How (dis-)similar are Different Citation Normalizations and the Fractional Citation Indicator? (And How it can be Improved)

Daniel Sirtes

sirtes@forschungsinfo.de
Institute for Research Information and Quality Assurance, Schützenstr. 6, Berlin, 10117 (Germany)

Introduction
It is conventional wisdom in bibliometrics that raw citation numbers cannot be used to make valid comparisons across different scientific disciplines. The citation densities of different fields of research can diverge to the extent that the same amount of citations of a paper can place it into the top tier of one field, whereas in another field it would be merely average (see Schubert, Glänzel & Braun, 1988).

A solution for this problem is to normalize the citation score. This means that the citation score should be rendered universal by comparing the raw citation score with the average for that kind of paper.

$$C_{normalized_{set}} = \frac{C_{raw}}{\overline{C_{set}}}$$

However, a problem arises: What is the proper mean to which this raw score should be compared? What are the relevant properties of a paper that makes it a member of a certain set of papers whose average should be taken as the normalization basis ($\overline{C_{set}}$)? Should the paper be compared to the papers in its scientific field or to those in the journal in which it was published? Is the document type (article, review, etc.) relevant? Should the publication type (journal, proceedings, etc.) be accounted for? How should different languages be treated?

Issues about the appropriate target set become especially contentious given the following questions:

1. Are the sets for normalization neutral regarding the quality of the papers?
2. How appropriate is the delineation of different target sets?

The first problem can be illustrated by the following example: Can it be assumed that a biochemical paper in German has on average the same quality as an English paper in biochemistry? Although the answer to this question is obviously negative, disqualifying language normalizations in the natural sciences, the ramifications of this problem go even deeper. Couldn't it be argued that the average paper in one *discipline* or *sub-discipline* has a higher average quality than in another? This would cast doubt on any kind of cross-disciplinary fairness test (see Radicchi & Castellano, 2012; and Sirtes, 2012).

The second problem is a constant source of lament in the bibliometrics community. In particular, the delineation of fields of research in the Web of Science database (WoS) subject categories (WoS-SC) has been heavily criticized (see e.g. van Leeuwen, van der Wurff & van Raan, 2001).

A proposed solution to the delineation problem is the 'source-normalized' fractional citation indicator that counts each citation fractionally, that is, as the reciprocal of the total number of references in the citing paper. This method should circumvent the problem of different citation densities in different fields. The idea being that the citation density is a function of the reference count (Leydesdorff & Opthoff, 2010).

We assessed quantitatively the similarities and differences between different normalizations and the fractional citation indicator and arrived at some sobering results.

Methods
We used all Swiss publications in the EFI WoS-SC super-categories 'biology', 'physics', and 'social sciences, economics' from 2002-2006 (24,764 publications). Citations for a three (not shown) and five year citation window were collected and 30 combinations of WoS-SC, EFI27, document type, publication type, language, and journal averages, with or without a restriction to 'citables' (articles, reviews, letters), were calculated as basis for normalization. Additionally, a three and five year fractional citation score was calculated (corrections, as citing document type, were excluded). The papers were then ranked according to their citation scores. Spearman's rank correlation coefficient was then calculated to arrive at the similarities between the different rankings. Finally, on this similarity matrix of the rankings, a multidimensional scaling (MDS) method (PROXSCAL) was performed in order to visualize the proximity of the different rankings.

