

Topic-centered Scientific Collaborations of Scholarly Journals in Information Science: A Pilot Study of Implementation

ABSTRACT

This proposal is the extended work in implementation of a framework of topic-centered collaboration network. A goal of this study is to investigate the question of: In which topics and in what extent researchers collaborate with others? Topic-centered collaboration networks for two scholarly journals in the field of information science are constructed using bibliographic datasets for the past five years. This proposal contributes to the areas of both collaboration social network and big metadata.

INTRODUCTION

Scientific collaboration (also referred to as *research collaboration*) can be defined as: “interaction taking place within a social context among two or more scientists that facilitates the sharing of meaning and completion of tasks with respect to a mutually shared, superordinate goal.” (Sonnenwald, 2007, p. 645) It has been a well-established research field for the past five decades. Since the earlier works by Hagstrom (1965) and Price (1965), its mainstream research agendas have focused on the identification of networks or clusters of authors, papers (Boyack *et al.*, 2005, Leydesdorff, 1987, McCain, 1991); the generation of discipline ‘maps’ (Newman, 2004); and the evolution and growth of collaboration networks (Banks & Carley, 1996, Snijders, 2001). Hence, a co-authorship network reflects and visualizes existing scientific collaborations but does not consider integrating or involving other essential collaborative factors, such as collaborative topic, for the following two reasons: (1) collaboration and topic in scientific publications have been taken as two separate research agenda; (2) considering topic features in collaboration network has been a challenging issue (Ding, 2011). In a recent study, Anonymous (2018) proposed a new framework of scientific collaboration that takes topics into account, called “topic-associating” or “topic-centered” collaboration network. This proposal is the extension of the framework into application – constructing topic-centered collaboration networks of journals in information science and comparing them in collaborative topics. Thus, this project will discover collaborative topics as well as its intensity from co-authorships in journals. This study investigates an answer to the question: in what similar/different topics and in what extent do researchers from journals collaborate with others?

TOPIC-CENTERED COLLABORATION NETWORK

Schema

Figure 1 through 3 demonstrate how the network is constructed from publication data. Figure 1 is a schematic representation of publication data with authors (circles) connected to the papers (double squares) they write. The data include three publications – two two-authored papers (one by A and B and another by C and D) and one three-authored paper (by A, C, and D). A collaboration network is a graph in which nodes represent authors of scientific publications and edges between nodes refer to scientific collaboration and, thus, exists when the corresponding authors published a paper (Moody, 2004). Intensity of edges of the network can be simply measured by the number of publications in collaboration by associating authors. Such a network is constructed on the basis of co-authorship from individual publications, which is also called a co-authorship network/graph. Figure 2 is a collaboration network corresponding to the publications depicted in Figure 1. Figure 3 is a topic-centered collaboration network derived from the collaboration network shown in Figure 2, where authors (circles) connect to topics (squares) with extent of its association (indicated by weight assigned to edge).

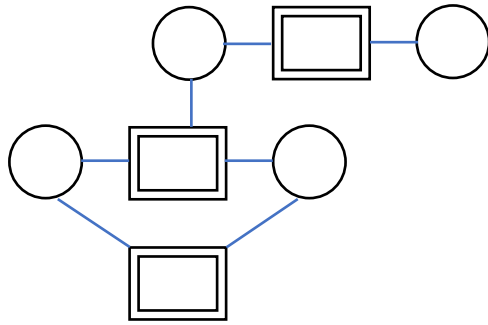


Figure 1. An example of individual publications.

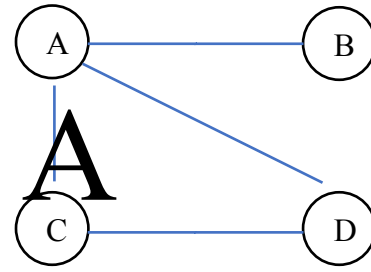


Figure 2. An example of a collaboration network.

P_1

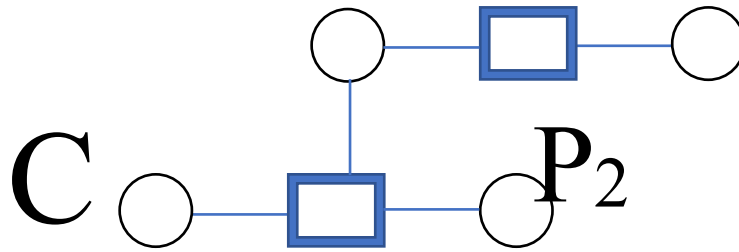


Figure 3. An example of a topic-centered collaboration network.

