding and high propensity to use external suppliers in industry)
	Introvert system (High degree of institution based public funding and low propensity to use external suppliers in industry)	**Private demand-oriented** (High degree of institution based public funding, but high propensity to use external suppliers in industry)
	Firms' use of external R&D-sources →	

The stylised model in table 1 suggests a framework for describing country differences in the degree of openness and market exposure in national R&D funding Introvert systems are here defined as systems with a high degree of institutional public funding and a high share of in-house business enterprise R&D. In contrast, extrovert systems are systems with a high share of public funding through open mechanisms and a high share of enterprises purchasing R&D from external suppliers. Using experimental indicators, we seek to classify countries according to this general model.

The public market for R&D-services

As part of an OECD project on improving cross country comparisons of public R&D funding, Van Steen has proposed experimental indicators which distinguish direct institutional funding from government project funding (Steen, 2012). The latter comprises both project funding through research councils and direct contract research. Available indicators reveal significant country differences in the degree of project funding, ranging from above 70 per cent (New Zealand) to below 25 per cent (Switzerland) (table 2). These findings are already presented as an experimental indicator in the OECD STI Scoreboard (OECD 2011).

The use of external R&D in industry

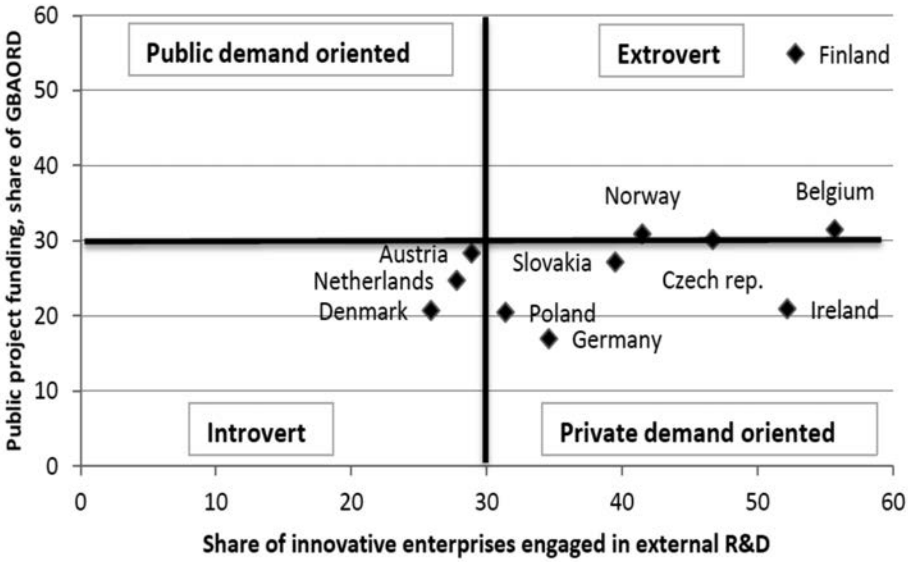
As stated in the Frascati manual, countries are recommended to distinguish between R&D performed within (intramural) and outside the enterprise or statistical unit (extramural) (OECD 2002). Furthermore, it is recommended to distinguish between extramural R&D purchased from entities within and outside the same enterprise group. In principle, these distinctions allow for comparisons of both scale and type of firms' procurement of R&D from external sources.

Unfortunately, relevant comparable data are scarce and incomplete. Only a handful of OECD-countries apply this distinction in their official R&D statistics. However, the Community Innovation Survey (CIS) provides information on the share of companies purchasing R&D from external partners.

Classifying extrovert and introvert systems

Figure 1 applies the proposed typology of systems by using Van Steen's indicators as a proxy for open public R&D-funding combined with data from the Community Innovation Survey (CIS 2010) as a proxy for companies' propensity to purchase R&D from external sources.

Figure 1. Public project funding as a share of GBAORD and share of innovative enterprises engaged in external R&D



Sources: 1) Steen 2012, 2) Eurostat/CIS 2010,

The 12 countries with available data vary considerably. On the one hand, countries vary slightly in terms of firms' propensity to purchase R&D from external sources, ranging from 55 per cent (Finland) to 17 per cent (Germany). On the other hand, countries differ a great deal in terms of the share of open competition based funding from public sources, ranging from 56 per cent (Belgium) to 26 per cent (Denmark).

A need for indicator development

However, the CIS-data have three major weaknesses. Firstly, they only cover R&D purchased from innovation active enterprises. Secondly, this is a yes/no question and does not provide information on the amounts of purchased R&D. Thirdly, and most importantly, this indicator also includes the procurement from other enterprises within the same group. There is reason to doubt that such internal transactions reflect a real orientation towards external suppliers of R&D.

We believe that data distinguishing extramural and intramural R&D in firms will provide a more accurate picture of private demand for external R&D. Furthermore, the experimental indicators distinguishing project based and institution based public R&D funding proposed by van Steen needs to be developed further and include more countries. Both aspects have been included in the on-going revision of the OECD Frascati manual.

If these dimensions are fully incorporated in official R&D-statistics, a further understanding of extrovert vs. introvert R&D funding systems will be feasible.

Preliminary conclusions and further work

Countries vary considerably regarding the size and the share of public/private funding in their domestic markets for R&D. Hence, the dichotomy of extrovert vs. introvert R&D funding systems seems fruitful.

However, these terms are not intended to be normative, pinpointing the extrovert system as the ideal one. Extensive use of open sources tends to create strong competition, and studies indicate that too much competition may lead to conformity. The policy challenge is therefore to strike the right balance between internal and external knowledge sources and between competition-based and direct funding. Comparable indicators in this area would allow for a better comparison of R&D funding systems, and a better understanding of the strengths and weaknesses of each system.

References

Chesbrough, H. (2006) *Open innovation – the new imperative for creating and profiting from technology*. Boston, MA:Harvard Business School Press.

Eurostat, Community Innovation Survey 2008–2010 (CIS 2010).

Lepori, B. et al. (2007) *Comparing the evolution of national policies: what patterns of change?*. Science and Public Policy, 34(6), 2007, 372–388.

OECD (2002) Frascati-Manual: *Proposed Standard Practice for Surveys on R&D*. 3rd Edition, Paris, OECD Publications.

OECD (2010) *The OECD Innovation Strategy: Getting a Head Start on Tomorrow*, Paris, OECD Publications.

OECD (2011) *Science Technology and Industry Scoreboard 2011: Innovation and Growth in Knowledge Economies*, Paris, OECD Publications.

Solberg E, Larssen K, Wiig O et al. (2012) *Markets for Applied Research: A comparative analysis of R&D-systems in five countries*, NIFU report 46/2012.

Steen, J. v. (2012) *Modes of Public Funding of Research and Development: Towards Internationally Comparable Indicators*, OECD Science, Technology and Industry, Working Papers, 2012/04, Paris, OECD Publishing.

Measuring Interdisciplinary Performance By Analyzing Mixed Node Publication Networks

André Calero Valdez calero-valdez@comm.rwth-aachen.de / Human-Computer Interaction Center, RWTH Aachen University, Am Theaterplatz 14, 52064 Aachen (Germany)

Anne Kathrin Schaar Human-Computer Interaction Center, RWTH Aachen University, Am Theaterplatz 14, 52064 Aachen (Germany)

Martina Ziefle Human-Computer Interaction Center, RWTH Aachen University, Am Theateplatz 14, 52064 Aachen (Germany)

Introduction

As Campbell (2005) states, "scientific knowledge is social". It exists as the union of individual knowledge with individual languages and idiosyncrasies. Due to effects like the confirmation bias, differences between mental models of terminology and their linguistic use are rarely recognized in interdisciplinary exchange but much more perceived in their differential occurrence and experienced as "friction" between team members. This can go as far as teams not being successful at the interdisciplinary goals they have set for themselves and external goals that have been set by funding organizations.

