



ARTICLE

Lena Lúcia de Moraes¹  <https://orcid.org/0000-0002-0778-9070>

Ivette Kafure²  <https://orcid.org/0000-0002-5758-8226>

¹ University of Brasília, Brasília, DF, Brazil / e-mail: lenamoraes@gmail.com

² University of Brasília, Brasília, DF, Brazil / e-mail: ivettek@unb.br

Bibliometrics and data Science

an example search and analysis of scientific information from the Web of Science (WoS)

ABSTRACT

This practical study shows, from the perspective of Information Science (CI), how the techniques and tools of Bibliometrics and Network Science can be used to map the knowledge of a research area or subject. Based on the problem situation of a researcher¹ who needs to make a bibliographic survey of the subjects: study of users and emotions involved in episodes of interaction with information systems, the study is developed. It seeks to answer questions such as: (i) what are the keywords that occur the most and how are they related? (ii) who are the most cited authors in the study area? (iii) what are the references shared by the authors of the documents? (iv) what are the main sources of the documents? Based on the subjects, the initial bibliographic survey and consultation on Google Books Ngram Viewer, a set of terms used in the search tools was determined: information science, user study, user experience, user emotion and user affect. The scientific databases were consulted: Web Of Science (WoS), Scopus, PubMed and SciELO. The metadata of 5427 documents retrieved from WoS was used in the VOSviewer tool to build, visualize and analyze the word co-occurrence networks (in English) and the co-citation networks of authors, references and sources. The results obtained show that the summarization, visualization and analysis of the networks allow the combination of elements for understanding information and knowledge of the studied area. They make it possible to explore general and specific aspects and, thus, point out possible paths and approaches to define the scope and delimitation of the research scope. In the context of searching, retrieving and analyzing scientific documents, the tools used were effective in identifying who the interlocutors are, what they discuss and their scientific production.

KEYWORDS

Bibliographic search. Bibliometrics. Network analysis. User studies.

Bibliometria e ciência de dados

um exemplo de busca e análise de dados da Web of Science (WoS)

RESUMO

Este estudo prático mostra, na perspectiva da Ciência da Informação (CI), como as técnicas e ferramentas da Bibliometria e Ciência de Redes podem ser usadas para mapear o conhecimento de uma área ou assunto de pesquisa. A partir da situação problema de um pesquisador que necessita fazer o levantamento bibliográfico dos assuntos: estudo de usuários e emoções envolvidas em episódios de interação com sistemas de informação, o

¹ In this study the term is broadly referring to students, beginner or experienced researchers and people seeking scientific information for some purpose.

estudo é desenvolvido. Busca responder questões como: (i) quais são as palavras-chave que mais ocorrem e como estão relacionadas? (ii) quem são os autores mais citados na área de estudo? (iii) quais são as referências compartilhadas pelos autores dos documentos? (iv) quais são as principais fontes dos documentos? A partir dos assuntos, do levantamento bibliográfico inicial e de consulta no Google Books Ngram Viewer, foi determinado um conjunto de termos usados nas ferramentas de busca: information science, user study, user experience, user emotion e user affect. Foram consultadas as bases de dados científicas: Web Of Science (WoS), Scopus, PubMed e SciELO. Os metadados de 5427 documentos recuperados da WoS foram usados na ferramenta VOSviewer para construir, visualizar e analisar as redes de co-ocorrência de palavras (em Inglês) e as redes de co-citação de autores, de referências e de fontes. Os resultados obtidos mostram que a sumarização, visualização e análise das redes permitem a combinação de elementos para entendimento de informações e conhecimentos da área estudada. Viabilizam a exploração de aspectos gerais e específicos e, assim, aponta caminhos e abordagens possíveis para definição da abrangência e delimitação do escopo da pesquisa. No contexto da busca, recuperação e análise de documentos científicos, as ferramentas utilizadas foram eficazes na identificação de quem são os interlocutores, o que discutem e sua produção científica.

PALAVRAS-CHAVE

Pesquisa bibliográfica. Bibliometria. Análise de redes. Estudos de usuários



JITA: BB. Bibliometric methods.

1 INTRODUCTION

Information and Communication Technologies (ICTs) have provided many benefits to all. Accessing, in real time, information about almost everything that exists and being able to establish direct contact with information sources represents a drastic change in human society (BRAGA, 2016). On the other hand, there are the harmful effects of the great availability of information and the constant interaction with ICTs.

For Wurman (2001), one of the manifestations that our channels of perception are short-circuiting is the appearance of disturbances such as information anxiety. Anxiety is the result of the increasing distance between what we understand and what we think we should understand. And several situations can cause it: not understanding information; feeling overwhelmed by its volume; not knowing if information exists, not knowing where to find it, not knowing how to use search tools, among others.

