Journal of Informetrics 12 (2018) 1133–1145

Contents lists available at ScienceDirect

Journal of Informetrics

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

Regular article

Four pitfalls in normalizing citation indicators: An investigation of ESI's selection of highly cited papers

Zhigang Hu[a], Wencan Tian[a], Shenmeng Xu[b], Chunbo Zhang[a], Xianwen Wang[a,*]

[a] WISE Lab, Institute of Science of Science and S&T Management, Dalian University of Technology, Dalian, 116024, China
[b] School of Information and Library Science, University of North Carolina at Chapel Hill, Chapel Hill, 27517, USA

ARTICLE INFO

Article history:
Received 8 April 2018
Received in revised form
10 September 2018
Accepted 10 September 2018

Keywords:
Normalized citation impact
Essential science indicators
Highly cited papers
Publication month
Online-to-print delays
Citation impact indicators
Journal-to-field classification

ABSTRACT

InCites Essential Science Indicators is becoming increasingly used to identify top-performing research and evaluate the impact of institutes. Unfortunately, our study shows that ESI indicators, as well as other normalized citation indicators, have the following flaws. First, the publication month and the online-to-print delay affect a paper's probability of becoming a Highly Cited Paper (HCP). Papers published in the earlier months of the year are more likely to accumulate enough citation counts to rank at the top 1% compared with those published in later months of the year. Papers with longer online-to-print delays have an apparent advantage for being selected as HCPs. Research field normalizations lead to the third pitfall. Different research fields have different citation thresholds for HCPs, making research field classification important for a journal. In addition, the uniform thresholds for both articles and reviews in ESI affect the reliability of HCP selection because, on average, reviews tend to have higher citation rates than articles. ESI's selection of HCPs provides an intuitive feel for the problems of normalized citation impact indicators, such as those provided in InCites and SciVal.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Citation data used as performance indicators are valid only after they have been normalized properly (Waltman, 2016a). It does not make sense to compare citation frequency between two papers published in different years or in different fields. Various previous works have proven that a paper's citation count is sensitive to citation time windows, publication types, and research areas (e.g., Colliander & Ahlgren, 2011; Didegah & Thelwall, 2013; Levitt & Thelwall, 2011; Thelwall & Fairclough, 2015; Thelwall & Wilson, 2014; Waltman, 2016a, 2016b; Waltman & van Eck, 2013; Wang, 2013). Thus, normalization, either by weighting papers' citation counts on the basis of citation windows or according to scientific subject fields, is essential for ensuring a fair comparison.

However, a number of studies have questioned whether the current normalization methods based on publication dates or research fields are adequate. First, traditional citation impact indicators are usually normalized for a publication year. Levitt and Thelwall (2011) argue that this normalization mechanism is problematic because it gives publications that appear in the early months of a year a significant advantage over those that appear in the last months. Donner (2018) also found

* Corresponding authors.
E-mail address: xianwenwang@dlut.edu.cn (X. Wang).

https://doi.org/10.1016/j.joi.2018.09.006
1751-1577/© 2018 Elsevier Ltd. All rights reserved.

that early-in-the-year publications have a substantial citation advantage that persists in terms of citations received in the first three years after publication. A paper by Haustein, Bowman, and Costas, (2015) raised an issue related to the changing definition of publication dates. In the digital era, papers are increasingly published online first and are citable before they have been published in a journal issue. The online date marks the time when articles are first made publicly available on the publishers' websites. Examples include the Springer's "Online First", Wiley-Blackwell's "Early View", and Nature Publishing Group's "Advance Online Publication". When, then, is an article actually published? Which date should be defined as the beginning of the citation window? Is it the date of online availability or print publication? At present, the Web of Science (WoS) and Scopus databases still record only the date of the journal issue and ignore the paper's online publication date. It has also been suggested that preprint availability could inflate the journal impact factor (Al & Soydal, 2017; Amaral & Tort, 2012; Echeverria, Stuart, & Cordon-Garcia, 2017; Yu, Wang, & Yu, 2005).

