Published in *Journal of Informetrics* **6** (2012), p. 448– 450, http://dx.doi.org/10.1016/j.joi.2012.01.008

Finding the Easter eggs hidden by oneself: Why Radicchi and Castellano's (2012) fairness test for citation indicators is not fair

Daniel Sirtes, iFQ Berlin, sirtes@forschungsinfo.de

Radicchi and Castellano (2012) suggest a method for assessing the fairness of citation indicators. The idea is simple and quite ingenious. If a set of publications is a representative sample of all publications in certain fields of research, then the top z percent of publications ranked by a fair citation indicator should have the same distribution of fields as the overall sample. For example, if the sample has 20 physics articles and 80 biology articles, then the top 10 should include 2 physics and 8 biology publications.

$$E\left(m_g^{(z)}\right) = n^{(z)} N_g / N$$

The expected number of papers in a category g ranked in the top *z%* ($(m_g^{(z)})$) is equal to the share of top ranked papers ($n^{(z)}$) according to the ratio of papers in this category (N_g) to the overall sample size (N). Although this suggestion rings true theoretically, there is a major weakness when it comes to actually applying it. The expectation is only sound if one can assure that the antecedent clause is met, that is, if one can ascertain that the sample used is actually a representative sample. If the sample is skewed so that the publications in the sample of a certain field are not representative of that field, but biased towards the highly or poorly cited, then the citation indicator should rank more or less of the papers in that field in the top *z%*, respectively.

Moreover, how Radicchi and Castellano (2012) calculate their proposed citation indicator is even more detrimental to the soundness of their results. They use a rescaled citation count. This indicator assesses the citation success by comparing it to the field's average $c_f = c/c_0$ defined as the total number of cites c received by the paper divided by the average of citations c_0 that corresponds to the category and year of publication of the paper (Radicchi, Fortunato, & Castellano, 2008). Thus, whatever the inequality of the samples of the different fields stems from, whether from genuine field-specific differences or from differences in the quality of the samples, the process of rescaling evens them all out. Now the problem becomes clear: As they use the same sample to calculate their average c0 as they use to test the fairness of their indicator, their 'test' becomes circular and begs the question.[1]

That this circularity does indeed lead to false conclusions, especially when used to criticize other indicators, can be shown by comparing their results to another source for assessing their sample, of which they do not more than claim that it "represent[s] an optimal benchmark for the study of citation patterns of publications within physics" (2012, p.122). As it turns out, their sample is not representative after all. They use the American Physics Society's categorization of research fields PACS (http://www.aip.org/pacs/pacs2010/ individuals/pacs2010 regular edition/index.html) and the

[1] For further problems with the proposed rescaling method, such as the non-inclusion of uncited papers or the question as to whether the rescaled indicator is really universal (i.e. applicable in all research fields), see Waltman, van Eck, & van Raan (2012).

American Physics Society's journals Physical Review Letters, Physical Review A, Physical Review B, Physical Review C, Physical Review D and Physical Review E. Using the Journal Citation Report from Thomson Reuters and their categorization one can assess the sample used in the paper (Figs. 1 and 2). By comparing the two figures, one can infer why the simple or fractional citation count (Leydesdorff & Opthof, 2010) seems to be unfair. The reason is that the sample the authors used to calculate the expected share of top articles in particular subject categories varies in quality from field to field. In certain fields, the APS journal is the major publishing venue for the top articles, while in others, it is only second rate. This explains why, for example, the simple citation count or the fractional citation count of 'PACS 30: Atomic and Molecular Physics', which finds its counterpart in JCR's 'Physics, Atomic, Molecular & Chemical', is lower than expected: 24% of all articles in this field are published in journals with a higher Impact Factor; or why the opposite is the case with 'PACS 70: Condensed Matter: Electronic Structure, Electrical, Magnetic, and Optical Properties', where the top 24% of JCR's 'Physics, Condensed Matter' articles are published in the APS journal. Thus, pointing