According to Braga (2016), for psychiatry, anxiety arises as a consequence of overstimulation that cannot be discharged through action. In the case of information anxiety, it can result from both excess and lack of information. What counts for the genesis of anxiety is how we feel about information or the lack of it.

The explosion of data on the Internet and in organizations has caused a growing interest among scientists, policy makers, professionals from various sectors (lawyers, journalists, among others.) in the development of research using Big Data, Data Science and Network Science techniques. For Miller (2013), these terms refer to analytical technologies that have existed for years, but are now applied on a large scale, quickly and accessible to thousands of users. In these circumstances, several collections and databases and tools have been built and made available on the Web.

At universities and research centers, students and researchers often need and seek information to develop their research. With the technological resources available today, it initially seems simple to search and retrieve scientific information in journals and journals because they are available and accessible on the Internet. Depending on the area and subject, a search may return thousands of documents or a very restricted set.

It is in this context of information avalanche and constant changes in ICT that people search and select the information that meets their needs. And this, as observed by Wurman (2001) and Braga (2016), can cause short circuits in our perception channels.

This study was developed from the problem situation, in which a researcher, when proposing a study or research project, needs to do the bibliographic survey and map the main theoretical and methodological approaches, main authors and sources of the documents related to the subjects approached by the research. There are several ways to carry out a bibliographic survey and, it is up to the researcher to evaluate and follow the path that is most appropriate to the context of the research. In this work, tools and technologies were used, thus giving a technological background to the bibliographic survey.

The objective of the article is to show, through a practical example, how a researcher can use the techniques and tools of Bibliometrics and Network Science to help him/her in mapping the knowledge of a research area or subject. To achieve this goal, the questions are answered: (i) What are the most common keywords and how are they related? (ii) who are the most cited authors in the study area? (iii) what are the references shared by the authors of the documents? (iv) what are the main sources of the documents?

The study starts with the definition of the terms to be used in consultations and data retrieval. The terms were defined from the theme used in the problem situation: study of users and emotions involved in episodes of interaction with information and communication

technologies. It should be noted that in this study, terms in English were used because most of the databases and data processing tools available on the Internet use the English language and do not include the Portuguese language. The terms were validated in Google Books Ngram Viewer and were then used in searches in scientific databases. The metadata of the 5427 WOS documents were exported, treated and used as inputs in the construction of the networks used to answer the questions presented in the objective.

This article provides a brief introduction to Network Science, Bibliometrics and Bibliometric Networks. An overview of the software tools is presented as they are used in the practical development of the study, which were used in the development of the study: Google Books Ngram Viewer, scientific database queries and VOS viewer. Next, the focus is on the search and treatment of meta-data from the documents used in the development of the study. Afterwards, networks of co-occurrence of terms (words) and co-citation of authors, sources and references are constructed and presented. Finally, the article is concluded with a discussion about the adequate use and limitations of bibliometric networks visualizations.

2 THEORETICAL CONTEXT

The following is a brief theoretical context on Network Science, Bibliometrics and Bibliometric Networks.

2.1 Network Science

We live in a connected world. We are members of several interconnected and complex systems. Looking at the natural world around us, we have several examples of complex systems and ecosystems: biological food chains, gene collections, human brain neurons, among others. And, humans in their interactions with each other create other complex systems of human institutions, such as: Internet, social networks, viral marketing, universities, institutions.

Given the important role that complex systems play in people's lives, science and economics, their understanding, mathematical description, prediction and eventually control are the main intellectual and scientific challenges today. Behind each complex system there is an intricate network that encodes the interactions between the components of the system. In the networks, the system components are represented by nodes (or vertices) and the interactions (or connections) are the links (or edges).

The National Research Council of the United States defines network science as "the study of network representations of physical, biological, and social phenomena that lead to predictive models of these phenomena" (NRC, 2005).

Although the study of networks has a long history and roots in graph theory and sociology, the modern chapter of network science emerged in the first decade of the 21st century. Technological advances and the Internet revolution have made it possible to create, share and analyze real network data.

It is an interdisciplinary field and is based on theories and methods from several areas, including: graph theory of mathematics, data mining and visualization of information from computer science, inferential modeling of statistics and social structure of sociology (BARABÁSI, 2016).

Sampaio (2015) discusses and advocates the use of network analysis to support knowledge assessment and management. According to Sampaio, the method of network analysis applied to scientific and technological production, which are products of science, supports the construction of knowledge about science. This support helps to identify who the interlocutors are, how they relate, what they discuss and their scientific production. Network analysis uses technology to evaluate innovation and collaboration in scientific and technological production of entities, research groups, and researchers. This enables the dissemination, appropriation, reallocation or restructuring of knowledge.

