Journal of Informetrics 10 (2016) 814–829

Contents lists available at ScienceDirect

Journal of Informetrics

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

Differing disciplinary citation concentration patterns of book and journal literature?

Pei-Shan Chi

KU Leuven, ECOOM and Dept. MSI, Naamsestraat 61, 3000 Leuven, Belgium

ARTICLE INFO

Article history:
Received 4 February 2016
Accepted 30 May 2016

Keywords:
Citation concentration
Book Citation Index
Monographs
Gini coefficient
Herfindahl–Hirschman Index
Citation window

ABSTRACT

A new data source providing the citation links of book publications, the Book Citation Index (BKCI), was recently released. A deeper understanding of the citation characteristics of book publications is needed before specific bibliometric indicators can be developed. In this study, the characteristics of citation distribution concentration in journal and book literature in Web of Science Core Collection (WoS), and the differences of these characteristics across fields, levels of aggregation and citation periods were probed to determine possible applications of this new data source for bibliometric studies. Even though the aggregation scheme is not sound for evaluation practices in books, aggregation matters much more for edited books in the sciences than for those in the social sciences and humanities. Journal articles have the least concentrated citation distribution in the sciences, while books play a more important role than other document types in the humanities. In the social sciences, both edited books and authored books have citation concentration distribution similar to journal articles. In addition, the Leimkuhler curves showed that citation window length (3 years vs. 9 years) does not affect the citation concentrations of most document types in journal and book literature significantly.

1. Introduction

The coverage of journal articles in the natural sciences and life sciences is relatively high in the Science Citation Index Expanded (SCIE). By contrast, the coverage of the Social Sciences Citation Index (SSCI) and the Arts & Humanities Citation Index (A&HCI) is too inconsistent to accurately represent the output of the social sciences and humanities in all countries (Hicks, 1999; Nederhof, Zwaan, De Bruin, & Dekker, 1989; Norris & Oppenheim, 2007). The general trend that can be observed from previous studies (Nederhof, Meijer, Moed, & van Raan, 1993; Butler & Visser, 2006; Sivertsen, 2009; Engels, Ossenblok, & Spruyt, 2012) is that the more important books are in a field, the less its literature is covered by WoS because of its restriction to journal literature contrary to the fragmented feature of works in the social sciences and humanities. The limited coverage of WoS will certainly lead to inaccuracies when standard bibliometric methods are applied to these fields. Studies investigating the citation characteristics of books considered either the citations of so-called non-source items in the references of WoS journal papers (Butler & Visser, 2006; Hammarfelt, 2011; Amez, 2013; Chi, 2014) or analyzed the citations in other alternative data sources such as Google Books or Google Scholar (Kousha & Thelwall, 2009; Kousha, Thelwall, & Rezaie, 2011; Samuels, 2011, 2013). However, large-scale bibliometric studies analyzing the citation patterns of

E-mail address: peishan.chi@kuleuven.be

http://dx.doi.org/10.1016/j.joi.2016.05.005
1751-1577/

book literature were rarely conducted during the last decades due to the lack of a reliable and comprehensive data source providing citation links.

The Book Citation Index (BKCI), a new collection in the WoS, was released by Thomson Reuters in 2011 to supplement the limited coverage of WoS. It allows users to discover book literature and trace its citation links alongside journal literature (Adams & Testa, 2011). BKCI covers over 60,000 editorially selected books starting from 2005 with an additional 10,000 new titles each year (Book Citation Index, 2015). After the release, some of its limitations already have been discussed in previous studies. For example, Gorraiz, Purnell, and Glänzel (2013) question the fuzzy boundaries of subtypes of books and how to treat new editions of works, and argue to distinguish the citations of an edited book and its book chapters since a global consensus on how to cite the book editor(s), the book author(s) or the author(s) of the book chapter is lacking. They also argue that 'book' might be considered to be at a higher hierarchical level than 'journal' instead of being treated as a document type in the system, and consequently point out the lack of cumulative citation counts from different hierarchies in BKCI. Even though it is possible to distinguish in the database between monographs and edited volumes among the type 'book', a normalization of the credit for a monograph depending on the document types is required (Leydesdorff & Felt, 2012). Additionally, problems from ignoring differences between book series and annual series were indicated by Leydesdorff and Felt (2012) and confirmed by Torres-Salinas, Robinson-García, Jiménez-Contreras, and Delgado López-Cózar (2012) and Torres-Salinas, Rodríguez-Sánchez, Robinson-García, Fdez-Valdivia, and García (2013).

With the new data source, a deeper understanding of books' citation characteristics is needed before bibliometric indicators can be developed, in order to adjust and adapt them appropriately to the bibliometric toolbox (Leydesdorff, 2009; Torres-Salinas, Robinson-García, Cabezas-Clavijo, & Jiménez-Contreras, 2014). One step toward this goal could therefore be comparing the citation distributions of journal and book publication groups, since the enormous inequality in researchers' productivity and the skewness of citation distributions are at the foundation of bibliometrics and most of the empirical laws in the field deal with the power law probability distributions of both publications and citations (i.e. Bradford's law, Zipf's law, Lotka's law and Price's law).

Many concentration measures, such as the Gini Index, the Leimkuhler curve, the Herfindahl–Hirschman Index and the Characteristic Scores and Scales, from different backgrounds have been used to investigate the bibliometric distributions for journal publications (e.g. Allison, 1974; Burrell, 1992; Egghe and Rousseau, 1990; Glänzel, 2007; Larivière, Gingras, & Archambault, 2009). For example, Allison (1974) found that the inequality of productivity of scientists assessed by the Gini index increased over time. For articles' citations decreases in the Herfindahl–Hirschman Index and the least percentage of papers for specific proportion citation over time were observed (Larivière et al., 2009; Yang, Ma, Song, & Qiu, 2010). Some constant features of citation concentration shown in different sample sets with different measures include the disciplinary differences and the independence of citation window length (cf. Allison, 1974; Glänzel, 2007; Larivière et al., 2009; Yang et al., 2010).

Following this framework, this study will investigate the citation concentration characteristics of both journal and book literature in the new WoS contents, and their differences across fields, levels of aggregation of publications and lengths of citation periods, to determine possible applications of different data sources for bibliometric indicators. The research questions of this study are as follows:

- Are citation distributions in book and journal literature comparable?
- Should citations to constitutive items of a book be aggregated at the book level?
- What is the influence of citation window length for book publications? What is the difference between citation concentrations with a long-term citation window (nine years) and with a short-term citation window (three years)?

2. Material and methods

2.1. Data sources

The complete 2005–2011 contents of the Web of Science Core Collection (WoS) including the three journal databases, SCIE, SSCI and A&HCI, as well as the BKCI have been processed as source documents. Overlap among proceedings, books and journals were excluded to obtain duplicate-free datasets. Citation counts were calculated from the 2014 version of WoS database of the Centre for Research & Development Monitoring (ECOOM), KU Leuven, which indexes 50,251 books and 30,058,730 journal papers in total.

2.2. Document type

Four document types were selected for further analyses at two comparison levels: books and articles. *Edited books* and *authored books* were compared with each other at individual and aggregated levels, while the document type 'Article' for *book chapters* and *journal papers* was used because of its central importance. Table 1 shows the detailed terms of each document type.

1) Books

Since citations to individual chapters could not always be identified when they were assigned to the edited book in the citing documents (Chi, Jeuris, Thijs, & Glänzel, 2015), the citation count of an edited book was not divided by the number of