Journal of Informetrics 12 (2018) 481–493

Contents lists available at ScienceDirect

Journal of Informetrics

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

Regular article

Examining the usage, citation, and diffusion patterns of bibliometric mapping software: A comparative study of three tools

Xuelian Pan [a], Erjia Yan [b], Ming Cui [a], Weina Hua [a,*]

[a] School of Information Management, Nanjing University, Nanjing, China
[b] College of Computing and Informatics, Drexel University, Philadelphia, USA

ARTICLE INFO

Article history:
Received 15 September 2017
Received in revised form 18 March 2018
Accepted 18 March 2018

Keywords:
Software citation
Bibliometric mapping software
Knowledge diffusion
Bibliometrics
Scholarly communication

ABSTRACT

This study investigates the use, citation and diffusion of three bibliometric mapping software tools (CiteSpace, HistCite and VOSviewer) in scientific papers. We first conduct a content analysis of a sample of 481 English core journal papers—i.e., papers from journals deemed central to their respective disciplines—in which at least one of these tools is mentioned. This allows us to understand the predominant mention and citation practices surrounding these tools. We then employ several diffusion indicators to gain insight into the diffusion patterns of the three software tools. Overall, we find that researchers mention and cite the tools in diverse ways, many of which fall short of a traditional formal citation. Our results further indicate a clear upward trend in the use of all three tools, though VOSviewer is more frequently used than CiteSpace or HistCite. We also find that these three software tools have seen the fastest and most widespread adoption in library and information science research, where the tools originated. They have since been gradually adopted in other areas of study, initially at a lower diffusion speed but afterward at a rapidly growing rate.

1. Introduction

Software is vital to scientific research: it assists scientists in identifying research questions, analyzing data, visualizing results and disseminating knowledge. Indeed, "just about every step of scientific work is affected by software" (Howison et al., 2015, p. 454). However, the academic role of software has long been undervalued or, worse yet, ignored in the current publication-driven scientific reward system. This issue is especially acute in recent years, as the variety of software available freely for academic use has increased tremendously (Hannay et al., 2009; Huang et al., 2013). As the value of data is increasingly recognized (Chao, 2011; Belter, 2014; Yu et al., 2015) and a significant amount of freely available software packages are used in the scientific community (Howison & Bullard, 2016; Thelwall & Kousha, 2016), some scholars argue that software too should be valued as an academic contribution (Hafer & Kirkpatrick, 2009; Piwowar, 2013). The U.S. National Science Foundation (NSF) has recognized software as a valid research output since 2013 (NSF, 2013), and the U.K. Research Excellence Framework 2014 (Research Excellence Framework, 2013) lists it as a type of scholarly contribution. Nonetheless, many funding institutions, policy makers and administrators have not yet followed suit (Piwowar, 2013). It is therefore

* Corresponding author at: School of Information Management, Nanjing University, NO. 163 Xianlin Road, Nanjing, Jiangsu, 210023, China.
E-mail address: huawn@nju.edu.cn (W. Hua).

https://doi.org/10.1016/j.joi.2018.03.005
1751-1577/

imperative to measure the impact of software, both to gain a better understanding of its value and to better reflect that value in research evaluations and scholarly communications.

Bibliometric indicators such as citation counts and journal impact factors are often used to evaluate the impact of papers (Cartes-Velásquez & Manterola Delgado, 2014), researchers (Fu & Ho, 2013; Havemann & Larsen, 2014; Jacob, Lehrl, & Henkel, 2007), and institutions (Abramo, D'Angelo, & Costa, 2011), because they make such evaluations less time-consuming and more objective (Thelwall & Kousha, 2016; Yu et al., 2015). The increasing significance of bibliometrics in research evaluation (Belter, 2014), along with "recent developments in computing and information services" (Ding et al., 2014, p. 1820), has led some scholars to suggest that bibliometric indicators can be used to measure the impact of a wider variety of knowledge entities, such as diseases, drugs, data sets, and software (Ding et al., 2013; Pan, Yan, Wang, & Hua, 2015; Urquhart & Dunn, 2013). However, recent studies on data citation have found that a significant number of data sets mentioned in the scientific literature were not formally cited (Mooney, 2011; Peters, Kraker, Lex, Gumpenberger, & Gorraiz, 2015). Likewise, our own previous study has found that more than 30% of mentioned software in articles published in *PLOS ONE* in 2014 received no formal citations (Pan, Yan, & Hua, 2016). Howison and Bullard (2016) have found that more than 50% of software mentions did not include references among the biology articles published in Web of Science (WoS) journals. Taken together, these prior studies evince a need to use alternative metrics in addition to citations when assessing the impact of software. Much research is yet needed before we can claim to have a comprehensive understanding of software's impact on scientific research.