2.2 Bibliometrics and Bibliometric Networks

Bibliometrics is a field that uses mathematical and statistical techniques to study patterns that arise in publications and use of documents (DIODATO, 1994). Furthermore, bibliometric is concerned with document substitutes and the relationships that can be derived or inferred from the production, manipulation or redistribution of information (NORTON, 2008).

According to Norton (2008, p. 75), bibliometric can be used to measure and describe documents and user behavior. The act of description and measurement can reveal aspects of information units, which can be explored for other applications or interpretations. By measuring and evaluating the resources of the information units it may be possible to infer patterns of intellectual activity or interest. Quotations or co-quotes can reveal research fronts, even disciplinary transformations. The application of measurement techniques does not guarantee that there will be significant results, only that new activity indicators or areas to be reviewed may emerge.

For Hjørland (2002), citation and co-quotation studies contribute to the understanding of a domain, understood as a reflection of a discursive community and its role in science. Bibliometrics opens a door and offers a way to examine the components of the information and communication enigma.

In Information Science (CI), citation networks or bibliometric networks have been widely studied in recent years (SAMPAIO, 2015). A bibliometric network consists of nodes and edges, in which nodes can be publications, journals, researchers or keywords. The edges indicate relationships between pairs of nodes and, in weighted networks, the strength of the relationship.

The most commonly studied types of relationships are citation relationships, keyword co-occurrence relationships, and co-authorship relationships. The citation relationships can be of direct quotation, co-quotation or bibliographic coupling. In bibliometric networks based on the co-authorship of researchers, research institutions or countries are linked to each other based on the number of publications they have created together. In the bibliometric networks of co-occurrences of keywords, the nodes are formed by keywords extracted from the title, abstract or list of keywords of a publication. The number of co-occurrences of two keywords is the number of publications in which both keywords occur together in the title, abstract or keyword list (GLÄNZEL, 2003) (ALVARADO, 2007).

In the networks of citation, citation or bibliographic coupling there is, in most cases, no social relationship between the nodes, since the relationship is based on the reference to the information transmitted or used and not on the researcher himself. In the case of co-authorship, a social relationship between the authors is assumed, since joint work is assumed (SAMPAIO, 2015).

The visualization of bibliometric networks has been studied since the beginning of bibliometric research and several advanced techniques and tools have been developed. The most popular visualization approaches are: distance-based; graph-based; timeline-based. There are several software tools to create, analyze and visualize networks. Some of these tools are network analysis (Pajek, Gephi, among others.) or specific tools for the visualization of bibliometric networks (CiteSpace, Sci2, VOSviewer etc.) (L3P, 2016).

3 METHODOLOGY, DATA COLLECTION, TREATMENT

This search is essentially an applied search, as it uses knowledge already developed to solve practical problems of searching, retrieving and selecting information in bibliographic databases.

Given the problem situation and in order to answer the research questions, actions were taken to obtain the results. To delimit, understand the domain, identify researchers and the most important journals on the subject under study were conducted: (i) Google Books Ngram Viewer, (ii) research in scientific databases; and, (iii) creation and analysis of co-occurrence networks of keywords and co-quote (authors, references and sources) built from the metadata of documents obtained from WOS.

3.1 Data Collection and Processing

Nowadays, faced with a need for information, the first thought that arises in people's minds is to search the most popular Internet search engine: Google² (NETMARKETSHARE, 2018).

Google Inc. has several indexing and information retrieval tools. Google Books Ngram Viewer³ is one such tool and allows you to query phrases or terms in the set of scanned books and documents. Queries can be by period and in specific languages. The search result is shown in a graph of how often the terms occur per time period. You can access the terms in the works through the links provided in the periods. In addition, the data can be downloaded for the researcher to develop his experiments (GOOGLE, 2013).

The first question that arises when encountering an Internet search system is to determine which terms to use so that the search returns the information that best suits your needs. From the subject of the problem situation and the initial bibliographic survey, a set of terms (in English) used in search engines was determined: information science, user study, user experience, user emotion and user affect. In order to validate the relevance of these terms, a Google Books Ngram Viewer query was performed. Figure 1 shows how the use of these terms varies over time.

In the academic and scientific area, besides consulting the information contained in the books, it is important to search for information in the collections, journals and bibliographic databases of the various areas of science. A bibliographic database is a digital collection that contains the records of published literature, with information on what was published, who published and where (RUAS; PEREIRA, 2014).

