Renata Cristina Gutierres Castanha - Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP)
University of Wisconsin-Milwaukee (UWM)
CAPES

Carla Mara Hilário - Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP)

Paula Carina de Araújo - Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP)
Universidade Federal do Paraná (UFPR)
University of Washington (UW)
Fulbright/CAPES

Maria Cláudia Cabrini Grácio - Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP)

Citation Analysis of North American Symposium on Knowledge Organization (NASKO) Proceedings (2007-2015)

Abstract

Knowledge Organization (KO) theoretical foundations are still being developed in a continuous process of epistemological, theoretical and methodological consolidation. The remarkable growth of scientific records has stimulated the analysis of this production and the creation of instruments to evaluate the behavior of science became indispensable. We propose the Domain Analysis of KO in North America through the citation analysis of North American Symposium on Knowledge Organization (NASKO) proceedings (2007-2015). We present the citation, co-citation and bibliographic coupling analysis to visualize and recognize the researchers that influence the scholarly communication in this domain. The most prolific authors through NASKO conferences are Smiraglia, Tennis, Green, Dousa, Grant Campbell, Pimentel, Beak, La Barre, Kipp and Fox. Regarding their theoretical references, Hjørland, Olson, Smiraglia, and Ranganathan are the authors who most inspired the event's studies. The co-citation network shows the highest frequency is between Olson and Mai, followed by Hjørland and Mai and Beghtol and Mai, consolidating Mai and Hjørland as the central authors of the theoretical references in NASKO. The strongest theoretical proximity in author bibliographic coupling network occurs between Fox and Tennis, Dousa and Tennis, Tennis and Smiraglia, Dousa and Beak, and Pimentel and Tennis, highlighting Tennis as central author, that interconnects the others in relation to KO theoretical references in NASKO. The North American chapter has demonstrated a strong scientific production as well as a high level of concern with theoretical and epistemological questions, gathering researchers from different countries, universities and knowledge areas.

1 Introduction

Knowledge Organization (KO) stands as a mediator core between the production and use of information in information science (IS). Scientific knowledge framing starts from a process made from the social and labor relations of individuals in a discourse community.

2

Discourse communities are distinct social groups that share similar theories, thought, language and knowledge. They are not autonomous entities, but social constructions designed by individuals naturally tied to cultural, social and historical dimensions (Hjørland and Albrechtsen 1995; Hjørland 2002). Knowledge is generated and disseminated through their documentary records and this practice has been the core study of KO.

Hjørland and Albrechtsen (1995) agree that the best way to understand and interpret the information and scientific dynamics in the context of IS is through the study of their knowledge domains related to its discursive communities, that highlight similar theories, thought, language and knowledge.

Dahlberg (2008) understands that KO has a scientific nature and is intended to systematize knowledge units (concepts) from the elements or characteristics inherent to them. This is related to the application of concepts and classes of concepts that may allow the transfer of content (subjects). "Knowledge by itself cannot be grasped or represented unless it is presented by knowledge units and their many possible combinations in word/terms or statements" (Dahlberg 1993, 211).

KO may be considered a scientific area whose theoretical foundations are still being developed in a continuous process of epistemological, theoretical and methodological consolidation. The epistemological inquiry of the concepts and authors that constitute the KO domain are contributing to the development and consolidation of scientific knowledge.

Scientists communicate their research through scholarly communication. Conferences are one of the main channels to communicate their research and to receive feedback about their findings. We consider citation analysis a fundamental tool to better understand scholarly communication.

Furthermore, by analyzing citations and their different relations we can identify epistemological, methodological and theoretical influences in a scientific domain. And, we agree with Smiraglia's (2015, 8) statement that citations are "a form of trace evidence of thematic or theoretical relationships, and these can be used to generate visualizations of thematic or theoretical paradigms within specific communities."

In this sense, we propose the domain analysis of KO in North America. For this reason, we analyze the citations of North American Symposium on Knowledge Organization (NASKO) proceedings (2007-2015). The following indicators are analyzed: citation, co-citation and author bibliographic coupling, in order to visualize and recognize the researchers that influence the scholarly communication in this domain.

2 Bibliometric and Domain Analysis

The growth of scientific knowledge records has increasingly encouraged the mapping of this disseminated knowledge through their scientific production. The publication of scientific results allows the socialization of the produced knowledge. Knowledge is the effect of these social relationships in a helical flow between science production and scholarly communication (Guimarães 2014).

In this context, the remarkable growth of scientific records has stimulated the analysis of this production and the creation of instruments to evaluate science became indispensable. Therefore, it is noteworthy to consider bibliometrics as a method of visualization and analysis of written communication processes. Through mathematical

3

and statistical analysis, bibliometrics analyzes the nature and development of scientific disciplines.

