Journal of Informetrics 8 (2014) 569–580

Contents lists available at ScienceDirect

Journal of Informetrics

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

Including cited non-source items in a large-scale map of science: What difference does it make?

Kevin W. Boyack[a,*], Richard Klavans[b]

[a] SciTech Strategies, Inc., Albuquerque, NM 87122, USA
[b] SciTech Strategies, Inc., Berwyn, PA 19312, USA

ARTICLE INFO

Article history:
Received 3 February 2014
Received in revised form 28 March 2014
Accepted 2 April 2014
Available online 17 May 2014

Keywords:
Science mapping
Direct citation
Non-source documents
Books

ABSTRACT

Cited non-source documents such as articles from regional journals, conference papers, books and book chapters, working papers and reports have begun to attract more attention in the literature. Most of this attention has been directed at understanding the effects of including non-source items in research evaluation. In contrast, little work has been done to examine the effects of including non-source items on science maps and on the structure of science as reflected by those maps. In this study we compare two direct citation maps of a 16-year set of Scopus documents – one that includes only source documents, and one that includes non-source documents along with the source documents. In addition to more than doubling the contents of the map, from 19 M to 43 M documents, the inclusion of non-source items strongly augments the social sciences relative to the natural sciences and medicine and makes their position in the map more central. Books are also found to play a significant role in the map, and are much more highly cited on average than articles.

1. Introduction

One of the goals of science mapping, whether mapping full databases or smaller local datasets, has been to map the associated topic space as accurately as possible. It is well known, however, that coverage of the scholarly literature in the dominant databases (e.g., Web of Science, Scopus, PubMed) varies widely by discipline. Coverage is typically high in natural science disciplines such as chemistry and physics, slightly lower in the medical sciences, lower still in technical fields such as engineering, and very low in the social sciences and humanities (Butler & Visser, 2006; Hicks, 2004; Moed, 2005; van Leeuwen, 2006). Thus, we can assume that while maps of scientific areas in chemistry and physics will have close to full coverage of the topic space, maps in other disciplines may miss some topics altogether simply because of lack of database coverage of the literature associated with those topics. Global mapping, or mapping of all of science, may be especially vulnerable to the effects of this variance in coverage by discipline because all disciplines are present in a single map.

Although most science mapping efforts to date have focused on what are referred to as source items (publications in sources indexed by the database provider), there are no inherent limitations in science mapping techniques that would preclude non-source items from being included in a map. Any item in the data, whether source or cited non-source, can be mapped provided there is some information about that item which links it to other items. For example, title words can

* Corresponding author. Tel.: +1 505 856 1267.
E-mail addresses: kboyack@mapofscience.com (K.W. Boyack), rklavans@mapofscience.com (R. Klavans).

http://dx.doi.org/10.1016/j.joi.2014.04.001
1751-1577/

be used to include non-source items in text-based maps, and citation links from source to non-source items can be used to include non-source items in citation-based maps.

In this study we characterize the effects of including large numbers of non-source items in a global map of science. Two maps generated from the same set of database records and using a similar mapping methodology are compared. One map includes only source items; the second map includes source items and those non-source items that are cited at least twice. The balance of this paper proceeds as follows. First, relevant literature is reviewed. Data and methods used are then described, followed by a characterization of the two maps. Significant differences between the maps and the literature they represent are presented. The paper concludes with a discussion of the implications of this work on the characterization of science and technology.

2. Background

2.1. Source vs. non-source items

References cited by source documents in citation databases can be divided into two types:

- source items – references for which an indexed source record exists in the database,
- non-source items – references for which an indexed source record does not exist in the database.

While source items typically comprise around 75% of all references for a single publication year in the Web of Science, these numbers vary dramatically by discipline, ranging from around 90% for molecular biology and chemistry to less than 20% for the humanities (Moed, 2005). Although exact numbers vary, other studies show similar fractions of non-source items for the same broad areas of science (Butler & Visser, 2006; van Leeuwen, 2006). Hicks (2004) shows that while 85% of the output from natural scientists is in the form of journal and conference papers, the number is only around 50% for social scientists. An earlier study by Hicks (1999) reports that books comprise between 40% and 60% of the social science literature.

Non-source items are known to consist of many different document types. These include journal articles from non-indexed sources, conference papers, books, handbooks, book chapters, monographs, working papers, corporate and government reports, software, and even articles from newspapers such as the *New York Times*. Of these many document types, books seem to be getting the most recent attention. Nederhof, van Leeuwen, and van Raan (2010) analyzed highly cited non-source items in psychology and political science, finding that for references published after 1980, books formed the majority of these highly cited non-source items. Huang and Chang (2008) surveyed previous studies showing that books comprised from 15% to 89% of cited sources in various fields in the social sciences and humanities; books comprised more than half of all cited sources in 17 of the 25 individual cases surveyed. Zuccala and Guns (2013) classified documents cited by articles in over 1000 humanities journals and found there were more citations to books than to other document types combined.

More work has been done to characterize the effects of including (or not including) non-source material on research evaluation than upon science mapping, particularly in the social sciences and humanities where citations to non-source items such as books are known to be prevalent. For example, Butler and Visser (2006) performed an extensive bibliometric analysis of non-source items published by Australian universities, finding that they can substantially augment publication and citation counts in the social sciences and humanities, and can have a significant effect on rankings. Nederhof (2006) reviews efforts to address research performance in social sciences and humanities using bibliometrics and concludes that non-source items need to be included. More recently, Chi (2013) found that the inclusion of non-source items in evaluation of political science researchers significantly increases the numbers of publications reported, but has a much milder effect on their H-index values. We note that Google Scholar is gaining traction as a source for such evaluations given that non-source items seem to be extensively covered (Franceschini & Maisano, 2011).

2.2. Mapping of non-source items

From their earliest days, science mapping efforts have routinely included non-source items. In fact, non-source items were far more prevalent in early science maps than they are today. The earliest common implementations of direct citation maps, Garfield's historiographies (Garfield, 1973), did not distinguish between source and non-source items. This was also true for early document co-citation (Small, 1973) and author co-citation (White & Griffith, 1981) maps. These early studies simply mapped documents or authors, and paid no attention to the distinction between source and non-source items. The way in which citation indexes evolved played a role in this. In the 1970s and 1980s, data for many science maps was extracted from print editions of the (Social) Science Citation Index, or from electronic compilations of these data in DIALOG. These sources included lists of cited items, enabling datasets and maps to be created based on cited documents and authors, many of which did not appear as source items in the data. As the citation indexes moved from print to CDROM versions, and finally to the fully searchable Internet-based platforms of today, datasets for mapping have increasingly been constructed based on searches of source items.

Mapping of non-source journals has rarely been done. Tijssen and van Leeuwen (1995) mapped a combined set of source and non-source journals in the area of manufacturing technology and management. This required merging of data from three sources – JCR, Compendex and Ulrich's International Dictionary of Periodicals. A source journal map based on the JCR