But what can one do in order to reap the benefits of interdisciplinary cooperation? Effective interdisciplinary research clusters need to rely on approaches to measure, steer and regulate their success with regard to the expectations of single researchers, funding agencies or industrial actors.

In the cluster of excellence (CoE) at RWTH Aachen University so-called Cross Sectional Processes (CSPs) are established to aid integration of the participating cluster domains, research fields and interdisciplinary researchers (Jooß et al. 2012). The CSPs provide and investigate services for the researchers within the cluster while at the same time enabling measuring scientific success and supporting steering and regulating of the cluster. A prospective outcome of this on-going process is a better understanding of how measurement of research processes can stabilize and increase scientific output by testable interventions, thus transforming scientific efforts into scientific outcomes. One measure is explained in this paper.

Criteria for interdisciplinary scientific performance

In order to establish a direction for steering to success one must first define success. In this case what is interdisciplinary scientific success? Success is surely multifaceted or multi-dimensional when looking at scientific success in general.

Welter et al. (2012) found that the overall objectives of scientific collaboration, such as for clusters of excellence or collaborative research centers, could be derived from the funding criteria of the DFG and the Wissenschaftsrat. Core criteria e. g. comprise:

(a) research quality,
(b) originality and coherence of the scientific program,
(c) interdisciplinary collaboration,
(d) influence on the field of research in future,
(e) options for transferability into practice,
(f) quality of scientists and their further options for development,
(g) integration of local research capacities,
(h) structural development of the university,
(i) international visibility.

Individual scientific disciplines though may define success in regard to these criteria differently. How does one determine research quality? The CSPs elaborate on these criteria using an approach that uses both qualitative and quantitative measures to triangulate the subject. In the following section a network-based approach for analyzing publication relationships is presented in the context of the CSP as a part of the CoE at RWTH Aachen University.

Mixed-Node Publication Network Analysis

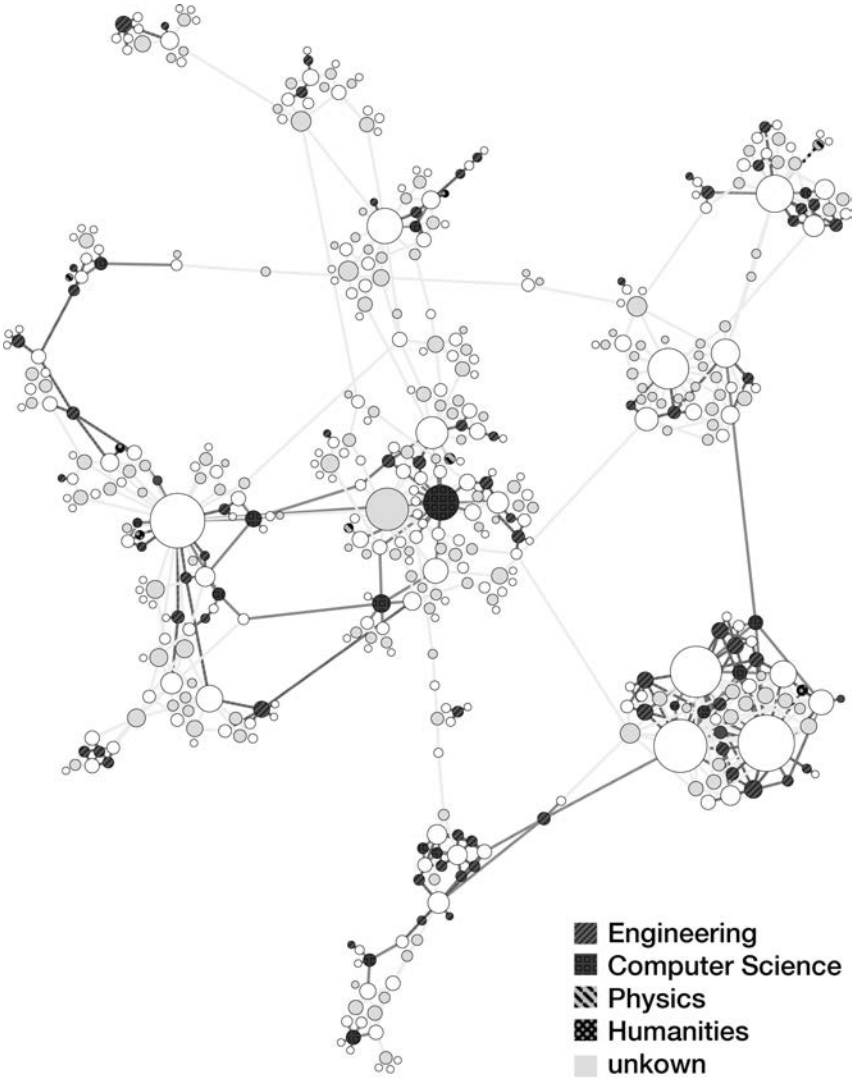
Publication analysis uses publication data in order to determine the quality of the output of scientific teams. Under the assumption that the peer review process works, publications are a natural indicator of performance (addressed criteria: a, d, and i).

Since no single metric seems to capture interdisciplinary performance well enough (Wagner et al. 2011), we try to implement a web-based self-measurement/self-management tool, which focuses on visualizing publication relationships using a bipartite author/publication graph. This data is furthermore enriched with sociometric data, spatial data and citation data, which is analyzed and evaluated for the individual researchers that logs into the web-application.

For this purpose various network metrics are used and visualized such as centrality measures (betweeness, degree, eigenvector) as well as graph entropy measures. Centrality measures enable detection of important weak connections (people that connect otherwise unrelated work groups). Possible candidates for entropy measures are topological information content as in Mowshowitz (1968), parametric graph entropies (Dehmer 2008), network entropies as in Solé (2004) and graph entropies based on ER models as in Ji et al. (2008). These measures are applied on the micro level in contrast typical outlet-focused analyses (Leydesdroff et al. 2011). Several of these measures can be compared between workgroups or even automatically detected communities in order to detect self-similarity, symmetry and connectedness.

Preliminary results show a high level of connectedness between authors in the cluster (Calero Valdez et al. 2012). A preliminary analysis of the partaking disciplines in the cluster is shown in Figure 1.

Figure 1. Example visualization of all publications in the CoE. White circles denote publications; colored circles denote authors and their discipline; circle size denotes degree centrality.



Furthermore the applicability of these graph visualizations as a self-measurement tool for the researchers will be investigated from an acceptance point of view.

References

Calero Valdez, A. et al. (2012): Using mixed node publication network graphs for analyzing success in interdisciplinary teams. In: *AMT*, 606–617.

Campbell, D.T. in Derry, S. J., Schunn, C. D., & Gernsbacher, M. A. (2005). Interdisciplinary Collaboration: An Emerging Cognitive Science. Psychology Press. 2nd Edition, 3–21.

Dehmer, M. (2008): Information processing in complex networks: Graph entropy and information functionals. *Applied Mathematics and Computation 201*(1–2), 82–94.

Ji, L., et al. (2008): Network entropy based on topology configuration and its computation to random networks. *Chinese Physics Letters* 25(11), 4177.