Research consists of a driving key element of the development of knowledge in each domain. Papers and other scientific publications are the products of scientific research. Thus, bibliometrics constitutes a research field that provides valuable information about a subject, as well as the relations between disciplines, revealing social standards of scholarly communication (Hjørland and Albrechtsen 1995).

Metric Studies (including bibliometrics, scientometrics, informetrics, among other metrics) are focused on the development of appropriate methodologies for the formulation of indicators that are essential in the analysis of science. A strong quantitative trend has been observed in many metric studies. Although bibliometrics is conceptualized in quantitative aspects, science itself is a social process in which its actions and behaviors are closely linked to its context, and its quantitative aspects need to be interpreted.

Scholarly communication is an important issue related to the evolution of science and research. Hjørland (2002) presented eleven approaches to domain analysis that make a unique competence for information specialists. And, one of these approaches is bibliometric which the author considers a tool and also a method to be used in domain analysis.

Hjørland (2002) considers bibliometric studies, one of the eleven approaches of domain analysis, to be objective and consolidated methodologies for analysis and identification of a scientific domain. He recognizes the advantages of articulating bibliometric studies with epistemological and historical approaches or others of qualitative nature. He also states that "bibliometrics is a strong approach to domain analysis because it is empirical and based on detailed analysis of connections between individual documents" (Hjørland 2002, 436).

"Bibliometric studies seek to quantify in replicable terms processes of scholarly discourse, based on the trace evidence of citation patterns. They are based on theoretical distributions that have been observed at a metalevel in all of science" (Smiraglia 2015, 8). These studies involve a broad set of indicators and among the different aspects analyzed, these indicators show the researchers, themes, institutions, and knowledge areas, as well as the research front of a knowledge field, collaboration networks among scientists, groups, institutions or countries and citation and co-citation networks.

Studies of citations and co-citations are relevant procedures for analyzing interlocution among researchers and their roles in different areas of science. They contribute to the visualization of communicative and interactive processes, as well as the underlying structure of a knowledge domain (Grácio and Oliveira, 2013).

In this sense, the set of references of scientific papers is a reflection of a discourse community and they constitute a domain. Its study is based on analysis of citation frequency, whether of authors or documents, and co-occurrence frequencies (co-citation) among them.

Citation is the prior indicator of scientific communication. Through citation we may identify groups of scientists and their publications in order to recognize the researchers with greater impact in the area. It is also possible to point out the paradigms considered, methodological procedures, and give visibility to theoretical references, as well as concepts, objects and methods. The visualization of different types of citation networks

4

allows us to explain the relationship between the researchers and elite researchers, those who are responsible for building new knowledge in the field (Grácio and Oliveira 2013).

The study of co-citations deals with the frequency in which two authors or documents are cited together in the scientific production of a certain area of knowledge. The analysis of co-citation produces valid representations of the intellectual structure of a scientific domain. According to Grácio and Oliveira (2013), co-citation analysis has as fundamental premise: when two or more documents, authors, or journals are cited together in a third work, there is, at least in the perspective of the citing author, a similarity of subject among those cited. The higher the frequency of co-citation, the closer the relationship between them.

That statement is supported by Small (2004). The author explains that when scientists agree on what constitutes relevant literature, they are in fact defining the structures of their communities. "The structure of science is generated by patterns of co-recognition" and "when documents are coined, citation authors are attributing co-recognition as well as creating an association of meanings" (Small 2004, p. 71)

On the other hand, bibliographic coupling establishes a connection between two articles or authors when they use the same references. In this sense, when there is a common reference item between two articles or authors, in this case, when they refer to at least one publication in common, they are said to be bibliographically coupled. Also for the author, the set of references used by the authors in their research evidences the intellectual environment in which they work and if two articles have similar bibliographies, there is an implicit relationship between the authors.

The intensity of the coupling of two articles depends on the number of references they have in common: the greater the number of references in common, the greater the force of connection between them. That is, there is greater similarity between them, which may be thematic, theoretical, methodological or other shared particularity.

Therefore, bibliographic coupling makes possible the study of the development of the lines of research, allowing to identify the epistemic communities, the researchers and the most influential articles in a scientific domain. Although the analysis of authors' co-citation is often applied to map the intellectual structure of a scientific domain, what is actually mapped in these studies is the structure of intellectual influences in a domain as perceived by its active authors (Egghe and Rousseau, 2002; Zhao and Strotmann 2008).