Results

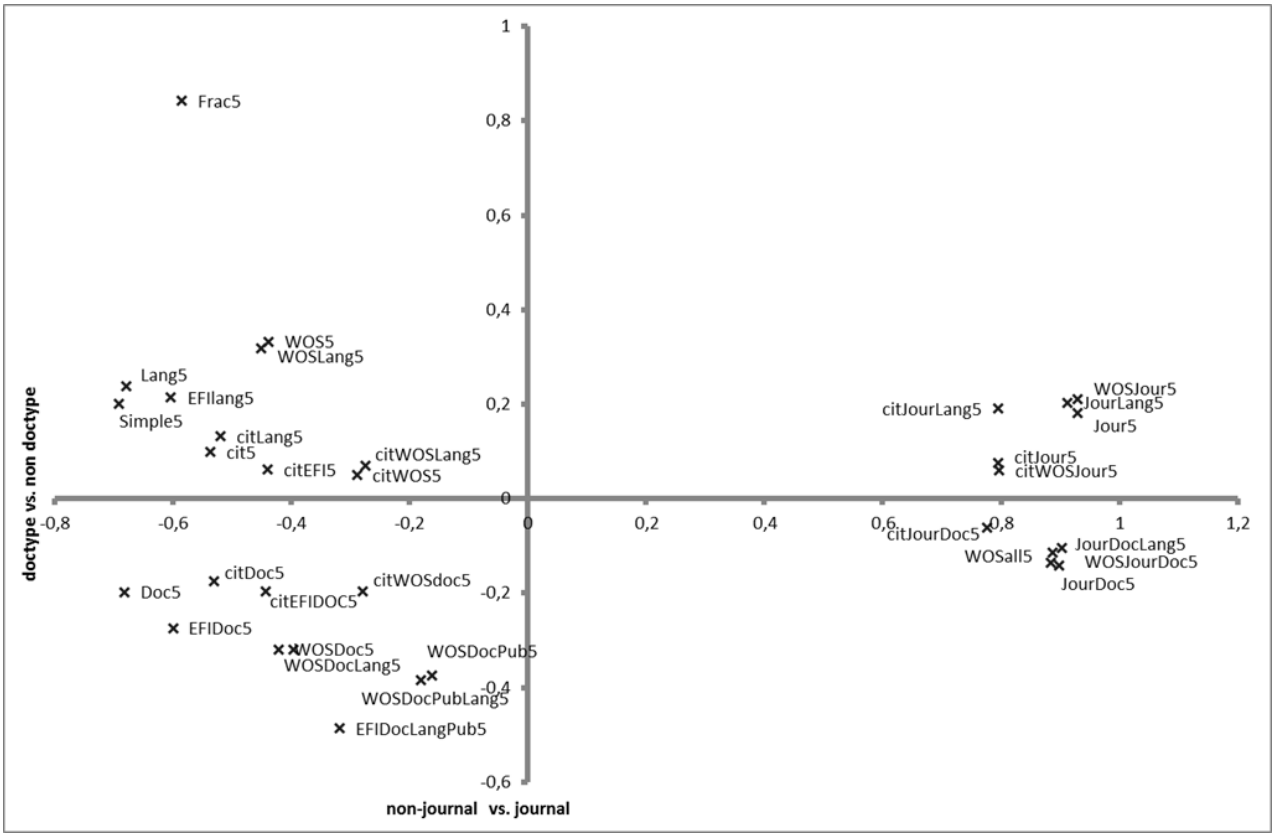

Figure 1. Multidimensional scaling of similarities between different normalized citation score rankings and the fractional citation indicator ranking for the Swiss publication output in 'biology', 'physics' and 'social sciences, economics' 2002-2006 according to the WoS database. The names of the score rankings constructed from the target set for normalizations: 'cit': only 'citables', WOS: WoS SC, EFI: EFI27 super-categories, Doc: document type, Pub: publication type, Jour: journal, Lang: language. Simple: raw citation scores, Frac: fractional citation indicator. 5: 5-year window for citations.

Two important results can be inferred from the MDS visualization (Fig. 1). First, besides the difference between journal and non-journal normalized rankings which form, as expected, the two most distant clusters, the clustering is stronger for document and non-document type normalizations than for any kind of field normalization or non-normalized rankings (Simple5/cit5). Second, it is clear that the rankings resulting from fractional scores are quite different from any of the other normalizations (Spearman's rank correlation coefficients to non-journal normalized rankings are 0.805-0.901, and to journal normalized are 0.722-0.777. In comparison, citWOSDoc5 coefficients are 0.813-0.841 to journal normalized and 0.930-0.996 to non-journal normalized rankings).

In order to correct for the document type specific reference count variances (for 2006: Median reference count for letters 6, articles 22, reviews 86) new indicators building on the fractional citation indicator were developed and tested: First, the source document type reference count normalized fractional citation indicators

(1) $C_{frac_{SA}} = \sum_{i=1}^{n} \frac{\overline{rc_D}}{rc_i}$

(2) $C_{frac_{SM}} = \sum_{i=1}^{n} \frac{\widetilde{rc_D}}{rc_i}$

where each fractional citation $\left(\frac{1}{rc_i}\right)$ is normalized according to its document type reference count mean $\left(\frac{1}{\overline{rc_D}}\right)$ or median $\left(\frac{1}{\widetilde{rc_D}}\right)$ of the published year, and second, source *and* target document type normalized fractional citation indicators

(3) $C_{frac_{DNA}} = \frac{C_{frac_{SA}}}{\left(C_{frac_{SA}}\right)_D}$

(4) $C_{frac_{DNM}} = \frac{C_{frac_{SM}}}{\left(C_{frac_{SM}}\right)_D}$

where the source document type normalized score $\left(C_{frac_{SA/M}}\right)$ were again normalized with the target document type average score $\left(\overline{\left(C_{frac_{SA/M}}\right)_D}\right)$.

