A systematic empirical comparison of different approaches for normalizing citation impact indicators

CrossMark

Ludo Waltman *, Nees Jan van Eck

Centre for Science and Technology Studies, Leiden University, The Netherlands

ABSTRACT

We address the question how citation-based bibliometric indicators can best be normalized to ensure fair comparisons between publications from different scientific fields and different years. In a systematic large-scale empirical analysis, we compare a traditional normalization approach based on a field classification system with three source normalization approaches. We pay special attention to the selection of the publications included in the analysis. Publications in national scientific journals, popular scientific magazines, and trade magazines are not included. Unlike earlier studies, we use algorithmically constructed classification systems to evaluate the different normalization approaches. Our analysis shows that a source normalization approach based on the recently introduced idea of fractional citation counting does not perform well. Two other source normalization approaches generally outperform the classification-system-based normalization approach that we study. Our analysis therefore offers considerable support for the use of source-normalized bibliometric indicators.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Citation-based bibliometric indicators have become a more and more popular tool for research assessment purposes. In practice, there often turns out to be a need to use these indicators not only for comparing researchers, research groups, departments, or journals active in the same scientific field or subfield but also for making comparisons across fields (Schubert & Braun, 1996). Performing between-field comparisons is a delicate issue. Each field has its own publication, citation, and authorship practices, making it difficult to ensure the fairness of between-field comparisons. In some fields, researchers tend to publish a lot, often as part of larger collaborative teams. In other fields, collaboration takes place only at relatively small scales, usually involving no more than a few researchers, and the average publication output per researcher is significantly lower. Also, in some fields, publications tend to have long reference lists, with many references to recent work. In other fields, reference lists may be much shorter, or they may point mainly to older work. In the latter fields, publications on average will receive only a relatively small number of citations, while in the former fields, the average number of citations per publication will be much larger.

In this paper, we address the question how citation-based bibliometric indicators can best be normalized to correct for differences in citation practices between scientific fields. Hence, we aim to find out how citation impact can be measured in a way that allows for the fairest between-field comparisons.

In recent years, a significant amount of attention has been paid to the problem of normalizing citation-based bibliometric indicators. Basically, two streams of research can be distinguished in the literature. One stream of research is concerned

---

* Corresponding author. Tel.: +31 715275806.
  E-mail addresses: waltmanlr@cwts.leidenuniv.nl (L. Waltman), ecknjpvan@cwts.leidenuniv.nl (N.J. van Eck).

with normalization approaches that use a field classification system to correct for differences in citation practices between scientific fields. In these normalization approaches, each publication is assigned to one or more fields and the citation impact of a publication is normalized by comparing it with the field average. Research into classification-system-based normalization approaches started in the late 1980s and the early 1990s (e.g., Braun & Glänzel, 1990; Moed, De Bruin, & Van Leeuwen, 1995). Recent contributions to this line of research were made by, among others, Abramo, Cicero, and D'Angelo (2012), Crespo, Li, Herranz, and Ruiz-Castillo (in press), Crespo, Li, and Ruiz-Castillo (2013), Li, Radicchi, Castellano, and Ruiz-Castillo (2013), Radicchi and Castellano (2012c), Radicchi, Fortunato, and Castellano (2008), and Van Eck, Waltman, Van Raan, Klautz, and Peul (2013).

The second stream of research studies normalization approaches that correct for differences in citation practices between fields based on the referencing behavior of citing publications or citing journals. These normalization approaches do not use a field classification system. The second stream of research was initiated by Zitt and Small (2008),[1] who introduced the audience factor, an interesting new indicator of the citation impact of scientific journals. Other contributions to this stream of research were made by Glänzel, Schubert, Thijs, and Debackere (2011), Leydesdorff and Bornmann (2011), Leydesdorff and Opthof (2010), Leydesdorff, Zhou, and Bornmann (2013), Moed (2010), Waltman and Van Eck (2013), Waltman, Van Eck, Van Leeuwen, and Visser (2013), Zhou and Leydesdorff (2011), and Zitt (2010, 2011). Zitt and Small referred to their proposed normalization approach as 'fractional citation weighting' or 'citing-side normalization'. Alternative labels introduced by other authors include 'source normalization' (Moed, 2010), 'fractional counting of citations' (Leydesdorff & Opthof, 2010), and 'a priori normalization' (Glänzel et al., 2011). Following our earlier work (Waltman & Van Eck, 2013; Waltman et al., 2013), we will use the term 'source normalization' in this paper.

Which normalization approach performs best is still an open issue. Systematic large-scale empirical comparisons of normalization approaches are scarce, and as we will see, such comparisons involve significant methodological challenges. Studies in which normalization approaches based on a field classification system are compared with source normalization approaches have been reported by Leydesdorff, Radicchi, Bornmann, Castellano, and De Nooy (in press) and Radicchi and Castellano (2012a). In these studies, classification-system-based normalization approaches were found to be more accurate than source normalization approaches. However, as we will point out later on in this paper, these studies have important methodological limitations. In an earlier paper, we have compared a classification-system-based normalization approach with a number of source normalization approaches (Waltman & Van Eck, 2013). The comparison was performed in the context of assessing the citation impact of scientific journals, and the results seemed to be in favor of some of the source normalization approaches. However, because of the somewhat non-systematic character of the comparison, the results must be considered of a tentative nature.

Building on our earlier work (Waltman & Van Eck, 2013), we present in this paper a systematic large-scale empirical comparison of normalization approaches. The comparison involves one normalization approach based on a field classification system and three source normalization approaches. In the classification-system-based normalization approach, publications are classified into fields based on the journal subject categories in the Web of Science bibliographic database. The source normalization approaches that we consider are based on the audience factor approach of Zitt and Small (2008), the fractional citation counting approach of Leydesdorff and Opthof (2010), and our own revised SNIP approach (Waltman et al., 2013).

Our methodology for comparing normalization approaches has three important features not present in earlier work by other authors. First, rather than simply including all publications available in a bibliographic database in a given time period, we exclude as much as possible publications that could distort the analysis, such as publications in national scientific journals, popular scientific magazines, and trade magazines. Second, in the evaluation of the classification-system-based normalization approach, we use field classification systems that are different from the classification system used in the implementation of the normalization approach. In this way, we ensure that our results do not suffer from a bias that favors classification-system-based normalization approaches over source normalization approaches. Third, we compare normalization approaches at different levels of granularity, for instance both at the level of broad scientific disciplines and at the level of smaller scientific subfields. As we will see, some normalization approaches perform well at one level but not so well at another level.

To compare the different normalization approaches, our methodology uses a number of algorithmically constructed field classification systems. In these classification systems, publications are assigned to fields based on citation patterns. The classification systems are constructed using a methodology that we have introduced in an earlier paper (Waltman & Van Eck, 2012). Some other elements that we use in our methodology for comparing normalization approaches have been taken from the work of Crespo et al. (2013, in press).

The rest of this paper is organized as follows. In Section 2, we discuss the data that we use in our analysis. In Section 3, we introduce the normalization approaches that we study. We present the results of our analysis in Section 4, and we summarize our conclusions in Section 5. The paper has four appendices. In Appendix A, we discuss the approach that we take to select core journals in the Web of Science database. In Appendix B, we discuss our methodology for algorithmically constructing field classification systems. In Appendix C, we provide an example illustrating a methodological problem in

---

[1] Some first suggestions in the direction of this second stream of research were already made by Zitt, Ramanana-Rahary, & Bassecoulard (2005).