Thus, qualitative and quantitative studies related to the construction of scientific intellectual spaces have a fundamental place in the investigative dimension of the domains of knowledge. This kind of study also permits verification of what is important or significant in a knowledge field, such as trends, processes, and agents, and their relationships can be identified and analyzed. Following these claims, we believe that this work contributes to better visualize the KO domain in NASKO through a bibliometric approach.

3 North America Symposium on Knowledge Organization (NASKO)

The first North American Symposium on Knowledge Organization (NASKO) was held in 2007 in Toronto, Canada. It is organized by the International Society of Knowledge Organization (ISKO), chapter for Canada and United States. Five conferences have been held: Toronto, 2007; Syracuse, 2009; Toronto, 2011; Milwaukee, 2013; and Los Angeles, 2015.

5

In previous research concerning the ISKO International conferences, Guimarães et al. 2014 points out that it was possible to observe that KO is an area or field of knowledge and/or application related to construction of specialized speeches. Its object lies in the structure of knowledge contained in documents (registered, socialized and published knowledge), with special emphasis on concepts and their modeling, as well as discursive activities in scientific domains, and social practices. The authors state that the predominant approaches are the questions related to cognitive, technological, logical-conceptual, sociocultural and management aspects, which point to the most pressing dialogues in the area.

4 Methods

NASKO proceedings are the object of this research. We retrieved all papers (76) presented at all the conferences throughout the period 2007-2015. We also chose not to perform a temporal threshold since there were only five conferences so far.

The references of the 76 papers presented at NASKO (2007-2015) were collected and organized in alphabetical order for citation counting. We consider only authors, excluding references with authorship attributed to groups and institutions, and self-citations.

The articles were then separated and their authors and references settled in a table using Excel software. A matrix was constructed with the most productive authors (10 authors, with a threshold of at least 3 papers) and their most cited referents (11 authors, with a threshold of at least 16 citations) in order to obtain the citation network.

From this, a symmetrical matrix with the authors concomitantly cited was constructed to obtain the co-citation network and another symmetrical matrix was constructed, with the authors and their concomitant references, in order to obtain the bibliographic coupling network. We also calculate the eigenvector centrality for measure the influence of the authors in the network. The eigenvector centrality is a measure that assigns relative scores to all nodes in the network based on the concept that connections from high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes, and their normalized values are between 0 and 1, 0 for no influence and 1 for total influence of na author in the network. All networks were developed in UCInet and VOSViewer software.

5 Results and discussion

Regarding the 76 papers analyzed, the most productive authors through NASKO conferences are Richard P. Smiraglia, with 6 papers, Joseph T. Tennis, with 5 papers, Rebecca Green and Thomas M. Dousa with 4 papers each, and D. Grant Campbell, David M. Pimentel, Jihee Beak, Kathryn La Barre, Margaret E. I. Kipp and Melodie J. Fox with 3 papers each.

The most cited author in the papers presented in NASKO (2007-2015) are: Birger Hjørland (50 citations), Hope Olson (38), Richard Smiraglia (32), and Shiyali R. Ranganathan (21). Clare Beghtol and Jens-Erik Mai are cited 19 times each, followed by José Augusto Chaves Guimarães and Joseph T. Tennis with 17 citations each, and Margaret Kipp, Francis Miksa and Rick Szostak, each with 16 citations.

6

Hjørland, Mai, and Olson also were the most cited authors in the Proceedings of ISKO Brazil, Spain, France and North America, according to Guimarães et al (2015), for the period from 2011 to 2013. Guimarães (2015) states that this result reveals two theoretical approaches to knowledge organization: the socio-cognitive and the ontological, in which the first stands out with the semiotics issues proposed by Mai and with the cultural studies proposed by Olson.

In order to a better visualization of the indicators, we built a citation network (Figure 1). Analyzing the citation indicators, we observe that the highest citation frequency (citing-cited authors) was found between Dousa and Miska (10), Tennis and Ranganathan (9), Dousa and Hjørland (8), Beak and Smiraglia (8) and Smiraglia and Hjørland (7), followed by La Barre and Ranganathan (6), Fox and Olson (6), Smiraglia and Beghtol (5), Tennis and Hjørland (4), and Tennis and Mai (4), as can be observed through the thickness of the lines in the network.