Field normalization of citations is another issue that deserves discussion (Herranz & Ruiz-Castillo, 2012; Li & Ruiz-Castillo, 2013; Opthof & Leydesdorff, 2010; Ruiz-Castillo, 2014). In practice, normalization of citation impact indicators is usually based on the Web of Science or Scopus subject categories system. Some scholars have also constructed subject classification systems for their own purpose (Ruiz-Castillo & Waltman, 2015; Waltman & van Eck, 2012; Waltman, Eck, & Van, 2018), including the well-known field-normalized citation impact indicators SNIP (Moed, 2010) and the crown indicator (Waltman, van Eck, van Leeuwen, Visser, & van Raan, 2011). However, it is questionable whether these subject categories are sufficiently accurate and homogeneous in terms of field normalization. For instance, Leydesdorff and Bornmann (2016) pointed out that WoS subject categories have too broad a scope to be suitable for normalizing citation indicators because this categorization method was originally developed for information retrieval and not citation analysis.

Unfortunately, all of these pitfalls in normalizing citation impact indicators are sometimes ignored in the citation-based systems InCites and SciVal. In this study, we will examine InCites' Essential Science Indicators (ESI) and the extent to which these potential pitfalls can cause errors and omissions.

2. Four Pitfalls in ESI's selection of Highly Cited Papers

InCites Essential Science Indicators (ESI), a publication-and-citation-based research analytic tool provided by Clarivate Analytics for identifying top-performing research in WoS-indexed items, is sometimes used to evaluate the impact of countries (Csajbók, Berhidi, Vasas, & Schubert, 2007; Fu, Chuang, Wang, & Ho, 2011), institutes (Chuang, Wang, & Ho, 2011; Ma, Ni, & Qiu, 2008) and scientists (Harzing, 2015). ESI Highly Cited Papers (HCPs) and Hot Papers (HPs), which represent the most influential research articles in one of 22 research fields, are two fundamental components of ESI. Specifically, Highly Cited Papers represent the papers in the top 1% of papers cited in their field and publication year.

Although widely used in research evaluation and literature retrieval, ESI's reliability in selecting Highly Cited Papers has rarely been examined in previous works. ESI normalizes citation counts according to the publication year and research fields. In this way, a Highly Cited Paper needs only to surpass its counterparts in the same cohort to rank higher in the research field. However, four pitfalls exist in ESI's selection of HCPs.

First, comparing the citation counts of papers published in different months of the same year introduces bias. With the acceleration of scholarly communication, the publication year is no longer a sufficient unit of time aggregation for bibliometrics and social media indicators. Papers published in January have an almost one-year longer citation window than those published in December. For example, as of today (April 2018), papers published in January 2017 have a 15-month citation window, while those published in December 2017 only have 4 months to accumulate their citations for the year. However, all papers published in all the months of 2017 will be compared according to ESI's method for selecting HCPs.

Second, ESI considers a paper's publication date the beginning of its citation window; however, this is not accurate for many papers. The online date can be a more meaningful for citation-based indicators than the publication date. Additionally, a paper could benefit from the time difference between its online date (the beginning of its citability) and its print date (the beginning of the citation window considered by ESI) because being published online first gives the paper extra time to accumulate citations. For example, if a paper was issued in 2017 but was online in 2015, it would have a longer citation window than papers that appeared both online and in print in 2017. Thus, the former paper has a higher chance of being selected as an HCP.

Third, ESI's research field classification has a considerable influence on the selection of HCPs. In ESI, the research category schema works at the journal level. More than 11,000 journals are assigned to only 22 research fields. This is a broader classification system than the WoS classification system, which consists of 251 subject categories, or the Scopus system, which consists of 334 subject areas. Article-level classifications are usually finer-grained than journal-level classifications. Ruiz-Castillo and Waltman (2015) proposed an article-level classification system with 5119 research fields (clusters) and argued that working with a few thousand fields may be an optimal choice for the field-normalized citation impact indicators. Thus, the ESI classification may be too coarse-grained to be used in the selection of HCPs.

Fourth, ESI selects Highly Cited Papers from only two types of publications: regular scientific articles and reviews. Reviews are, on average, cited three times more frequently than original research articles (Glänzel, 2008). However, ESI uses a common citation baseline for articles and reviews rather than separate baselines for each. It is reasonable to expect that the common baseline gives reviews an advantage in being selected as HCPs. In the following investigation, the ratio of reviews to articles selected as HCPs will be calculated, and the extent to which reviews benefit in the selection of HCPs will be investigated.

Download English Version:

https://daneshyari.com/en/article/11031553

Download Persian Version:

https://daneshyari.com/article/11031553

Daneshyari.com