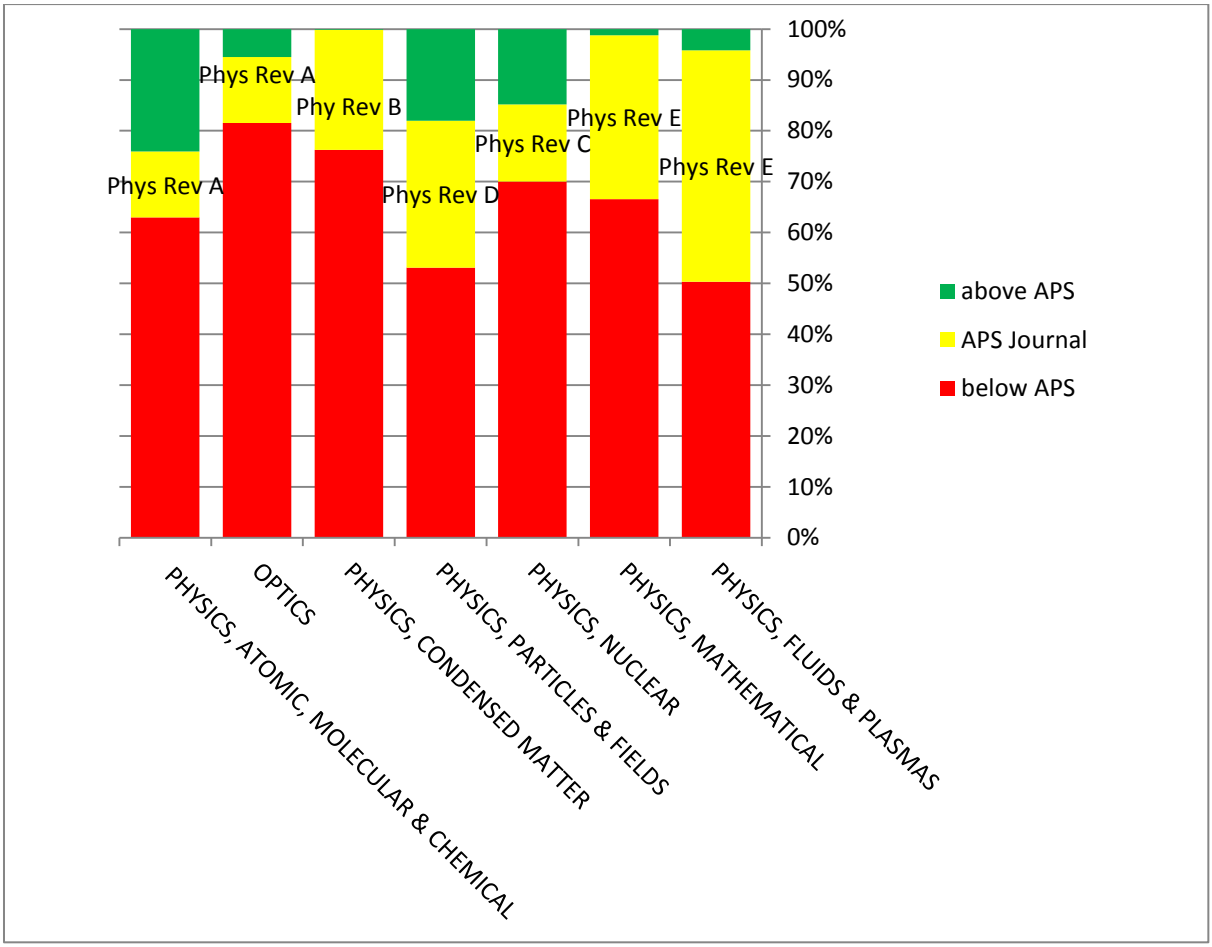

Fig. 1. Percentage of articles published in Web of Science subject categories in journals with lower, higher, and the American Physical Society (APS) journals' Impact Factor. The percentage of articles published in Physical Review A, Physical Review B, Physical Review C, Physical Review D and Physical Review E in their respective subject category ordered by the journals' Impact Factor according to the 2000 Journal Citation Report (Thomson Reuters, 2001).