Jooß, C. et al. (2012): Scientific Cooperation Engineering in the Cluster of Excellence Integrative Production Technology for High-Wage Countries at RWTH Aachen University. In: *ICERI2012 Proceedings*, 3842–3846.

Leydesdorff, L. Rafols, I. (2011), Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations, *Journal of Informetrics*, Volume 5, Issue 1, 87–100.

Mowshowitz, A. (1968): Entropy and the complexity of graphs: I. an index of the relative complexity of a graph. 30, 175–204.

Solé, R., Valverde, S. (2004): Information Theory of Complex Networks: On Evolution and Architectural Constraints. In Ben-Naim, E., Frauenfelder, H., Toroczkai, Z., eds.: *Complex Networks. Volume 650 of Lecture Notes in Physics*. Springer, Berlin, Heidelberg, 189–207.

Wagner, C.S. et al. (2011) Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature, *Journal of Informetrics*, Volume 5, Issue 1, January 2011, 14–26.

Welter, F. et al. (2012): Scientific Performance Indicators – Empirical Results from Collaborative Research Centers and Clusters of Excellence in Germany. *XVI IRSPM Conference*, 11–13 April 2012, University of Torvergata, Rome, Italy, 1–22.

Academic Career Structures – Historical Overview Germany 1850–2013[1]

Cathelijn J. F. Waaijer c.j.f.waaijer@cwts.leidenuniv.nl / Centre for Science and Technology Studies, Wassenaarseweg 62A, Leiden, 2300 AX (The Netherlands)

Introduction

The current academic career system is characterized by long selection periods and small initial chances of achieving a successful lifelong career in academia. Young academic researchers work in temporary jobs for a longer period than university graduates in other sectors (Statistisches Bundesamt 2013, Van de Schoot, Yerkes & Sonneveld 2012). In this study we investigate how this system has developed from the time research was incorporated into universities until now, and start by tracing the historical development of the academic career system in Germany from the early 19th century until now, as Germany was the leading country in science from the 19th century until the end of the Weimar Republic in 1933.

Academic positions in Germany

In the 19th century, only three official academic positions existed: the ordinary professorship (*ordentlicher öffentlicher Professoriat or Ordinariat*), the extraordinary professorship (*außerordentlicher Professoriat or alternatively, Extraordinariat*) and the private lectureship (*Privatdozentur*). At first, only ordinary and extraordinary professors were paid by the university; private lecturers only received lecture fees from their students (Ben-David & Zloczower 1961, Weber 1946). Before a man could fill an official academic position, he needed to obtain first a PhD and after that a *Habilitation*, based on original research. The work associated with both theses was unpaid and persons working towards these theses were not considered to be academics just yet (von Ferber 1956, Weber 1946).

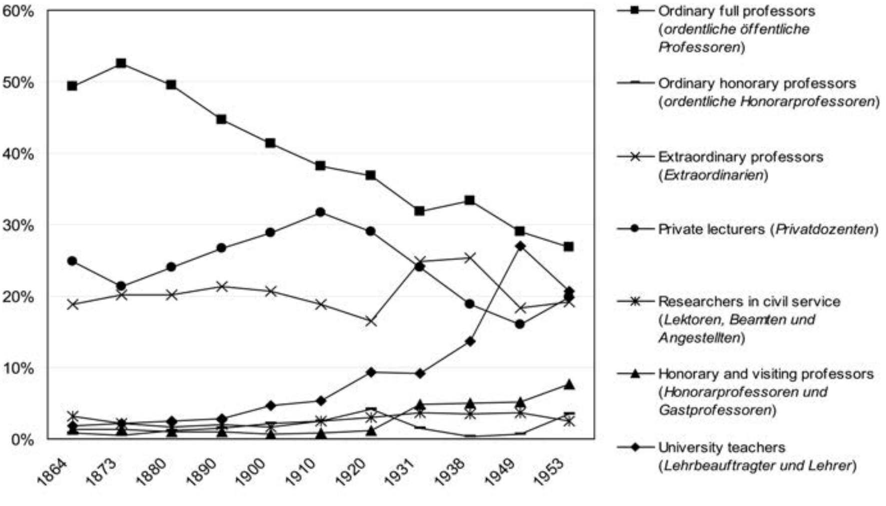
Over the course of the 19th century, private lecturers became paid universities employees, too. With the emergence of larger laboratories, research affiliateships emerged (*wissenschaftliche Assistenturen*, Bock 1972). At first, research affiliates (called "assistants" in German) worked as literal assistants to professors, by aiding them during lectures or practical experiments. Later affiliateships became formal positions for researchers working on their PhD or *Habilitation* dissertation (Bock 1972).

1 Cornelis A. van Bochove, Anthony F.J. van Raan and Inge C.M. van der Weijden are gratefully acknowledged for fruitful discussion of the study and manuscript.

Development of academic positions 19th century until mid 20th century

When looking at the distribution of all positions, the main trend is that the relative number of ordinary professorships declined, while other positions below the professorship (private lecturers, extraordinary professors and a heterogeneous group of [part-time] university teachers and researchers) grew, but took turns in "filling the gap" left by the relative decrease in ordinary professorships (Figure 1). Unfortunately, no data about the number of research affiliates before 1952 were available.

Figure 1. Distribution of academic positions at German universities 1864–1953. For 1953 West Germany only.



Source: von Ferber 1956 pp. 195, 210.

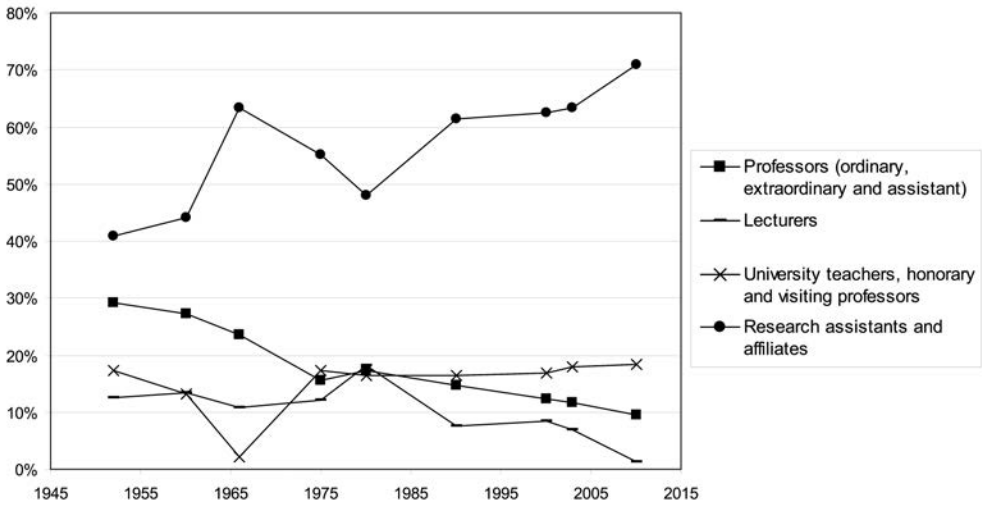
Development of academic positions mid 20th century until now

From the end of World War II until now the relative number of professorships declined even further, from 30% to 10% (Figure 2). On the other hand, the relative number of research affiliates and assistants (*wissenschaftliche Hilfskräfte*) rose from 40% to 70%. At first many of them, like faculty staff, were employed on a permanent contract (50% in 1980; Statistisches Bundesamt 1982). However, after this period a trend emerged to employ research affiliates on temporary contracts (only 30% employed on permanent contracts in 1990; Statistisches Bundesamt 1992).

