Journal of Informetrics 7 (2013) 933–944

Contents lists available at ScienceDirect

Journal of Informetrics

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

Which percentile-based approach should be preferred for calculating normalized citation impact values? An empirical comparison of five approaches including a newly developed citation-rank approach (P100)

Lutz Bornmann[a,*], Loet Leydesdorff[b], Jian Wang[c,d]

[a] Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany
[b] Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Kloveniersburgwal 48, 1012 CX Amsterdam, The Netherlands
[c] Institute for Research Information and Quality Assurance (iFQ), Schützenstraße 6a, 10117 Berlin, Germany
[d] Center for R&D Monitoring (ECOOM), Department of Managerial Economics, Strategy and Innovation, Katholieke Universiteit Leuven, Waaistraat 6, 3000 Leuven, Belgium

ARTICLE INFO

Article history:
Received 19 June 2013
Received in revised form
17 September 2013
Accepted 17 September 2013

Keywords:
Citation impact normalization
Percentile
Percentile rank class
P100
Citation rank

ABSTRACT

For comparisons of citation impacts across fields and over time, bibliometricians normalize the observed citation counts with reference to an expected citation value. Percentile-based approaches have been proposed as a non-parametric alternative to parametric central-tendency statistics. Percentiles are based on an ordered set of citation counts in a reference set, whereby the fraction of papers at or below the citation counts of a focal paper is used as an indicator for its relative citation impact in the set. In this study, we pursue two related objectives: (1) although different percentile-based approaches have been developed, an approach is hitherto missing that satisfies a number of criteria such as scaling of the percentile ranks from zero (all other papers perform better) to 100 (all other papers perform worse), and solving the problem with tied citation ranks unambiguously. We introduce a new citation-rank approach having these properties, namely P100; (2) we compare the reliability of P100 empirically with other percentile-based approaches, such as the approaches developed by the SCImago group, the Centre for Science and Technology Studies (CWTS), and Thomson Reuters (InCites), using all papers published in 1980 in Thomson Reuters Web of Science (WoS). How accurately can the different approaches predict the long-term citation impact in 2010 (in year 31) using citation impact measured in previous time windows (years 1–30)? The comparison of the approaches shows that the method used by InCites overestimates citation impact (because of using the highest percentile rank when papers are assigned to more than a single subject category) whereas the SCImago indicator shows higher power in predicting the long-term citation impact on the basis of citation rates in early years. Since the results show a disadvantage in this predictive ability for P100 against the other approaches, there is still room for further improvements.

* Corresponding author. Tel.: +49 89 2108 1265.
E-mail addresses: bornmann@gv.mpg.de (L. Bornmann), loet@leydesdorff.net (L. Leydesdorff), Jian.Wang@kuleuven.be (J. Wang).

1751-1577/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2013.09.003

1. Introduction

For comparisons of citation impacts across fields and over time, bibliometricians normalize the observed citation counts with reference to an expected citation value. For many years, this expected citation value was commonly calculated as the (arithmetic) average citations of the papers in a reference set; for example, the citation counts of all papers published in the same subject category and year as the paper(s) under study can be averaged. However, the arithmetic average of a citation distribution hardly provides an appropriate baseline for comparison because these distributions are extremely skewed (Seglen, 1992): the average is heavily affected by a few highly cited papers (Waltman et al., 2012).

Percentile-based approaches (quantiles, percentiles, percentile ranks, or percentile rank classes) have been proposed as a non-parametric alternative to these parametric central-tendency statistics. Percentiles are based on an ordered set of citation counts (i.e., papers are sorted in ascending order of citation counts), whereby the fraction of papers at or below the citation counts of a focal paper is used as an indicator for the relative citation impact of this focal paper. Instead of using an average citation impact for normalizing a paper under study, its citation impact is evaluated by its rank in the citation distribution of similar papers in a reference set (Leydesdorff & Bornmann, 2011; Pudovkin & Garfield, 2009).