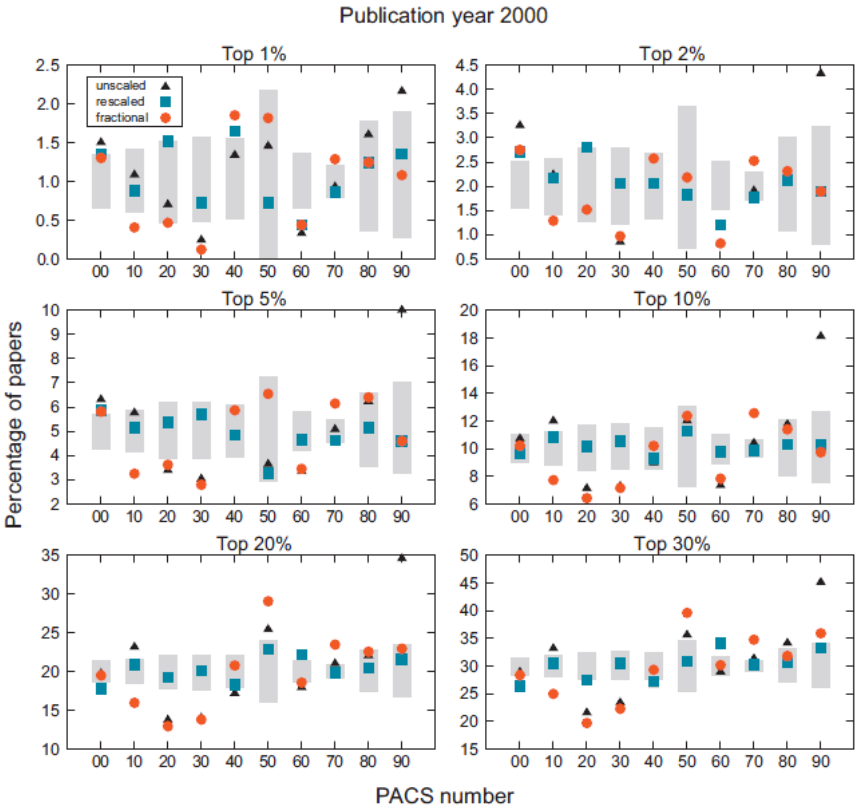

Fig. 2. Percentage of papers belonging to the top *z%* for different PACS classification codes and different values of *z* in the year 2000. Black triangles represent the results obtained with raw citation counts, blue squares stand for the results obtained with the indicator based on rescaled citation counts, while red circles indicate the results obtained using the indicator based on fractional citation counts. Gray areas bound the 90% confidence intervals. (For further details see Fig. 3 of Radicchi and Castellano (2012)) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

out that "[c]ounting citations fractionally is not an effective way to remove biases" is wrongheaded (Radicchi & Castellano, 2012, p.129), as the bias was introduced by their sample itself. Their rescaled citation indicator obfuscates this problem and allows them to declare that their indicator is fair, while their results confirm nothing beyond what was already hidden in their premises. Although the viciousness of the fairness test's circularity could be partly defused by a truly representative random sample of publications in different fields, the fact remains that there is still no good model for understanding and quantifying the multitude of reasons that lead to papers in one research area to be cited more than in others. Nonetheless, I believe Radicchi & Castellano are right to criticize fractional citation counts for not capturing the problems raised by cross-disciplinary comparisons (manuscript in preparation), however how they do it is unfair.

REFERENCES:

Leydesdorff, L., & Opthof, T. (2010). Normalization at the field level: Fractional counting of citations. *Journal of Informetrics*, *4*(4), 644-646. doi:10.1016/j.joi.2010.05.003

Radicchi, F., & Castellano, C. (2012). Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts. *Journal of Informetrics*, *6*(1), 121-130. doi:10.1016/j.joi.2011.09.002

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences*, *105*(45), 17268 -17272. doi:10.1073/pnas.0806977105

Thomson Reuters. (2001). *2000 Journal Citation Reports® Science Edition*.

Waltman, L., van Eck, N. J., & van Raan, A. F. J. (2012). Universality of citation distributions revisited. *Journal of the American Society for Information Science and Technology*, *63*(1), 72-77. doi:10.1002/asi.21671