Conclusions

Our results show a professionalization of the scientific enterprise in Germany from the 19th century until today. At first academic positions were only found in the higher echelons of academic career ladder. Over time positions below the professoriate became more proper academic positions of their own. Finally, our results suggest the relative number of researchers on temporary contracts has increased.

Figure 2. Distribution of academic positions 1952–2010.



Source: Statistisches Bundesamt, *1953, 1966, 1969, 1976, 1982, 1992, 2004 & 2011. For 1953–1980 West Germany only.*

Further lines of investigation

The relative availability of academic positions for PhD graduates influences the chances of achieving a lifelong career in science. We will compare the ratio of PhD graduates to the number of academic positions. These results will be shown at the conference.

References

Ben-David, J. & Zloczower, A. (1961). The Idea of the University and the Academic Market Place. *European Journal of Sociology*, 2:303–314.

Bock, K.D. (1972). *Strukturgeschichte der Assistentur: Personalgefüge, Wert- und Zielvorstellungen in der deutschen Universität des 19. und 20. Jahrhunderts*. Düsseldorf: Bertelsmann Universitätsverlag.

Ferber, C. von (1956). *Die Entwicklung des Lehrkörpers der deutschen Universitäten und Hochschulen* 1864–1954, vol. III, edited by H. Plessner. Göttingen: Vandenhoeck & Ruprecht.

Schoot, R. van de, Yerkes, M. & Sonneveld, H. (2012). The employment status of doctoral recipients: An exploratory study in the Netherlands. *International Journal of Doctoral Studies*, 7:331–48.

Statistisches Bundesamt (1953). Statistische Berichte: *Die Lehrpersonen und das wissenschaftliche Hilfspersonal an den wissenschaftlichen Hochschulen des Bundesgebietes und West-Berlin im Wintersemester 1952/53*. Stuttgart: Kohlhammer Verlag.

Statistisches Bundesamt (1966). *Bevölkerung und Kultur: Hochschullehrer und sonstiges wissenschaftliches Personal an den Wissenschaftlichen Hochschulen 1960*. Stuttgart: Kohlhammer Verlag.

Statistisches Bundesamt (1969). *Bevölkerung und Kultur: Hochschullehrer und sonstiges wissenschaftliches Personal an Wissenschaftlichen und Pädagogischen Hochschulen 1966*. Stuttgart: Kohlhammer Verlag.

Statistisches Bundesamt (1976). *Bevölkerung und Kultur: Personal an Hochschulen 1975*. Stuttgart: Kohlhammer Verlag.

Statistisches Bundesamt (1982). *Bildung und Kultur: Personal an Hochschulen 1980*. Stuttgart: Kohlhammer Verlag.

Statistisches Bundesamt (1992). *Bildung und Kultur: Personal an Hochschulen 1990*. Stuttgart: Metzler-Poeschel.

Statistisches Bundesamt (2004). *Bildung und Kultur: Personal an Hochschulen 2003*. Stuttgart: Metzler-Poeschel.

Statistisches Bundesamt (2011). *Bildung und Kultur: Personal an Hochschulen 2010*. Stuttgart: Metzler-Poeschel.

Statistisches Bundesamt (2013). *Jobs ohne Befristung*. https://www.destatis.de/DE/Publikationen/STATmagazin/Arbeitsmarkt/2013_05/2013_05Beschaeftigte.html. Retrieved 28 May 2013.

Weber, M. (1946). *Science as a vocation* (129–156). Edited by H. H. Gerth and C. Wright Mills. From Max Weber: Essays in Sociology, New York: Oxford University Press

Core Patents Subject Mining Based On Citations and Subject Terms Cross Co-Occurrence

Xu Haiyun xuhaiyun@mail.las.ac.cn / Chengdu Library of the Chinese Academy of Science, Yihuan Road, Chengdu, 610041 (P. R. China)

Yue Zenghui yuezenghui@mail.las.ac.cn / Chengdu Library of the Chinese Academy of Science , Yihuan Road, Chengdu, 610041 (P. R. China)

Fang Shu yuezenghui@mail.las.ac.cn / Chengdu Library of the Chinese Academy of Science , Yihuan Road, Chengdu, 610041 (P. R. China)

Introduction

As an important research field of the international scientometrics, Patentometrics is getting more and more concern from international scientific metrologists. According to the statistical data of WIPO, about 90%–95% of the world's inventions can be found in the patent literatures, and many inventions can only be found in the patent literatures. As the most important manifestation of the achievements of the technology research and development, patents can represent a well-defined output of the technological innovation process and can be measured through lots of statistical indicators. Patents are legal documents used for describing the intellectual property rights which are granted by the statutory authority through legal procedures, they can be used for indicating and evaluating the technical innovation and progress. Patents are not only the important measurement indicators of the achievements of R&D activities, both also the important economic means for the market competition among enterprises. Now, we can know the spatial and temporal distribution of the countries, enterprises and inventors, which own the main patented technology in a certain technological area, through patentometrics. Then we can analyze the transfer of technology centres or the evolution of some enterprises. Patentometrics can also be used to realize the international patent strategy of a particular country or enterprise, and then provide an important reference for countries to develop and perfect the relevant patent protection system and patent strategy.

At present, the output of the global patent literatures grows rapidly. While the few core patented technology play a key role for the technological advances among the massive patent applications. Studies have shown that the evaluation value of the patents is quite small and it shows great value skewness. Schankerman (1987) estimated that about 5% to 10% of the patents represent a half of the total ones. Therefore, how to effectively judge the value of patents, and find out the core patent literatures, then conduct the deep analytics of the details of the technology rights, are the important prerequisites for grasping the level of technological development and the breakthrough direction of the technology foresight. The core patented technology serves as a key link between the past and future in the technosphere. How to find out the core patented technology are the important prerequisites for the study of the scope of patent rights, tracking technology

development trends and grasping the technical distribution trends. Fang (2007) described Patentometrics like this "Analysis the subsequent cited patents of the core patents and classify the subsequent cited patents according to the degree of similarity of technical topics will help to understand the origin of the various technology components in a technical field and the related research progress, and analyze the future direction of the technology development".

Core patent Mining

Means of mining core patents

The subject terms extracted from the patent data can reflect the concepts and features for various industries. A simple statistical analysis can roughly show the major industrial technologies and the extent of technical research, but it can not show the relationship between these technologies. Law (1988) proposed that with the continuous improvement of the information technology, we are no longer satisfied with the simple statistical analysis of the patents, and we hope to gain the deeper technical knowledge through the mining of particular text information, such as the patent titles, abstracts, rights specifications, even the full literatures. The industrial technologies are composed of many technical thematic priorities, and there are also some relationships between the thematic priorities. Technical terms can in a way to characterize the core concept of a technical topic. A technical topic, however, often contains multiple characteristics, and its implementation usually requires the cooperation of a number of technologies. A single technical term is unable to show the whole technical panorama. That means we need to deeply learn about the technical terms of a technical topic, as well as the relationships between them.

The methods of science frontiers research among the bibliometric ones are mainly based on the paper citations, the high co-occurrence of subject terms characteristics or clustering of strong ties, revealing the science frontiers in related field through the time evolution. Such high co-occurrence or strong ties usually indicate a high degree interaction among network actors, and the actors are more intimate in some interactive form. Therefore, they are more easily identified in the scientific frontiers exploration. Agrawal and Imieliński (1993) proposed that association rule mining can discover the interesting associations between mass data items. If there are associations among multiple properties, one property can be predicted based on the value of the other ones. The support degree of the association rules equals to the ratio: the number of transactions which contain both X and Y divided by the number of all transactions in the transaction set. The support degree reflects the frequency of items in X and Y contained in the same transaction set simultaneously. The confidence coefficient of the association rules equals to the ratio: the number of transactions which contain both X and Y divided by the number of transactions which contain X in the transaction set. The confidence coefficient reflects the conditional probability that Y appears in a transaction including X.