² www.google.com

³ <https://books.google.com/ngrams>

In order to understand how research is carried out on subjects related to the example adopted in the problem situation, consultations were carried out in the scientific databases⁴: Web of Science (WOS), Scopus, PubMed and Scielo.

The terms, previously defined, were combined in logical expressions through the logical connectives "and" and "or" to direct the focus of the results to the areas of interest. The advanced search engine option was used, adapting the logical expressions according to the settings required by the database search tools. No filters or restrictions of year, language, country and document type were applied. The queries returned a large number of documents of various categories and areas of knowledge (Table 1).

Table 1. Results of searches on WOS, Scopus, PubMed and Scielo bases.

Base	Logical query expression used in advanced search of scientific database sites	Total number of articles	Principais categorias/áreas de pesquisa e número de artigos por área
WOS	<i>TS = (Information Science AND (User Study OR User Experience OR User Emotion OR User Affect))</i>	5427	<i>Information Science Library Science (1428), Computer Science Information Systems (1132), Computer Science Interdisciplinary Applications (428), Computer Science Theory Methods (381), Education Educational Research 354), etc.</i>
Scopus	<i>("Information Science") AND ("User Study" OR "User Experience" OR "User Emotion" OR "User Affect")</i>	10.669	<i>Computer Science (7.970), Social Sciences (3.386), Engineering (1.881), Mathematics (1.333), Decision Sciences (655).</i>
PubMed	<i>((("Information Science")) AND ("User Study" OR "User Experience" OR "User Emotion" OR "User Affect"))</i>	109	not available
SciELO.ORG	<i>(Information Science) AND ((User Study) OR (User Experience) OR (User Emotion) OR (User Affection))</i>	84	<i>Applied Social Sciences (63), Health Sciences (19), Humanities (3), Agricultural Sciences (1)</i>

Source: the authors

When performing database searches, depending on the area and subject, a search may return thousands of documents, or a very restricted set. In the results presented in Table 1, the high variability in the number of documents returned by the search tools can be verified. This variability may occur due to errors in the information or other factors, such as: type of data storage and indexing and different technological and algorithmic approaches adopted in search engines.

In the present study, based on the analysis of the main research areas and the number of documents per area (Chart 1), it was decided to use the WOS documents because WOS classifies the information in a specific area related to CI - Information Science Library and has a significant number of documents.

⁴ The consultations were carried out through the CAPES Periodical Portal (<http://www.periodicos.capes.gov.br>), with remote access via CAFE. Access at: 15 Jun. 2020.

The metadata of the documents and the references cited were exported from the WOS website to the local machine. The options of full record and cited references and text without formatting were selected. WOS allows the export of 500 records at a time, so the export process was repeated for groups of 500 records (1 to 500, 501 to 1000 ... 5001 to 5427), resulting in 10 text files.

In a text editor, the 10 text files were joined (copy and paste) into a single simple text file (.txt). It should be noted that it is necessary to remove the header lines (2 first lines) and the end (last line) of the intermediate files.

To build and visualize the bibliometric networks, it was chosen to use the VOS viewer⁵. It is a graphical tool that presents a good visualization of the networks. It was developed specifically for the construction and visualization of bibliometric networks. These networks can, for example, include journals, researchers or individual publications, and can be built based on citation, bibliographic coupling, co-citation or co-authorship relationships (Chart 2). In addition, it offers text mining functionalities that can be used to build and visualize networks of co-occurrence of terms extracted from a body of scientific literature (VAN ECK and WALTMAN, 2010, 2011, 2014).

Table 2. Types of bibliometric networks that can be created in *VOSviewer*.

Type of Analysis	Analysis Unit	Description of the Relationship
Co-authored by	Authors, organizations or countries.	The list of items is determined on the basis of co-authorship of documents.
Co-occurrence	All keywords, author's keywords, keywords plus.	The list of items is determined on the basis of the number of documents in which they occur together.
Citation	Documents, sources, authors, organizations or countries.	The list of items is determined on the basis of the number of times they are cited in the documents.
Bibliographic coupling	Documents, sources, authors, organizations or countries.	The list of items is determined based on the number of references they share in the documents.
Co-citation	References cited, sources cited, authors cited (first author)	The list of items is determined on the basis of the number of times they are mentioned together in the documents.

Source: the authors

The VOSviewer builds the networks from the data co-occurrence matrix. Co-occurrence is a concept that refers to the common presence, frequency of occurrence and proximity of similar keywords in documents. Co-occurrence networks are generally used to provide a graphical visualization of possible relationships between concepts or entities represented in electronic documents. The process of building a network consists of three steps: first, a similarity matrix is calculated based on the co-occurrence matrix; second, the network is built by applying the VOS mapping technique to the similarity matrix; and in the third step, the network is translated, rotated and reflected.