This percentile-based approach arose from a debate in which we argued that frequently used citation impact indicators based on using arithmetic averages for the normalization—e.g., "relative citation rates" (Glänzel, Thijs, Schubert, & Debackere, 2009; Schubert & Braun, 1986) and "crown indicators" (Moed, De Bruin, & Van Leeuwen, 1995; van Raan, van Leeuwen, Visser, van Eck, & Waltman, 2010)—had been both technically (Lundberg, 2007; Opthof & Leydesdorff, 2010) and conceptually (Bornmann & Mutz, 2011) flawed. The non-parametric statistics for testing observed versus expected citation distributions were further elaborated by Leydesdorff, Bornmann, Mutz, and Opthof (2011). Various statistical procedures have been proposed to analyze percentile citations for institutional publication sets (Bornmann, 2013a; Bornmann & Marx, in press; Bornmann & Williams, in press).

In this study, we pursue two related objectives—an analytical and an empirical one: (1) although different percentile-based approaches have been developed (see an overview in Waltman & Schreiber, 2013), an approach is missing that satisfies, in our opinion, a number of important criteria such as scaling from the percentile rank 0 (all other papers perform better) to 100 (all other papers perform worse), solving the problem with tied citation ranks (papers with the same number of citations in a reference set) unambiguously, ensuring that the mean percentile value is 50, and symmetrically handling the tails of the distributions. Bibliometricians hitherto use different approaches, but unfortunately no approach for the computation of percentiles in the case of discrete distributions is without disadvantages (Hyndman & Fan, 1996).

In the following, we propose a new citation-rank approach which, in our opinion, solves the two major problems analytically: (i) unambiguous scaling from 0 to 100, and (ii) equal values for tied ranks. In order to achieve this objective, we use a distribution other than the citation distributions, namely the distribution of *unique* citation values. On this basis, papers with tied citations obtain necessarily the same rank. Furthermore, it seems to us that the scale can run from 0 to 100 with the highest-ranking paper in a reference set at 100 and the lowest at 0. In other words, we derive the ranks from the empirical distribution of the unique citation counts in the reference set. By definition, the highest-ranking papers in different reference sets all obtain the highest rank of 100, the lowest ones are fixed at zero, and thus, the distributions are made comparable. By defining the citation ranks empirically as intervals between the two extremes (of 0 and 100), one gains a degree of freedom which enables us to solve the problem of the otherwise floating maximum or minimum values of percentiles in discrete distributions. We propose to use the abbreviation "P100" for our new approach.

(2) In the second part of this study we compare the reliability of our new citation-rank approach with percentile-based approaches, among which are the approaches developed by the SCImago group for the ranking of universities and research-focused institutions and by the Centre for Science and Technology Studies (CWTS) for the ranking of universities. Using all papers in Web of Science (WoS, Thomson Reuters) published in 1980 (Wang, 2013), we retrieve for each subsequent year the citation counts of all these papers. The citation impact is then calculated based on the different percentile-based approaches and P100. The results show differences among the approaches in assigning the papers to various rank classes (e.g., the 10% most frequently cited papers in a certain subject category) and in the ability of estimating the long-term citation impact of the papers (their cumulative citation impact in the final [31st] year (2010)). We are most interested in the degree of agreement between the cumulative citation impact in the final year and the impact in early years (especially the first few years after publication).

There is no reason to assume that the first three to five years of citation can be used as a predictor of the long-term citation impact of papers. Wang (2013) found low correlations between citation counts in short time windows and total citations counts in 31 years, and these correlations are even lower for field-normalized citation counts and highly cited papers. By applying Group-based Trajectory Modelling (Nagin, 2005) to citation distributions of journals, Baumgartner and Leydesdorff (in press) showed that papers within an otherwise similar set can vary significantly in terms of how they are cited in the longer run of ten or fifteen years.

Given the long time-window of 31 years in this study, one can assume that we capture the entire citation impact in the long run (in other words, with the long run we will have a valid estimate of a paper's "true" citation impact). The analytical argument about the rules for defining percentiles or citation ranks, however, is a different one from the empirical usability of one or another (percentile-based) approach as a predictor in evaluation studies. The latter is a correlation which does not imply a causal explanation for the preference of the one or the other approach.