AD

Approach

A topic-centered collaboration network illustrates the scientific collaborative efforts in topics (Anonymous, 2018). Here are the steps for creating a topic-centered collaboration network:

First, each publication is converted into a vector in a vector space through the Latent Semantic Analysis, an information retrieval technique that discovers hidden (latent) semantic structure embedded in documents. The combination of title and abstract of each publication is utilized to represent a publication in a vector space. For vector representation, a word-publication matrix is first constructed. The value $v_{i,j}$ of a cell (i, j) (for word w_i and publication p_j) in the matrix is expressed as: $v_{i,j} = g_i \times l_{i,j}$, where g_i is a global weight obtained by (the frequency of the word w_i occurring in publication p_j)/(the frequency of the word w_i occurring in the entire publication collection) and $l_{i,j}$ is a local weight calculated by $\log_2(1 + (\text{the frequency of the word } w_i \text{ occurring in publication } p_j)/(\text{the number of words in publication } p_j))$.

Second, representative topic vectors are captured by considering and from the clusters of publication vectors. In the case of large number of publication vectors, hierarchical agglomerative bottom-up clustering algorithm will be used. Hierarchical clustering does not require any pre-set number of clusters. Clustering will stop at a pre-specified level of similarity set by experiments. A topic vector is obtained by averaging all publication vectors within a cluster.

Third, each author's associative extent into each topic is calculated by the cosine of the angle between a representative topic vector and a publication vector that is a natural metric for the distance between any two vectors.

Fourth, a topic-centered collaboration network is produced. The network consists of two different types of nodes that represent authors and topics. Unlike a general collaboration network, in this network, no link (i.e., edge) exists between any two authors. But links only exist between topic node and author node, if any association between the two exists.

EXPERIMENT

Topic-centered collaboration networks for two scholarly journals in the field of information science are constructed using bibliographic datasets for the past five years: (1) Information Research (IR) (<http://www.informationr.net/>), an open access, international, peer-reviewed, scholarly journal in a wide range of information-related disciplines; and (2) Journal of the Association for Information Science and Technology (JASIST) (<https://onlinelibrary.wiley.com/journal/23301643>), a leading international forum for peer-reviewed research in information science and a journal for the largest research community in information science. For the data collection, we collected author names, titles, and abstracts of research articles from the journal websites. By the nature of collaboration network, the bibliographic data were harvested from research articles only. Titles and abstracts were then normalized to conflate varied name spellings, including stemming and utilized as input to the collaboration networks. The presentation will report and visualize similar/distinct features of topic-centered collaboration across two networks.

SIGNIFICANCES/CONTRIBUTIONS

The significances and contributions of the proposed project can be summarized in the following tracks. First, **realization of topic-centered collaboration among researchers**. This proposal demonstrates both *collaboration* and *topic* in a same network – precisely, collaboration in the context of topic. Second, **metric/methodology in big data (a pilot study of big metadata in this case)**. Big data exists in bibliographic, indexing, and research data repositories and is an important part of the cyberinfrastructure. As big data is pivotal for the data-intensive research, there has been a lack of discussion on what theoretical, methodological, and how it might be applied to support decision making. This proposal contributes to the discussion as an application of big metadata.

REFERENCE

- Banks, D. L. & Carley, K. M. (1996). Models for network evolution. *The Journal of Mathematical Sociology* 21(1-2), 173-196.
- Boyack, K. W., Klavans, R., Borner, K. (2005). Mapping the backbone of science. *Scientometrics* 64(3), 351-374.
- Ding, Y. (2011). Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of informetrics*, 5(1), 187-203.
- Hagstrom, W. (1965). *The Scientific Community*. New York: Basic Books.
- Leydesdor, L. (1987). Various Methods for the Mapping of Science. *Scientometrics* 11, 291-320.
- McCain, K. W. (1991). Mapping Economics through the Journal Literature: An Experiment in Journal Cocitation Analysis. *Journal of the American Society for Information Science* 42(4), 290-296.
- Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American sociological review*, 69(2), 213-238.
- Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America* 101, 5200-5205.
- Price, D. J. D. (1965). Networks of scientific papers. *Science* 149, 510-515.
- Sonnenwald, D. H. (2007). Scientific collaboration. *Annual review of information science and technology*, 41(1), 643-681.
- Anonymous. (2018). Latent Semantic Analysis for Discovering Topics in Collaboration Networks. *Proceedings of the 81st Annual Meeting of the Association of Information Science and Technology, Big Metadata Analytics Workshop*, November 14, 2018, Vancouver, BC, Canada.