In net visualizations, the distance between two nodes indicates approximately the kinship of the nodes. A shorter distance usually indicates a stronger relationship. By default, it classifies nodes in the network into clusters and uses colors to indicate the cluster to which a node has been assigned. A cluster is a set of closely related nodes (VAN ECK and WALTMAN, 2010).

In VOSviewer, the text file with the 5427-document metadata has been uploaded to generate the networks. Of the types of networks that can be created in VOSviewer (Chart 1),

⁵ Available for free at: <http://www.vosviewer.com/>

the keyword co-occurrence networks and the author citation, reference and source cited networks were constructed and analyzed.

In the first attempts to create the word co-occurrence network, analyzing the list of extracted words, it was realized that it was necessary to standardize the words and use the Thesaurus option of VOSviewer. Then, the mandatory words in the consultation (information, science and information Science) were removed because they had a high frequency of occurrence and could make it difficult to verify and visualize the other information in the network.

The generated networks and the results are presented in the following section.

4 RESULTS AND DISCUSSION

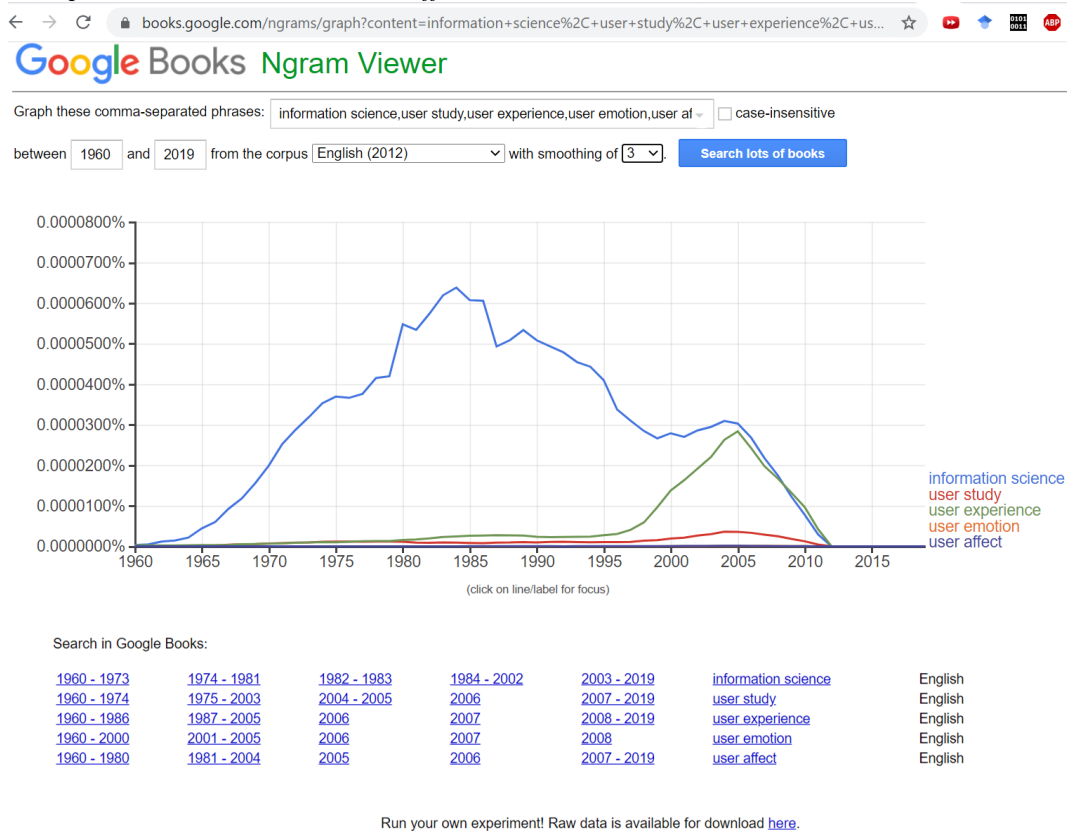
The result of the Google Books Ngram Viewer query (Figure 1) shows that the term information science appears in books from 1960 onwards, with the emergence of Information Science (CI) (BORKO, 1968) and with the evolution and expansion of computer use.

In the beginning, CI focused on information storage and retrieval systems (system-centered). As Araújo (2018) notes, the CI has been changing and, sub-areas have been consolidating according to the changes experienced by humanity, especially (but not only) with the development of technologies. The technologies have solved a series of problems, but brought other problems related to human issues (social, cultural, political, economic, legal).

