Science Mapping

Chaomei Chen
Drexel University, USA

Rachael Dubin
Drexel University, USA

Timothy Schultz
Drexel University, USA

INTRODUCTION

Science mapping is a fast growing interdisciplinary field originated in information science and technology. Science mapping is the development and application of computational techniques to the visualization, analysis, and modeling of a broad range of scientific and technological activities as a whole. This is an interdisciplinary field emerging from traditional library information science in the areas of scientometrics, citation analysis, and computer science in the areas of information visualization, visual analytics, data mining and knowledge discovery. A wide variety of mapping results and tools are increasingly accessible to not only researchers within the field, but also to analysts, scholars, and investigators across a wide range of domains (Cobo et al., 2011).

In order to take the full advantage of the expanding world of science mapping tools, analysts and end users need to understand their design principles and to develop competences in interpreting their results. This article provides an easy-to-access overview of the basic concepts and fundamental theories influencing the development of historical and contemporary science mapping techniques.

BACKGROUND

In the scientific literature and in our society more broadly, we are increasingly presented with an overwhelming volume of information that challenges our ability to keep abreast of current events and the state of the art. In addition to text-based information retrieval tools like Google Scholar and PubMed, one frequently used metric in science mapping is citation. A citation to a published article is often seen as a sign that the article is potentially useful to the scientific community. A citation does not even have to be positive or supportive to be considered as an impact indicator because, after all, a citation shows that the scientific community cannot simply choose to ignore the work. In this sense, any citations are good citations. The worst-case scenario for a scientific publication is that no one found it worthwhile to cite. The role of citation data in science mapping is that citations provide insights to how published articles are valued by the scientific community. Analyzing topics and language patterns in scientific literature alone is unlikely to reveal this type of evaluative insight. In this article, we will primarily focus on outlining the foundation of citation-based science mapping.

Quantitative studies of scientific literature belong to a field called scientometrics, which is part of the broad discipline of information science and technology. Scientific literature is not only large in size but also complex and dynamic. The fundamental principle of science mapping is to represent the body of scientific literature in a tangible form so that one can handle it more effectively. A science map provides an overview of the scientific landscape which can be used to support exploration, description, or explanation of the state and development of scientific knowledge and practices. We can visualize the relationships between these features at different levels of granularity, investigating direct links between individual works, conceptual relations between groups of articles, similarities between authors, and connections among larger scholarly organizations such as journals and institutions.

DOI: 10.4018/978-1-4666-5888-2.ch410

Revealing the Structure and Dynamics of Scientific Knowledge

The central goal of science mapping is to reveal the structure and dynamics of scientific knowledge (Morris & van der Veer Martens, 2008; Chen, 2013). Numerous quantitative methods have been proposed and applied to the study of various aspects of scientific knowledge, including co-word methods (Callon et al., 1986), hybrid methods of co-citation and word analysis (Braam, Moed, & van Raan, 1991; Klavans & Boyack, 2010), cited authors (McCain, 1990), the evolution of networks (Tijssen & van Raan, 1994; Rosvall & Bergstrom, 2010) to name a few.

A key concept in science mapping is the notion of citation mentioned in the previous section. Citations are references made by an author to other works. There are a wide variety of reasons why an author chooses to cite a particular article. Research has identified some of the major reasons, including acknowledgement, agreement or criticism, presentation of associated content, or favor towards personal friends or colleagues (White, 2011, p. 3347). Small (1978) pioneers the idea that one can use these citation links to find out something beyond merely who is referencing who. He suggests that citations can be used by authors as symbols of concepts. Instead of using complex or lengthy descriptions, highly cited references serve as a simple and clear symbol of the underlying concept.

The chain of citations connects intellectual ancestors and descendants. Such chains allow us to trace the trail of intellectual influence over time across generations of researchers and answer questions such as “What ideas and methods influenced this work?” and “How have current scholars incorporated this work into updated theories?” This is the basic idea of citation indexing, first introduced by Garfield (1964). Citation indexing addresses a number of weaknesses of other indexing techniques. Text-based indexing relies on the assumption that the user is able to specify precisely the target of search. In reality, we often only have a very vague idea when we start our search, and we can hardly tap into the knowledge of experts who may know the concept space very well.

For example, a keyword-based search for “science mapping” may find us articles that contain the term in text. Modern retrieval techniques are powerful. However, what if an author develops a novel technique relevant to your interest in science mapping, but the author’s vocabulary is totally different because of his/her discipline? In such a scenario, citation indexing could be a valuable tool; a citation to a typical science mapping article could lead you to this new work. Of course, vice versa, it is conceivable that two articles have related contents but fail to citation any articles in common. Hybrid approaches have been studied.

One of the oldest and simplest types of science map, the citation network, depicts articles as nodes and citations as edges, and may include other variables, such as an article’s total number of times cited mapped to node size. Citation networks can allow us to recognize at-a-glance the answers to questions such as “What are the most influential works inspired by a particular article?” and “When was a particular line of research initiated?” The use of a citation network rather than a textual list to visualize records in a citation index efficiently presents a user with a broad overview of the content and facilitates the immediate recognition of interesting network features (such as the most cited article or an abrupt shift in citation practices over time). The citation network-based science map was pioneered by Garfield (1964), who first created a “historiograph” by hand to illustrate prominent events in the study of DNA (Figure 1). This diagram displays citation linkages over time, positioning each article node along the y axis based on its date of publication. Today, a similar technique is used within Thompson Reuters’ HistCite application, which produces historiographs based on citation data exported from Web of Science (and is described in further detail later in this article).

The citation network allows us to see the role that one document plays for another as a “bit” of information, however an idea is not generally made up of only one such “bit.” If we consider that any particular work will contain an average of 15 citations (Price, 1965, p. 511), it becomes evident that each such work in fact consists of a heap of numerous referenced concepts. The co-citation relation, initially described by Small (1973), can be used to represent the similarity between these referenced works (Figure 2). We create a co-citation link between two documents each time they appear together within another document’s cited works. We can use this co-citation relation to understand what conceptual clusters make up a particular body of literature, how literature is interpreted and reinterpreted within different scholarly communities, and the changing roles played by articles over time.