Co-word analysis and co-citation analysis

There are two main methods of the co-occurrence analysis: co-citation analysis and co-word analysis (i.e. word co-occurrence analysis). Small (1973) proposed the co-citation analysis and its basic definition like this: When the two literatures (authors) are cited by the third one at the same time, then the two literatures (authors) are co-cited. Co-word analysis is a content analysis technique, which uses the co-occurrence of words or phrases to describe the relationship between the concepts. Co-citation analysis is based on the reference relationships between the literatures, so it is useless if the literatures failed to provide a reference relationship, so there are some limitations in co-citation implementation. According to Matsuo (1974) co-occurrence model has been widely applied to text classification, clustering, keyword extraction, duplication check of text and automatic summarization. Seidel (1949) described the idea of patent citations. Garfield (1966) studied the patent citations indicators. Michel (2001) conducted a deep analysis of the patent citations from the point of the patent search report writing.

Zhao (2009) proposed that Co-citation analysis (CA) is an important method to study the knowledge structure in a field, explore the disciplinary structure evolution. It is an important research direction of information science and widely used in many fields. Scientists will communicate with each other closely in various forms, reference citations is one of the main ways. Citations can reflect the compactly selection process when the scientists are writing papers, and the relationship between their works and those of others. Citations can also show the researchers' insights on emerging topics. According to the reference relationship between the literatures, we can identify those core papers cited by most authors, the highly-yield research teams and their activeness, etc. The references in highly cited papers may hide the relevance among innovative ideas.

However, due to the complexity of knowledge, there is another phenomenon appeared in the research frontier exploration which is worthy our in-depth researching. For instance, many data mining algorithms for the analysis of subject terms usually neglect or discard the data objects, which are not comply with or well support the data model, so some important data records are cleared as the noises in the data set. Actually, the important information hided in these outliers or isolated points, especially those themes with low support degree and high confidence coefficient sometimes is worthy more attention. Secondly, as co-word analysis usually uses high-frequency words in the topic clustering process to explore research fronts, thus the classes composed by relatively low frequency topics can not be reflected in the clustering, so the panorama of the subject can not be reflected. Finally, the information in the strong co-occurrence relations usually is duplicated, and it is easy to form a closed system, which is not conducive to the development of science frontiers. However, the "weak ties" characterized by low-cost and high-effect in the dissemination of information usually can provide better access to knowledge innovation. They transform the potential knowledge into a new hotspot or research frontier. Therefore, we need to find the critical low support degree while high confidence coefficient topic clustering and form an analytical method of comprehensive mining patent topics.

Cross Co-occurrence of subject heading and patent Citation

By a co-citation analysis, we can get a cluster of literature with the same research theme, and by a social network analysis, we can discover hot research topics in a certain field as well as the relationship between the topics. Although through the co-citation analysis, we can get the associated knowledge and their structure, but can not directly reflect the theme of the network. Usually, analysts apply a direct co-citation analysis to keywords and citations literature to get the content of the network structure in micro-level analysis. Such a direct co-occurrence way can reveal the theme of the cited literature to some extent, that the theme can play as an identification role to the citied literatures. However, the way of direct co-occurrence theme mining is easily to be dominated by high-frequency words and high cited literatures, leading to neglect to those low-frequency words or phrases as well as weak ties. Wang (2012) proposed that if we add the subject heading factor into the literature citation network, through the literature and subject heading cross co-occurrence analysis, we can identify the research topic of the co-citation network. What's more, we can identify the research topics around the same subject heading. Cross co-occurrence put the citation literature which had no direct co-occurrence into a group, with the same subject terms. So cross co-occurrence can be a supplement to the insufficient theme mining only through the high frequency direct co-occurrence method.

This study proposes the cross co-occurrence analysis methods to the indexing subject terms and cited patent. The indexing subject terms and cited patent cross co-occurrence network belongs to double dimensions data structures, which contains both the subject terms and the cited patents. Multiple cited patent literature with the same subject terms establish indirect co-occurrence network. So the indexing subject terms play as the bridge in the patent cited network, and are able to reflect the associated topics over two indirect cited patents.

The patent literatures in the field of "Medical Microscopy Endoscopy system and technology" are selected as the analysis data and DII database as the retrieval database. First, through literature research and communication with the field experts, we construct retrieval strategies and finally got 2697 patent family items. Through the artificial discrimination as well as the communication with field experts, the patent data about "Medical Microscopy Endoscopy system and technology" can be basically divided into 32 technology and efficiency subthemes. Patent literature unlike the scientific literature with indexing keywords from their authors, so we do a segmentation to the titles and summaries of the 2697 patent records. Then based on the segmentation, we proceeded a data cleaning process to the topic phrase by compiling vocabulary and removing the phrases with little mean in our analysis, the final Subject Terms and their frequency are shown in Table 1:

Table 1. Technology and Efficiency themes of "Medical Microscopy Endoscopy system and technology" (Fr.: frequency)

Topics	Fr.	Topics	Fr.	Topics	Fr.
Image Camera	765	Cost Reduction	88	display device	25
Imaging Methods	461	S/N Enhancing	81	Position/Distance Measurement	22
Light/Source	452	Image Display	72	Miniaturization	20
High-resolution	435	Convenient Structure	67	Reduce Trauma	20
Magnification	285	Sight-View Improving	47	Signal/Image transmission Pipe-lines	19
Work-Station	194	General Structures	45	manufacturing method	12
Picking up Circuits	188	Others	41	Operation Methods	10
Probe	152	Mannufacturing Mate-rials	39	Long Working Distance	7
Imaging Sensors	146	Easy Accessibility	33	Energy Efficiency	5
Image Processing	136	SurgicalInstruments	32	Good Color-reduction Degree	3
Fast Imaging	125	Optical Properties Improving			

After got the 32 subject terms, we selected 182 patent records with cited frequency up to 6, then got a 182 * 32 two-mode cross co-occurrence matrix for patents – subject terms. To get a weighted cross co-occurrence map, we made a matrix transformation to the two-mode cross co-occurrence matrix: first, we increase the rows and columns of the matrix to get a weighted cross co-occurrence matrix data. Second, we set "0" to the patent co-occurrence matrix, also set "0" to subject terms co-occurrence matrix. Such an operation respectively adds one 32 * 32 and one 182 * 182 zero square matrix module into the 182 * 32 two-mode cross co-occurrence matrix. So the raw two-mode matrix data is formed into a bipartite two-mode data, and that is a square matrix. Then we extracted the co-occurrence pairs, whose co-occurrence frequency up to 7, the patent cited frequency and subject terms are as shown in Table 2.