Figure 1 - Citation network
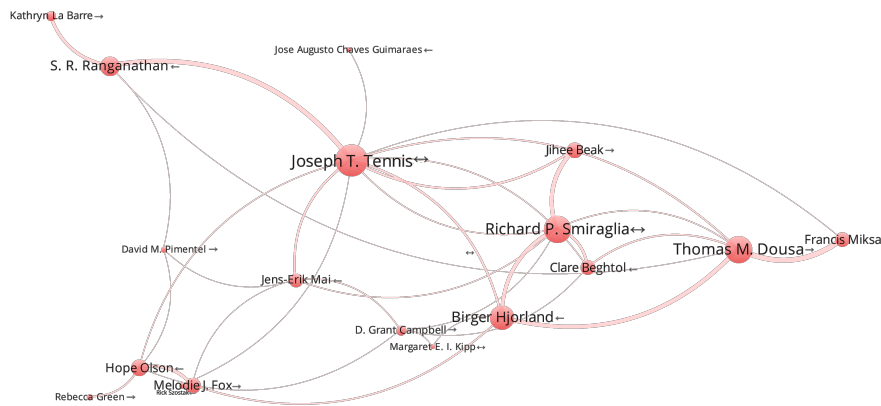
Caption: citing author →, cited author ←, both citing and cited author ↔

Small (1973, 265) states that "the number of identical citing items defines the strength of co-citation between the two cited papers. [...] Therefore, co-citation is the frequency with which two items of earlier literature are cited together by the later literature."

We can analyze the co-citation network in Figure 2. The highest frequency is between Olson and Mai (4), followed by Hjørland and Mai (3) and Beghtol and Mai (3). There is also a strong relation between Beghtol and Hjørland (3) and Tennis and Hjørland (3).

These results consolidate Mai and Hjørland as central authors of the theoretical references in KO in NASKO, as both are recognized as important referents in the area. These authors hold eigenvector values of 0.478 and 0.423, respectively.

Figure 2 - Co-citation network
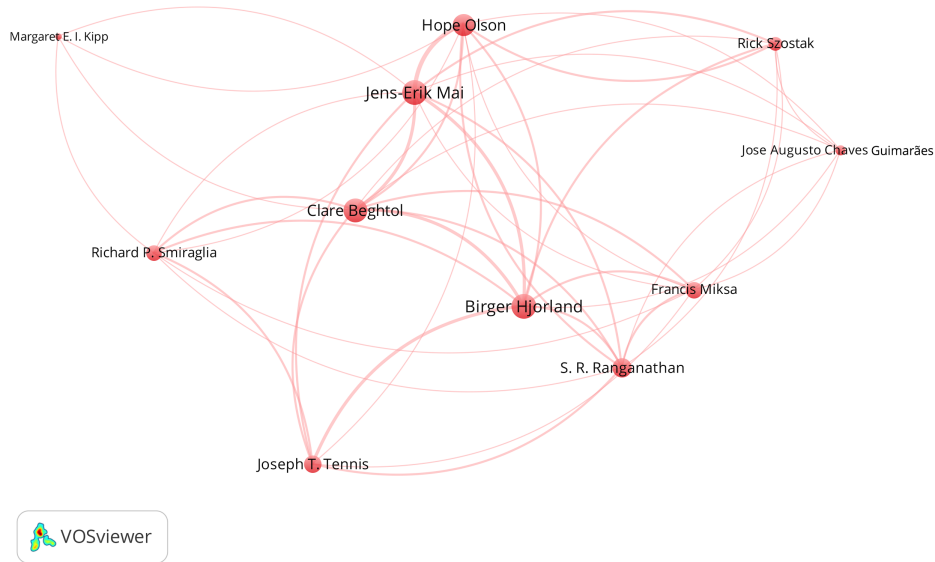
The author bibliographic coupling analysis is "a method for mapping active authors in a research field for a more realistic picture of the current state of its research activities" (Zhao and Strotmann, 2008, 2071).

The bibliographic coupling network is presented in Figure 3. The strongest theoretical proximity occurs between Fox and Tennis (4), and Dousa and Tennis (4). Fox and Tennis both cite Hjorland, Mai, Olson and Szostak, both considered some of the most important authors in KO. Dousa and Tennis cite Beghtol, Hjorland, Miksa and Ranganathan. The author bibliographic coupling is consolidated by the citing authors that are references to the domain.

There is also a representative connection between Tennis and Smiraglia (3), Dousa and Beak (3), and Pimentel and Tennis (3). By analyzing the bibliographic coupling network (Figure 3), we recognize Tennis as the author that links the others in regard to KO theoretical references in NASKO.

