Journal of Informetrics 8 (2014) 419–420

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Letter to the Editor

On the use of sampling statistics to advance bibliometrics

Bornmann & Leydesdorff's letter (Bornmann and Leydesdorff, 2014) questions the validity of the findings of our study on the universality of scholarly impact metrics (Kaur, Radicchi, & Menczer, 2013). The main contribution of our paper is a formal methodology to measure discipline bias in scholar impact metrics. This contribution, which is not addressed in Bornmann & Leydesdorff's letter, extends prior work on discipline bias and normalization of paper citations (Castellano & Radicchi, 2009; Radicchi & Castellano, 2011, 2012; Radicchi, Fortunato, & Castellano, 2008).

Our second contribution is the formulation of a simple discipline normalization of h, yielding a "universal" metric h_s. The criticism of Bornmann & Leydesdorff in this respect is that we compute the average h without taking academic age into account in an explicit way. There is no reason to suspect a dependence of average scholar age on discipline, in general or in any bibliographic database (Radicchi & Castellano, 2013). But let us assume such a bias existed in the database used in our analysis; suppose scholars from field A were on average more senior compared to scholars from field B. Therefore the average h of field A would be higher, and our normalization would neutralize that bias. If a metric favored a discipline or another due to a dependence of scholar age on the field, then our universality measure would detect such a bias, and yield a low value of universality. Our paper shows this not to be the case. That is exactly our main point – that we can measure such a bias. Therefore it does not seem necessary, statistically speaking, to apply more complex normalization schemes based on academic age, as suggested by Bornmann & Leydesdorff.

That said, one could of course use our proposed methodology to compare the discipline bias of scholar impact metrics that take age into account and those that do not. We invite our colleagues to perform such an analysis and share the results with the community.

The third contribution of our paper is an empirical evaluation of the universality of various metrics from the literature. This is where Bornmann & Leydesdorff focus their strongest criticism, arguing forcefully that the only way to normalize a metric is to use a reference set that includes all scientists.

As discussed extensively by Hoang, Kaur, and Menczer (2010), Kaur et al. (2012) and Sun, Kaur, Possamai, and Menczer (2013), the Scholarometer database used to carry out the empirical analysis in our study is based on crowdsourced data. Like all sources of bibliometric data, this too has strengths and weaknesses. Our paper explicitly mentions such caveats. Our sample is smaller than desirable, especially for certain disciplines. On the other hand, crowdsourced annotations can be extremely effective (Hoang et al., 2010). Indeed, one is able to capture up-to-date discipline annotations at a granularity that is not available in other bibliographic sources (Kaur et al., 2012), as well as links between scholars and disciplines that are not available in any other dataset (Sun, Kaur, Milojevic, Flammini, & Menczer, 2013). We will be delighted if our colleagues would be able to repeat our analysis with other sources, such as WoS or Scopus. While those datasets do not cover all scholars, they are certainly larger. On the other hand, their field classifications are very outdated and based on annotations of journals rather than authors (Pudovkin & Garfield, 2002). In any case, alternative empirical evaluations confirming or disproving our universality results will advance our understanding of impact metrics more than criticism alone.

The final criticism of Bornmann & Leydesdorff is about the quality of Google Scholar as a source of bibliographic data. While our methodology is not dependent on any one bibliographic source, plenty of studies have found that Google Scholar has good coverage of citations in most fields and better coverage than WoS and Scopus in some fields, such as computer science and social sciences (Harzing, 2014; Radicchi & Castellano, 2013; Sun, Kaur, Possamai, et al., 2013). Some studies cited by Bornmann & Leydesdorff arguing otherwise appear to be based on anecdotal rather than systematic analyses. That said, we acknowledge that Google Scholar too has its limitations; most crippling in our view is the closed nature of the data. We hope that our colleagues will be able to extend our analysis to other data sources.

In closing, we take strong exception with Bornmann & Leydesdorff's claim that the only way to normalize a metric is to use reference sets that include all scientists. While we cannot claim that the sample of scholars from any discipline in any given database is representative, the criticism seems to ignore elementary sampling statistics. To use a crude analogy, where would science be today if Newton had to wait for all apples to fall on his head?

1751-1577/$ – see front matter
http://dx.doi.org/10.1016/j.joi.2014.01.010

References

Bornmann, L., & Leydesdorff, L. (2014). On the meaningful and non-meaningful use of reference sets in bibliometrics. *Journal of Informetrics*, *8*, 273–275.
Castellano, C., & Radicchi, F. (2009). On the fairness of using relative indicators for comparing citation performance in different disciplines. *Archivum immunologiae et therapiae experimentalis*, *57*, 85–90.
Harzing, A. (2014). *Publish or Perish*. http://harzing.com/pop.htm
Hoang, D. T., Kaur, J., & Menczer, F. (2010). Crowdsourcing scholarly data. In *Proceedings of the Web Science Conference: Extending the Frontiers of Society On-Line (WebSci)* http://journal.webscience.org/321/
Kaur, J., Hoang, D. T., Sun, X., Possamai, L., JafariAsbagh, M., Patil, S., Menczer, F., et al. (2012). Scholarometer: a social framework for analyzing impact across disciplines. *PLoS ONE*, *7*, e43235.
Kaur, J., Radicchi, F., & Menczer, F. (2013). Universality of scholarly impact metrics. *Journal of Informetrics*, *7*, 924–932.
Pudovkin, A. I., & Garfield, E. (2002). Algorithmic procedure for finding semantically related journals. *Journal of the American Society for Information Science and Technology*, *53*, 1113–1119.
Radicchi, F., & Castellano, C. (2011). Rescaling citations of publications in physics. *Physical Review E*, *83*, 046116.
Radicchi, F., & Castellano, C. (2012). A reverse engineering approach to the suppression of citation biases reveals universal properties of citation distributions. *PLoS ONE*, *7*, e33833.
Radicchi, F., & Castellano, C. (2013). Analysis of bibliometric indicators for individual scholars in a large data set. *Scientometrics*, *97*, 627–637.
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*, 17268–17272.
Sun, X., Kaur, J., Milojevic, S., Flammini, A., & Menczer, F. (2013). Social dynamics of science. *Scientific Reports*, *3*.
Sun, X., Kaur, J., Possamai, L., & Menczer, F. (2013). Ambiguous author query detection using crowdsourced digital library annotations. *Information Processing and Management*, *49*, 454–464.

Jasleen Kaur*
Filippo Radicchi
Filippo Menczer
Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University, Bloomington, USA

* Corresponding author.
E-mail address: jakaur@umail.iu.edu (J. Kaur)

20 January 2014

Available online 5 March 2014