Table 2. Patents – Subject terms cross co-occurrence

No.	patent citations	Subject terms	Co-Fr.	C-Fr.	T- Fr.
1	JP 9090244	high resolution	13	14	435
2	US 4312572	Image Camera	9	11	765
3	US 6485413	Imaging Methods	8	22	461
4	US 5120953	Imaging Methods	8	20	461
5	US 4841977	high resolution	8	10	435
6	US 4951677	high resolution	8	10	435
7	US 4312572	Magnification	8	11	285

… Continuation Table 2

No.	patent citations	Subject terms	Co-Fr.	C-Fr.	T- Fr.
8	US 5120953	Image Camera	7	20	765
9	US 4781448	Image Camera	7	10	765
10	US 5321501	Imaging Methods	7	15	461
11	US 5459570	Imaging Methods	7	15	461
12	US 6485413	Light/Source	7	22	452
13	US 5120953	high resolution	7	20	435
14	US 4917097	high resolution	7	11	435
15	US 5000185	high resolution	7	10	435
16	US 3938502	high resolution	7	10	435
17	US 4494549	high resolution	7	10	435
18	US 4587972	high resolution	7	10	435

Co-Fr. is short for cross co-occurrence frequency, C-Fr. is short for citation frequency, T- Fr. is short for terms frequency.

To show the differences between cross co-occurrence and a single direct co-citation analysis, we get the co-occurrence matrix for the 182 patent literatures, in which co-occurrence pairs with a co-occurrence frequency up to 7, see Table 3.

Table 3. TOP20 Highly Cited Patents

No.	patent citations	Fr	No.	patent citations	Fr
1	US 6485413	22	15	US 4917097	11
2	US 5120953	20	16	US 4312572	11
3	US 6294775	16	17	US 4841977	10
4	US 5321501	15	18	WO 9838907	10
5	US 5459570	15	19	US 4951677	10
6	US 6069698	15	20	US 5000185	10
7	JP 9090244	14	21	US 5295477	10
8	US 5034613	13	22	US 5603687	10
9	US 6134003	13	23	US 5749830	10
10	US 6370422	13	24	US 3938502	10
11	US 4364629	13	25	US 4494549	10
12	US 2001055462	12	26	US 4587972	10
13	US 5742419	12	27	US 4781448	10
14	US 4862873	11			

Table 2 lists the co-occurrence pairs with the patent indexing subject terms co-occurrence frequency up to 7, a total of 18. As you can see from the table, there is a direct relationship

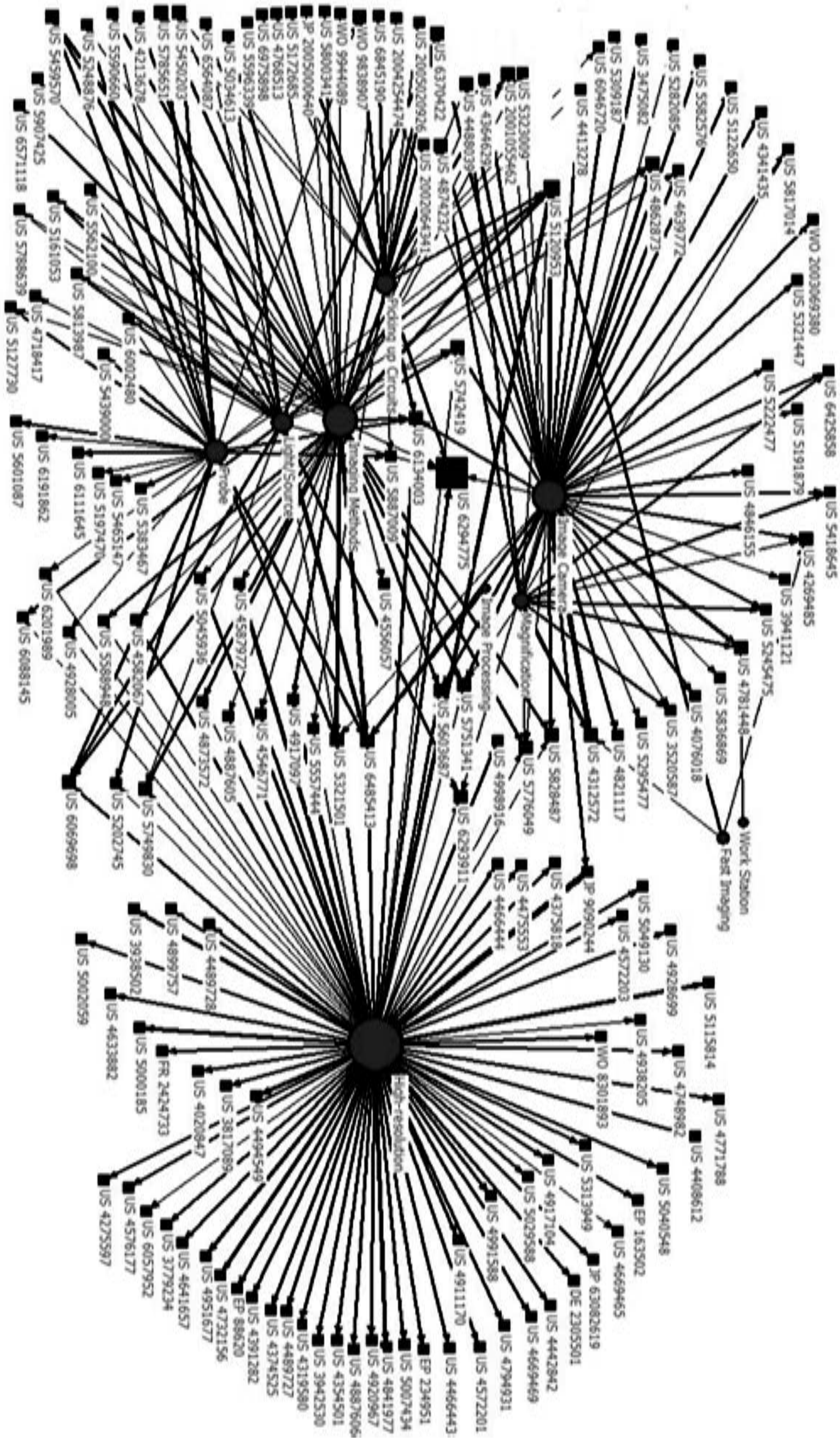
between the frequency of the cited patent and indexing subject terms and the frequency of patent citations, indexing subject terms. Highly cited papers and high-frequency indexing subject terms are easy to form a high co-occurrence frequency. Among the 18 co-occurrence pairs with the co-occurrence frequency up to 7,there are 15 cited patents ranking in the top 27, and 5 subject terms ranking in the top 5. However, the co-occurrence subject terms frequency has changed in contrast with the absolute frequency. From Table 2, we can also see that as the decrement of the co-occurrence frequency, the number of co-occurrence pairs with the same occurrence value increases rapidly.

The cross co-occurrence and co-occurrence network comparison

Patent – subject terms cross co-occurrence map is a bipartite graph. Bipartite graph is a graph, in which the vertices can be divided into two disjoint collections with different attributes; the vertices in the same collection are not adjacent. Using mathematical language, it can be expressed as: Graph $G = \langle V, E \rangle$, for any vertex V, there is $V1 \cup V2 = V$, $V1 \cap V2 = \Phi$, for any edge' two vertices $(x, y) \in E$ in G, there are $x \in V1$, $y \in V2$. Or $x \in V2$, $y \in V1$, then G is a bipartite graph.

Network would be visualized using social network software Ucinet, and in this process, the two-mode square matrix would be seen as one mode data. Node size is the symbol of the network node degree, see Figure 1 below, where the red dot is the indexing subject terms, while those blue square are patent literatures. In order to do a comparison analysis, we do a patent citation co-occurrence analysis to those 182 patents, see Figure 2 below.