8

Figure 3 - Bibliographic Coupling Network
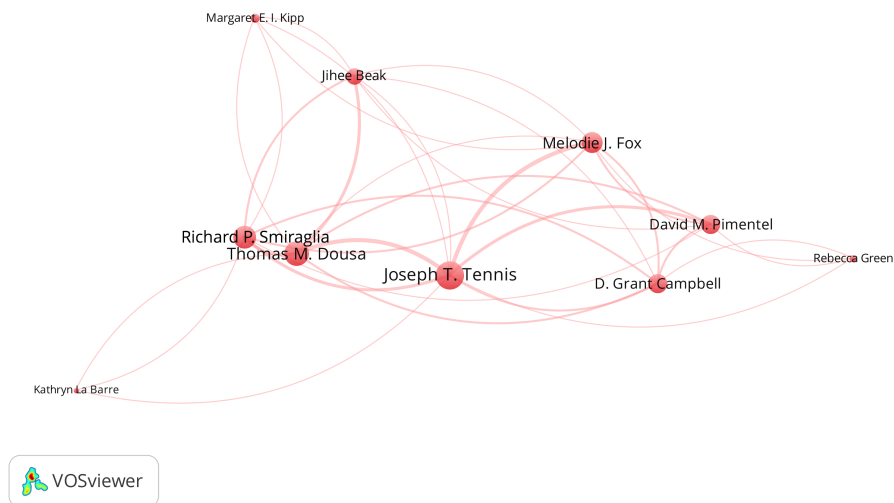
Tennis' research interests in classification theory, the foundations of KO and ethics in indexing explain why he is a central author in the bibliographic coupling network, followed by Dousa, once they hold eigenvector values of 0.478 and 0.423, respectively. Fox also researches about classification theory and ethics in indexing. Pimentel's research is also focused on classification. The implicit relationship between Dousa and Tennis is explained by their interest in the epistemological, theoretical and methodological issues in KO.

6 Conclusion

We understand that bibliometric studies are an important approach in domain analysis. Their use in combination with epistemological studies leads to better qualitative and quantitative analyses that take into account social and contextual aspects of indicators in order to better aid in the analysis. It also provides a valuable way to understand the information design, the visualization of the domain and the theoretical underpinnings of the social processes that permeate the information.

In this paper, the analysis demonstrated that the most prolific authors through NASKO conferences were Smiraglia, Tennis, Green, Dousa, Grant Campbell, Pimentel, Beak, La Barre, Kipp and Fox. Regarding their theoretical references, Hjørland, Olson, Smiraglia, and Ranganathan are the authors who most influence the conference's studies. The co-citation network shows the highest frequency is between Olson and Mai, followed by Hjørland and Mai and Beghtol and Mai, consolidating Mai and Hjørland as the central authors of the theoretical references in NASKO.

The strongest theoretical proximity in author bibliographic coupling network occurs between Fox and Tennis, Dousa and Tennis, Tennis and Smiraglia, Dousa and Beak, and

9

Pimentel and Tennis. Tennis is the author that interconnects the others in relation to KO theoretical references in NASKO.

The North American chapter demonstrates, over time, a strong scientific production in the area as well as a high level of concern with the theoretical questions concerning the organization of knowledge, gathering researchers from different countries, universities and knowledge areas.

References
Dahlberg, Ingetraut. 1993. Knowledge organization: its scope and possibilities. *Knowledge Organization,* 20(4):211-222.

Dahlberg, Ingetraut. 2008. The Information Coding Classification (ICC): A modern theory-based fully-faceted, universal system of knowledge field. *Axiomathes,* 18(2):161-176.

Guimarães, José Augusto Chaves. 2014. Domain analysis as a methodological perspective in knowledge organization. *Ciência da Informação,* 43(1):13-21.

Guimarães, José Augusto Chaves, Daniela Fernanda de Oliveira Matos, André Ynada dos Santos and Rodrigo de Sales. 2015. A dimensão conceitual da organização do conhecimento no universo científico da ISKO: uma análise de domínio a partir dos congressos da ISKO-Brasil, ISKO-Espanha, ISKO-América do Norte e ISKO-França. *Scire*, 21(2):36-26.

Oliveira, Ely Francina Tannuri, Maria Cláudia Cabrini Grácio. 2014. Studies of author cocitation analysis: a bibliometric approach for domain analysis. *IRIS,* 2(1):12-23.

Hjørland, Birger. 2002. Domain Analysis in information Science: eleven approaches traditional as well as innovative. J*ournal of Documentation*, 58(4):422-462.

Hjørland, Birger and Hanne Albrechtsen. 1995. Toward a new horizon in information science: domain-analysis. *Journal of the American Society for Information Science*, 46(6):400-425.

Small, Henry. 1973. Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the Association for Information Science and Technology*, 24(4):265-269.

Smiraglia, Richard P. 2015. *Domain Analysis for knowledge organization*: tools for ontology extraction. Chandos Information Professional Series. Waltham: Elsevier Chandos Pub.

Zhao, Dangzhi, & Strotmann, Andreas. 2008. Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology*, 59(13):2070-2086.