The study of knowledge diffusion through citations has become a standard topic in the field of library and information science (LIS) (Liu & Rousseau, 2012). Researchers have explored the diffusion of scientific knowledge on multiple levels, ranging from that of the individual paper (Liu & Rousseau, 2012), to journals (Zhao & Wu, 2014), fields of study (Yan, 2016), institutions (Börner, Penumarthy, Meiss, & Ke, 2006), and countries (Lewison, Rippon, & Wooding, 2005). In these studies, citations are generally treated as an indication of knowledge flow from the cited entity to the citing one; specifically, the cited and citing entities are usually considered as the source and target of diffusion. A variety of knowledge-diffusion approaches have been proposed to measure the impact and diffusion patterns of such research outputs as papers (Liu & Rousseau, 2010), patents (Nomaler & Verspagen, 2008), handbooks (Milojević, Sugimoto, Larivière, Thelwall, & Ding, 2014), and databases (Yu et al., 2015). However, few studies have sought to apply these same approaches to software. In this article, we aim to go beyond an analysis of the citation of software in scientific literature. Using several quantitative diffusion indicators, we investigate software diffusion patterns as well as trends in academic impact.

In this article, we consider a piece of software to be diffused in the academic communication system when it is used in scientific articles. The software used and the paper using it are considered as the source and target of diffusion, respectively. In other words, the software influences the articles that make use of it. Based on the above hypothesis, we employ knowledge diffusion indicators to explore how bibliometric mapping software tools are used and diffused in scientific papers. Bibliometric mapping software tools, sometimes called science mapping software tools, are programs that have been developed for carrying out bibliometric mapping analysis (Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011). Bibliometric mapping, which aims at presenting the structural and dynamic aspects of scientific research, is an important research topic in the field of bibliometrics, which in turn is generally viewed as a branch of LIS (Börner, Chen, & Boyack, 2003; van Eck & Waltman, 2010). Many bibliometric mapping software tools have been created and used in the scholarly community (Cobo et al., 2011), but for this article, we select three widely used examples as the targets of our analysis: CiteSpace, VOSviewer, and HistCite. We conduct a content analysis of a sample of more than 800 English-language journal papers that cite or mention the selected software tools, thereby gaining insight into the software tools' usage, citation, and diffusion patterns. The following research questions drive the investigation.

1. How are the three bibliometric mapping software tools used and cited in scientific literature?
2. What is the academic impact of the three software tools as measured by several diffusion indicators?
3. What are the diffusion patterns of the three software tools?

The answers to these questions will provide a better understanding of the impact of software on science. Though framed as a case study, our analysis is considerably broader in its implications: it employs these popular tools as a research instrument to reveal the broader landscape of software use in bibliometric research. Using the tools as a relatable pivot point, this study is able to provide a context in which to understand usage and citation statistics. Though acknowledged as a vital complement to data-driven bibliometric research, such context has been lacking in prior studies of software use and impact. Moreover, this study treats software entities as knowledge units, explores the diffusion patterns of software entities in the academic communication system, and helps present a more complete picture of the communication patterns which surround diverse research outputs.

2. Literature review

2.1. Evaluations of the impact of software

Although indicators such as number of users, downloads, reviews, and subscribers might be used to assess the academic impact of software, such data are rarely available (Thelwall & Kousha, 2016). Moreover, some of these indicators cannot measure the impact of software very well (Howison, Deelman, McLennan, Da Silva, & Herbsleb, 2015) because, in many