Figure 1. Patents – Subject terms cross co-occurrence

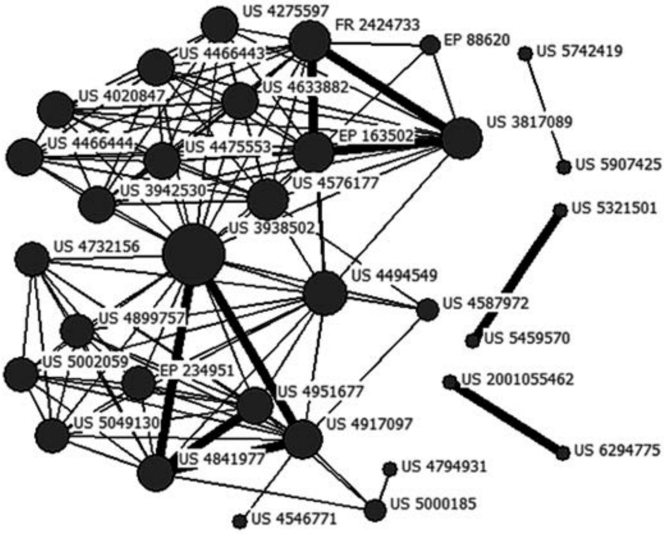


It shows that in figure 1 "High-resolution", "Imaging Methods", "Image Camera", "Picking up Circuits", "Light/Source", "Probe", "Magnification", "Fast Imaging", "Work Station", "Image Processing" and these subject terms formed into several obvious clusters. Those clusters are able to represent current protection themes about "Medical Microscopy Endoscopy system and technology". "High-resolution" has the highest network degree, and proves that High-resolution is one of the most important effectiveness themes in this research and patent protection area. "Imaging Methods" and "Image Camera" are also have high network degree, and this can illustrate imaging methods and imaging camera are an important technical topics. At the same time we can see the presence of a large number of patents connecting three technical-efficacy subject terms "High-resolution","Imaging Methods" and "Image Camera". It is not difficult to understand that the improvement of Imaging Methods and Image Camera production material and process can both improve the High-resolution efficacy. Removing those leaf nodes with single theme, the nodes left are patents having two or more subject terms. Therefore, through the patent literatures bridging technical and efficacy, we can effectively explore the efficacy of corresponding technical topics, or the effectiveness of technical topics may be achieved. At the same time, we can find the same patent clusters with the same technical-efficacy subjects.

Patent US6294775 is not the most highly cited patent, but it has the largest degree in the cross co-occurrence network. This patent has strong co-occurrence strength with "High-resolution", "Imaging Methods", "Image Camera" and "Picking up Circuits", indicating that the patent covers these subjects terms and has more weight on them. What's more, besides Imaging Methods and Image Camera technology improvement, Picking up Circuits technology and system improvements can also increase High-resolution efficacy. The cited patent US6485413 has strong co-occurrence relationship with "High-resolution", "Imaging Methods", "Image Camera", "Probe" and "Light/Source". So it shows that patent US6485413 involves Imaging Methods, Image Camera, Probe, Light/Source technology subject, while these several topics are important to improve the effectiveness of High-resolution.

Cross co-occurrence of patent and subject terms can directly and clearly reveal technology and efficacy of the topics reached in the protection field. Also this method can reveal the subject relationship among core patents. Compared with co-occurrence analysis, cross co-occurrence is somewhat weaker in finding the subjects clusters, for after the indexed subject terms filter, some actually important nodes or subject clusters may be weaken or even disappear. For example, in the co-occurrence analysis network (Figure 2), patent US3938502 bridge two largest clusters, but it is not the most cited patent (citation 10), so the cross co-occurrence offers different information with the single co-occurrence analysis method, although may also not be aware of the important patent literature. In figure 1 of the cross co-occurrence network, Patent US3938502 only form a strong co-occurrence relation with "High-resolution", under certain co-occurrence threshold value, this patent is not the centre of the network. This may because of the distinction and limitation of the subject terms, patents which originally play a connecting role, disappear in the cross co-occurrence network. Therefore, cross co-occurrence analysis is able to directly reveal the subject, and has higher readability.

Figure 2. Patent citations co-occurrence



Summary

Patent–indexing subject terms cross co-occurrence generated two-dimensional element network, and in such a network, some cited patents formed indirect patent co-citation network through the delivery of the indexing subject, while some patens is filtered out. This article respectively visualized the indexing Patent – indexing subject terms cross co-occurrence and the single patents co-citation analysis, and compared the similarities and differences between the two analysis methods. Important node of the patent – indexing subject terms cross co-occurrence (CKCP) can identify the important nodes with low correlation that direct co-occurrence will filter out, while it can analysis the efficacy constitute of indirect co-occurrence core patented technology. In addition, cross co-occurrence put the citation literatures having no direct co-occurrence into a group, with the same subject terms. So Cross co-occurrence can be a supplement to the insufficient theme mining only through the high frequency direct co-occurrence method. However, the cross co-occurrence analysis may filter out some important documents or clusters, since the subject terms formed a strong filtering effect. Therefore a combination of direct co-occurrence and the cross co-occurrence analysis is able to make the core patents subject mining more complete.

References

Schankerman, M. & Pakes, A. (1987). Estimates of the Value of Patent Rights in European Countries During thePost-1950 PeriodNational Bureau of Economic Research Cambridge, Mass., USA

Shu, F & Xian Z & Guohua Xiao. (2007). Patent Intelligence Analysis Method Research and Application. DOCUMENT,INFORMATION & KNOWLEDGE.

Agrawal, R. & Imieliński, T. & Swami, A. (1993).Mining association rules between sets of items in large databases. *ACM SIGMOD Record*. 22, 207–216.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24, 265–269.

Matsuo, Y. & Ishizuka, M. (2004).Keyword extraction from a single document using word co-occurrence statistical information. *International Journal on Artificial Intelligence Tools*, 13, 157–169.

Seidel, A. H. (1949).Citation system for patent office. Journal of the Patent Office Society, 31,554.

Garfield, E. (1966).Patent Citation Indexing and the Notions of Novelty, Similarity, and Relevance. *Journal of chemical documentation*, 6, 63–65.

Michel, J. & Bettels, B. (2001). Patent citation analysis. A closer look at the basic input data from patent search reports. *Scientometrics*, 51,185–201.

Chen, C. & Ibekwe - SanJuan, F., Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple perspective cocitation analysis. *Journal of the American Society for Information Science and Technology*, 61,1386–1409.

Wang, L & Leng, F. (2012). Keywords and citations of academic co-occurrence relationship analysis and empirical research. *Information studies:Theory & Application*, 35, 82–86.

Regional outreach and engagement of universities: a scientometric view

Alfredo Yegros-Yegros a.yegros@cwts.leidenuniv.nl / Centre for Science and Technology Studies, Leiden University, PO Box 905, 2300 AX Leiden (the Netherlands)

Robert J. W. Tijssen tijssen@cwts.leidenuniv.nl / Centre for Science and Technology Studies, Leiden University, PO Box 905, 2300 AX Leiden (the Netherlands)

Introduction

Complementary to teaching and research, universities have during the last decades also been emphasizing their contribution to economic and social development through their so-called "third mission" (Molas-Gallart, et al., 2002). Such engagement with external partners and stakeholders outside the university system usually takes place within a regional context and can materialize in different ways (technology transfer activities, contracted research with non-academic partners such as industry or other kind of entities, continuing education programs, etc.). Universities are expected to contribute to the socio-economic development of the regions where they are located while at the same time they are competing to achieve high quality research results at the international level (Goddard, 2007).

