Journal of Informetrics 11 (2017) 1080–1094

Contents lists available at ScienceDirect

Journal of Informetrics

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

Regular article

Mapping science using Library of Congress Subject Headings

Fei Shu[a,*], Jesse David Dinneen[b], Banafsheh Asadi[a], Charles-Antoine Julien[a]

[a] School of Information Studies, McGill University, Canada
[b] School of Information Management, Victoria University of Wellington, New Zealand

ARTICLE INFO

Article history:
Received 16 March 2017
Received in revised form 30 August 2017
Accepted 31 August 2017

Keywords:
Map of science
LCSH
Gephi
Assignment
Citation

ABSTRACT

Maps of scientific knowledge are generally created by analyzing scientific literature including journal articles, conference proceedings, books, and monographs. Although citation analysis is the most popular method for generating maps of science from scientific journal articles and their citations, other relationships between scientific topics can be used to map science. This study offers a map of science generated from examining non-fiction book topics and their relationships as defined by Library of Congress Subject Heading (LCSH) co-assignments. The resulting map reveals which sub-disciplines of science must be learned together, showing that *Physics* and *Mathematics* are the central topics required to practice science, which is not revealed by previous studies. This novel LCSH-based science map reveals new relations between the major sub-disciplines of science to produce a more complete representation of scientific domains and how they interact.

1. Introduction

Maps of science are meant to visualize the structure and evolution of scientific inquiry (Börner, Theriault, & Boyack, 2015; Klavans & Boyack, 2015) by classifying science and relating the classes, which are generally derived from the analyses of elements of scientific literature such as authors, journals, disciplines, or other information (Klavans & Boyack, 2009). One of the challenges of mapping science is to create a valid placement of scientific domains and their relationships (Suominen & Toivanen, 2016): relatedness can be identified by examining, for example, expert judgements, citations, subject categories, topic modeling, or course descriptions. Each of these has limitations but the Library of Congress (LC) Subject Headings (LCSH), the most widespread knowledge organization in the world (Klavans & Boyack, 2009), has never been used to map science. A collection organized by LCSH topics offers a new way to map knowledge that reflects content published in non-fiction books. This differs from traditional citation-based maps of science that reflect how research disciplines collaborate to produce new knowledge, while an LCSH-based map has the potential to uniquely reveal which topics must be learned together as expressed by the topic co-assignments of non-fiction books. This study presents a map of science generated from LCSHs assigned to a representative non-fiction science book collection and investigates differences with previous maps of science.

1.1. Mapping science

The earliest maps of science were created using expert judgment and drawn by hand. Bernal (1939), Ellingham (1948) and Balaban and Klein (2006) presented maps that explicitly represent the hierarchical structure of science topics, but their

* Corresponding author at: School of Information Studies, McGill University, 3661 Peel, Montreal, QC H3A 1X1, Canada.
E-mail address: fei.shu@mail.mcgill.ca (F. Shu).

http://dx.doi.org/10.1016/j.joi.2017.08.008

Table 1
List of knowledge mapping approaches using citation analysis.

Mapping Approach	Unit of Analysis	Topics	Relationships	Studies
Direct citation	Journals	Controlled topics are assigned to journals	Citations between journals, articles or journal disciplinary categories create a relationship between their respective topics.	Bassecoulard and Zitt (1999), Leydesdorff (2015), Leydesdorff, Moya-Anegón and Nooy (2015), Leydesdorff and Rafols (2012)
	Papers	Controlled topics are assigned to individual articles		Boyack and Klavans (2014b); Pan, Zhang and Wang (2013); Waltman and Eck (2012)
	Categories	Controlled topics are assigned to journal disciplinary categories.		Leydesdorff, Carley and Rafols (2013), Leydesdorff and Rafols (2009), Zhang, Liu, Janssens, Liang, and Glänzel (2010)
Bibliographic coupling	Journals	Controlled topics are assigned to journals	Bibliographic coupling of articles or journals create relationships between their respective topics.	Boyack (2008)
	Papers	Controlled topics are assigned to individual articles		Boyack (2008)
co-citation	Papers	Controlled topics are assigned to individual articles	Co-cited articles or categories create relationships between their respective topics.	Boyack and Klavans (2014a), Klavans and Boyack (2007), Klavans and Boyack (2008), Small (1993, 1999), Small and Garfield (1985), Small and Griffith (1974)
	Categories	Controlled topics are assigned to journal disciplinary categories.		Moya-Anegón et al. (2007); Moya-Anegón et al. (2004)
Hybrid approach	Journals	Combination of different citation-based approaches including direct citation, co-citation, bibliographic coupling etc.		Boyack, Klavans and Börner (2005), Leydesdorff (1987), Gomez-Nunez, Vargas-Quesada, Moya-Anegon, Chinchilla-Rodriguez and Batagelj (2016), Tijssen, Raan, Heiser and Wachmann (1990)
	Papers			Boyack and Klavans (2010), Braam (1991a, 1991b), Janssens, Glänzel and De Moor (2008); Persson (2010), Tijssen et al. (1990)

main disciplines and map placement differ (Klavans & Boyack, 2009). The exception to these early manual approaches was Small and Griffith (1974), who created the first citation-based map of science; co-citation analysis can express the extent to which disciplines cite each other to create new knowledge. Disciplines can be placed on a 2D map where, for example, proximity or edge thickness expresses higher co-citation rates.

Citation analysis is currently the dominant method used to generate data for knowledge maps. Table 1 presents existing approaches to generate citation-based map data. Maps created with direct citation analysis between journals or their disciplinary categories represent a broad, discipline-level structure within science that offers limited detail (Waltman & Eck, 2012), while finer questions are answered using maps derived from document-level analyses like direct citation or co-citation (Boyack & Klavans, 2014a). Hybrid approaches combining two or more different citation analysis are reported to be more accurate (Boyack & Klavans, 2010).

Table 1 presents the predominant knowledge map data generation approaches beyond which there are alternatives. For example, co-words analysis, regarded as an alternative to co-citation analysis, generates map data from the co-occurrence of words in titles, abstracts or keywords (Ding, Chowdhury, & Foo, 2001; Leydesdroff, 1989; Peters & van Raan, 1993a, 1993b; Rip & Courtial, 1984). Balaban and Klein (2006) mapped science as it was represented in undergraduate course pre-requisites at Texas A&M University, and Suominen and Toivanen (2016) developed a map of science using topic modeling to visualize latent patterns in texts retrieved from the Web of Science (WoS). Taken as a whole, the knowledge mapping literature shows a clear preference for citation-based maps, while acknowledging that other mapping approaches are necessary to provide a comprehensive understanding of the relative importance of knowledge disciplines and how they might be related.

1.2. Study objective

Citation-based maps use data provided by citation indexing databases (e.g., WoS, Scopus, or Google Scholar) that generally index journal articles while excluding other types of documents such as books.[1] The resulting maps of knowledge reflect how disciplines draw upon each-other to produce new scientific knowledge; however, a different story is likely to emerge if only books are considered. Non-fiction scientific books/monographs differ from scientific articles: books tend to cover broader

[1] Citations to books and other materials are now included is some citation databases (e.g., WoS and Scopus), but they are still rare in mapping science practice.