Table 1. Spearman's rank correlation coefficients between different (non-journal-normalized) rankings (see caption of Fig. 1) of four new indicators. FracDocS are reference count source normalized according to mean (Avg) or median (Med) reference counts of the citing document type of the year. FracDocDN are double normalized on source reference count and on the average target document type scores themselves.

Indicators	Frac5	Frac DocS Med5	Frac DocS Avg5	Frac Doc DN Med5	Frac Doc DN Avg5
WOS5	.874	.889	.884	.842	.840
EFIlang5	.891	.911	.905	.862	.859
Lang5	.891	.914	.908	.866	.861
Simple5	.895	.919	.913	.866	.866
citWOS5	.886	.900	.900	.862	.869
citLang5	.890	.913	.913	.873	.878
cit5	.894	.919	.919	.873	.884
citEFI5	.901	.919	.919	.881	.887
WOSDoc5	.843	.851	.851	.878	.889
citWOSdoc 5	.865	.876	.876	.890	.896
EFIDoc5	.858	.866	.868	.903	.913
Doc5	.855	.869	.869	.906	.914
citDoc5	.872	.890	.890	.912	.916
citEFIDOC 5	.878	.892	.892	.912	.917

As can be seen in Table 1, the double normalized fractional citation indicator rankings have a greater similarity to the document type normalized indicators, with the highest similarity of FracDNAvg5 to citEFIDoc5 with a Spearman coefficient of 0.917. By contrast, the source normalized fractional indicators are more similar to the non-document type normalized indicators, with the highest similarity of both FracDocSs to citEFI5 and the raw 'citables' citation score with a Spearman coefficient of 0.919. All four new indicators are more similar to the previously calculated indicators than the simple fractional indicator is.

Discussion

The stronger clustering of document type vs. non-document type normalizations to field vs. non-field normalizations is somewhat surprising as the main discussions in the literature have been about the problems of field normalizations and the fractional citation indicator was introduced in order to circumvent the problems of field delineation. Thus, even given the problems of document type assignment (see Harzing, 2012), it is reasonable to argue that the variation of document type citing behavior does not reflect a difference in quality between these types, and consequently a good citation indicator should account for it.

The patent dissimilarity between the fractional citation indicator and normalized indicator results is readily explainable by the fact that the differences in document type reference counts on the citing side are so large. The document type average reference counts mentioned above lead to the result that, for example, a citation by a letter has more than fourteen times more value than one by a review. This eclipses any differences in field specific reference counts and so cannot be justified. In order to fix this problem, new fractional indicators were constructed that normalize either only on the citing document type's average reference count or in addition on the average cited document type's citation scores. Although these new indicators clearly improve on the original fractional indicator by coming closer to normalized indicators without falling prey to the problems of field delineation, more work needs to be done on finding ways to iron out the artifacts produced by fractional citation counting.

Acknowledgments

The author would like to thank Jörg Neufeld, Jian Wang, Jana Schütze, Christian Klode, Thed van Leeuwen, Haiko Lietz, Marion Schmidt and Eric Oberheim for very helpful discussions and suggestions.

References

Harzing, A.-W. (2012). Document Categories in the ISI Web of Knowledge: Misunderstanding the Social Sciences? *Scientometrics*, http://dx.doi.org/10.1007/s11192-012-0738-1

Leydesdorff, L. & Opthof, T. (2010). Normalization at the field Level: Fractional Counting of Citations. *Journal of Informetrics*, 4 (4), 644-646.

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.

Schubert, A., Glänzel, W. & Braun, T. (1988). Against Absolute Methods: Relative Scientometric Indicators and Relational Charts as Evaluation Tools. In Van Raan, A.F.J. (Ed.), *Handbook of quantitative studies of science and technology* (pp. 137-169). Amsterdam: Elsevier.

Sirtes, D. (2012). Finding the Easter Eggs Hidden by Oneself: Why Radicchi & Castellano's (2012) Fairness Test for Citation Indicators is not fair. *Journal of Informetrics*, 6 (3), 448-450.

van Leeuwen, T.N., van der Wurff, L.J. & van Raan, A.F.J. (2001). The Use of Combined bibliometric-methods in Research Funding Policy. *Research Evaluation*, 10 (3), 195-201.