Developing university performance indicators of 'regional engagement' proves to be a challenge although concerted international initiatives, such as U-Multirank, are now on-going (Van Vught and Ziegeler, 2012). Previous empirical studies have focussed on research collaboration of universities (e.g. Benavent-Perez, et al., 2012) but without a specific focus on regional collaborations between universities and non-university partners. Other studies, however, have a clear focus on the collaborations of universities at the regional level but focus mainly on research partners from the private sector (Ramos-Vielba, et al., 2010; Abramo, et al., 2012; Tijssen, 2012).

The objective of this study is to explore, from a broad analytical perspective, the degree of *Regional Outreach and Engagement* (ROE) between universities and non-university partners located in the same region and also to gauge the relationship between a university's ROE performance and its research performance.

Method

The selected publications comprises of all research articles and research reviews, extracted from the *Web of Science* database (WoS), which were published by 60 universities in the Netherlands

and Spain – 13 research universities in the Netherlands and 47 public universities in Spain[i]. These two countries were selected as case studies because of their significant differences in terms of their national system of regional administrative units (NUTS region system) and organization of the higher education system.

The author addresses in their 2006–2010 WoS-indexed publications have been analysed to identify regional research partners. The comparative analysis is done at the country level. The bibliometric indicators in this study are:

– Total output: number of documents published by each university.
– Regional Outreach and Engagement (ROE): the share of publications in which a given university collaborates with a non-university partner (e.g. company, hospital, regional government, etc.) located within the same NUTS2 region as the university or university campus.
– Mean Normalized Citation Score (MNCS): average number of citations of the publications of a university normalized for differences in fields, publication year and document type.

We classify ROE publications according to the geographical locations of all participating institutions into the following four categories:

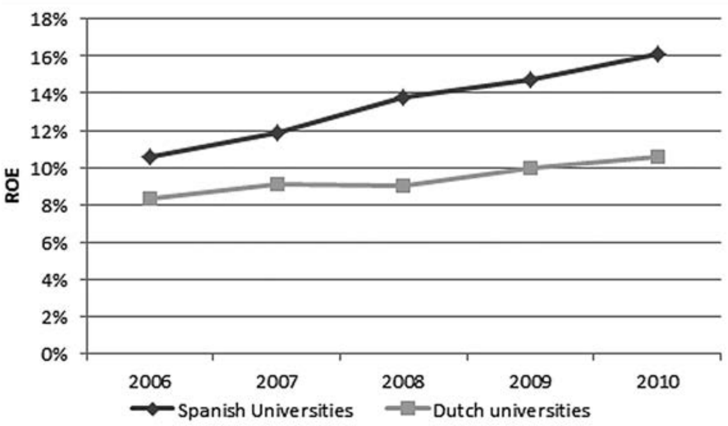
– Only Regional (R): when the university publication involves collaboration with one or more non-university partners located within the same NUTS2 region.
– Regional + Domestic (R+D): The non-university partners are located in the same region and also at least one partner is located in a different region of the same country.
– Regional + Foreign (R+F): The non-university partners are located in the same region and also at least one partner is located in a different country.
– Regional + Domestic + Foreign (R+D+F): The university collaborates at least with three different non-university partners, one or more located in the same region, in another region of the same country and in another country.

Results

General trend

The 60 Dutch and Spanish universities included in the analysis published 269,554 articles and reviews in the period 2006–2010. As shown in Figure 1, the regional outreach and engagement both in Spanish and Dutch universities has increased steadily over time, especially in Spain.

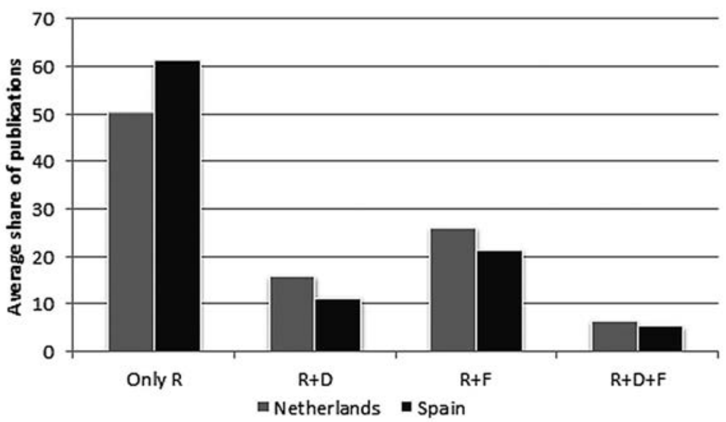
Figure 1. Increase of ROE publication output at Dutch and Spanish universities



How regional is the regional engagement?

The set of ROE publications has been analysed in order to identify the geographical location of non-university partners for each university. As shown in Figure 2, in both countries the highest share of these publications includes intra-regional external partners only. However, on average, an important proportion of publications (50% in the case of Netherlands and 39% in Spain) also include at least another non-university partner from a different geographical location, often a foreign country.

Figure 2. Location of non-university partners within ROE publications



The statistical analyses performed show a non-significant statistical relationship (Pearson correlation $r = 0.03$) between the 60 ROE and MNCS scores.

Conclusions

Our results show that regional research outreach and engagement of universities is increasing over time, at least when this engagement is measured through institutionally co-authored scientific publications within the *Web of Science*.

Regional engagement and citation impact seem unrelated in the Netherlands and Spain. Universities that are more regionally engaged in research are, on the whole, not underperforming compared the other universities in terms of impact.

When focussing on ROE publications, we find that regional cooperation often also involves partners in other geographical locations, either within other regions of the country or abroad. We observe a similar pattern between Dutch and Spanish universities as to the co-location of non-university partners.

Further research will focus on: alternative definitions of a "region"; extended identification and labelling of research partners (i.e. institutional sector); and determining the extent to which geographic collaboration patterns depend on characteristics of the higher education systems or the general type of university involved.

References

Abramo, G., D'Angelo, C., Solazzi, M. (2012). A bibliometric tool to assess the regional dimension of university-industry research collaborations. *Scientometrics*, 91,955–975.

Benavent-Perez, M., Gorraiz, J., Gumpenberg, C., Moya-Anegon, F. (2012). The different flavors of research collaboration: a case study of their influence on university excellence in four world regions. *Scientometrics*, 93, 41–58.

Goddard, J. (2007) Higher education and regions: Globally competitive, locally engaged. Paris: OECD.

Molas-Gallart, J., Salter, A., Patel, P., Scott, A. y Duran, X. (2002). Measuring Third Stream Activities. Final report to the Russell Group of Universities. SPRU, University of Sussex.

Ramos-Vielba, I., Fernandez-Esquinas, M., Espinosa de los Monteros, E. (2010). Measuring university-Industry collaboration in a regional innovation system. *Scientometrics*, 84,649–667.

Tijssen, R.J.W. (2012). Co-authored research publications and strategic analysis of public-private collaboration, *Research Evaluation*, 21, 204–215.

Van Vught, F. & Ziegele, F. (2012). Multidimensional Ranking: The Design and Development of U-Multirank. Springer Series: Higher Education Dynamics, Vol. 37.

i Long-distance teaching universities were not included in the study: Open University (Netherlands), Universidad Nacional de Educacion a Distancia (Spain), as well as Universidad Internacional de Andalucia and Universidad Internacional Menendez Pelayo.

Author Index

SUPPORTED BY

Stifterverband für die Deutsche Wissenschaft

ELSEVIER

THOMSON REUTERS™