At the end of the 1970s, user-centered information systems began to emerge. Cunha and others (2015, p.49), address aspects of the shift in the focus of user studies, which used to be system-centered and from the end of the 1970s on to user-centered. The situations and contexts of system use, cognition, interactivity and user interests are prioritized.

And, since 2000, the informational explosion caused the growth of studies focused on the user experience, considering cognitive, affective and emotional aspects of users (NORMAN, 2004), (HASSENZAHN, 2008).

Figure 1. Google Books Ngram Viewer query result with the terms: information science, user study, user experience, user emotion, user affect.



Fonte: <https://books.google.com/ngrams>. Access on 15/06/2020.

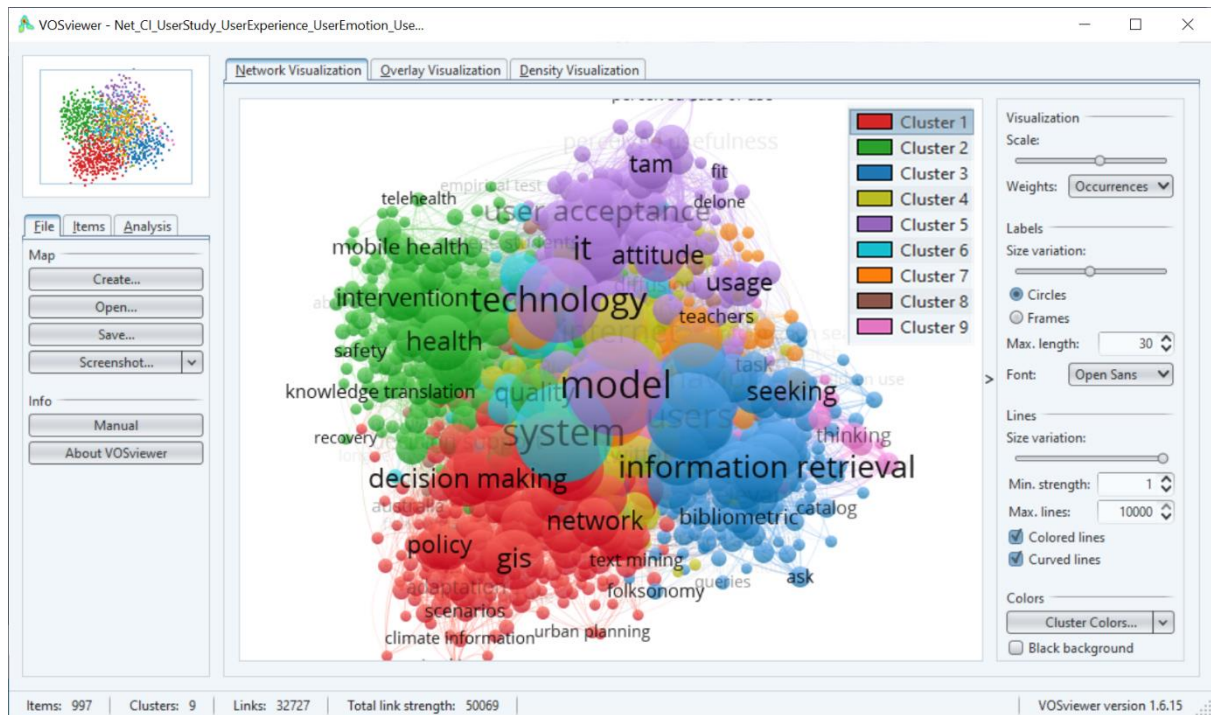
4.1 Keyword Co-occurrence Network

The keyword co-occurrence network was built to answer the question: (i) which keywords occur most and how are they related? The answer to this question is presented in Figure 3 and Table 3.

Using text mining and natural language processing techniques (Apache OpenNLP⁶), the VOSviewer extracts keywords from the metadata (title, abstract and keyword list) of scientific publications. A keyword is defined as a sequence of nouns and adjectives (ending with a noun).

⁶ <http://opennlp.apache.org>

Figure 2. Network of 997 keywords with at least 5 occurrences.








Fonte: as autoras

Figure 2 shows the network of co-occurrences of 997 keywords, with at least 5 occurrences, extracted from the title, abstract and keyword list of the 5427 WOS documents. In the network, the nodes and larger words reflect their highest occurrence, the colors indicate clustering, and the lines show the interrelationship of the keywords.

Table 3 presents the main keywords in descending order of link strength per cluster. From the clusters, one can assume the thematic inter-relationship that characterizes specific areas or applications of user studies. In the network visualization, it can be verified that the nodes of the clusters are grouped and close, characterizing the interchange and diversity of user studies.

Table 3. Keywords per cluster in descending order of link strength.

	Cluster 1 – 249 words - <i>knowledge, management, framework, decision making, policy, gis, climate change, challenge, participation, network, uncertainty, visualization, social science, citizen science, conservation, decision support, database, classification, tool, big data, organizations, future adaptation.</i>
	Cluster 2 – 200 words - <i>implementation, health, quality, care, intervention, risk, systematic review, support, people, health care, e-health, mobile health, children, review, complexity, program, cancer, adolescent, physicians, validity, meta-analysis, health information, usa, public health.</i>
	Cluster 3 – 150 words - <i>users, behavior, impact, information retrieval, web, library, students, retrieval, user study, www, seeking, patterns, needs, information seeking, digital library, academic library, university, relevance, resource, information seeking behavior, journals, access,</i>
	Cluster 4 – 113 words - <i>internet, communication, social media, motivation, online, twitter, media, social network, facebook, information literacy, coverage, security, credibility, science communication, engagement, emotion, web 2.0, algorithm, skills, simulation, business, psychology, content analysis.</i>

	Cluster 5 – 99 words - <i>model, technology, it, user acceptance, information system, perception, adoption, attitude, tam, acceptance, usage, satisfaction, dss, perceived usefulness, intention, success, ease, determinants, technology acceptance, e-learning, perceived ease, user satisfaction, work, self-efficacy.</i>
	Cluster 6 – 67 words - <i>system, innovation, trust, design science, community, evaluation, collaboration, design science research, knowledge management, methodology, diffusion, science research, requirements, business intelligence, analytics, expert system, information systems research.</i>
	Cluster 7 – 64 words - <i>design, performance, usability, education, experience, strategy, environment, user interface, hci, issues, computer, human factors, navigation, gender, augmented reality, literacy, virtual reality, instruction, cognitive style, usability evaluation, hypertext, language, view, representation.</i>
	Cluster 8 – 33 words - <i>service, preference, benefits, developing countries, systematic literature review, choice, customer satisfaction, user evaluation, growth, resistance, digital divide, e-commerce, perceived risk, effectiveness, loyalty, market, insights, service science, electronic mail, internet banking.</i>
	Cluster 9 – 22 words - <i>perspective, thinking, of-the-literature, end user, judgment, criteria, notion, information quality, relevance criteria, expertise, cognitive model, research project, dimensions, document use, image retrieval, alignment, development, domain knowledge, empirical study, pubmed.</i>

Source: the authors

The data in Figure 2 and Chart 3 show that in Information Science, according to the data recovered from WOS, there are 9 main areas of user studies, each one focused on specific aspects. The predominance of studies in information technology, systems, Internet and Web is noted.

4.2 Co-Quotation Networks

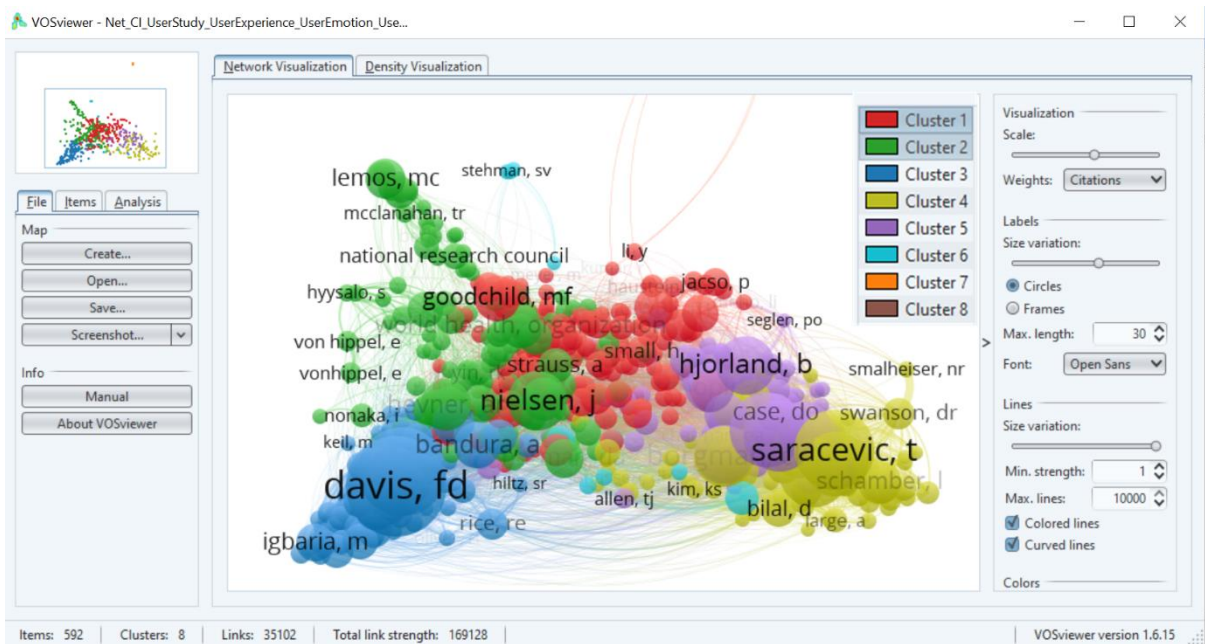
Co-citation analysis studies, based on the frequency with which two authors or documents are cited together in the scientific production of an area, show how the knowledge structure of an area is perceived by researchers. Its principle is that, when two documents or authors are cited together in a subsequent work, there is, from the perspective of the cited author, a proximity of subject between those cited. Thus, the greater the frequency of co-citation, the closer the relationship between these cited authors.

In VOSviewer, the co-citation networks of authors, references and sources were generated. To build the networks were selected authors, references and sources with at least 10 occurrences.

4.2.1 Authors' Co-Citation Network

The authors' co-citation network was built to answer the question: (ii) who are the most cited authors in the study area? The answer to this question is presented in Figure 3 and Table 4.

Figure 3. Network of 592 authors with at least 20 citations.










Source: the authors

Figure 3 shows the 592 first author only network, with at least 20 occurrences, extracted from the 5427 WOS documents. In the network, the nodes and names of the larger authors reflect their highest occurrence, the colors indicate cluster groupings, and the lines show the interrelationship of the authors.

Table 4 presents the main authors in descending order of link strength citation by cluster.

Table 4. Main authors by cluster in descending order citation link strength.

	Cluster 1 – 185 authors - <i>white, hd; thelwall, m; mccain, kw; garfield, e; small, h; chen, hc; goodchild, mf; bar-ilan, j; chen, cm; haklay, m; deci, el; leydesdorff, l; fogg, bj; preece, j; rieh, sy; tversky, a; ding, y; metzger, mj; zhang, y; pirolli, p; newman, mej; white, rw; jacso, p; kim, j; burton-jones, a.</i>
	Cluster 2 – 145 authors - <i>hevner, ar; orlikowski, wj; nielsen, j; latour, b; strauss, a; simon, ha; norman, da; gregor, s; shneiderman, b; glaser barney, g; callon, m; lemos, mc; march, st; cash, dw; peffers, k; carroll, jm; checkland, p; star, sl; hyysalo, s; walls, jg; brown, js; ericsson, ka; miles, mb; card, sk;</i>
	Cluster 3 – 110 authors - <i>davis, fd; venkatesh, v; ajzen, i; delone, wh; fishbein, m; igbaria, m; bandura, a; rice, re; markus, ml; rogers, em; ives, b; fornell, c; taylor, s; benbasat, i; agarwal, r; nunnally, jc; doll, wj; culnan, mj; chin, ww; hair, jf; mumford, e; gefen, d; goodhue, dl; zmud, rw; moore, gc; rogers, e;</i>
	Cluster 4 – 77 authors - <i>saracevic, t; belkin, nj; spink, a; ingwersen, p; borgman, cl; schamber, l; marchionini, g; harter, sp; barry, cl; bilal, d; fidel, r; swanson, dr; jansen, bj; wang, pl; ford, n; salton, g; borlund, p; wilson, p; park, tk; kuhlthau, c; lancaster, fw; janes, jw; cooper, ws; large, a; su, lt;</i>
	Cluster 5 – 62 authors - <i>dervin, b; wilson, td; kuhlthau, cc; bates, mj; ellis, d; vakkari, p; hjorland, b; savolainen, r; taylor, rs; case, do; talja, s; tenopir, c; bystrom, k; choo, cw; julien, h; cronin, b; pettigrew, ke; chatman, ea; nicholas, d; cole, c; brookes, bc; bawden, d; jarvelin, k; allen, tj; vickery, bc; wilson, t;</i>
	Cluster 6 – 10 authors - <i>nahl, d; anderson, jr; picard, rw; lopatovska, i; newell, a; ekman, p; stehman, sv; foody, gm; wulder, ma; breiman, l;</i>
	Cluster 7 – 2 authors - <i>zaidan, aa; zaidan, bb</i>

Cluster 8 – 1 authors - *faniel, im*

Source: the authors.

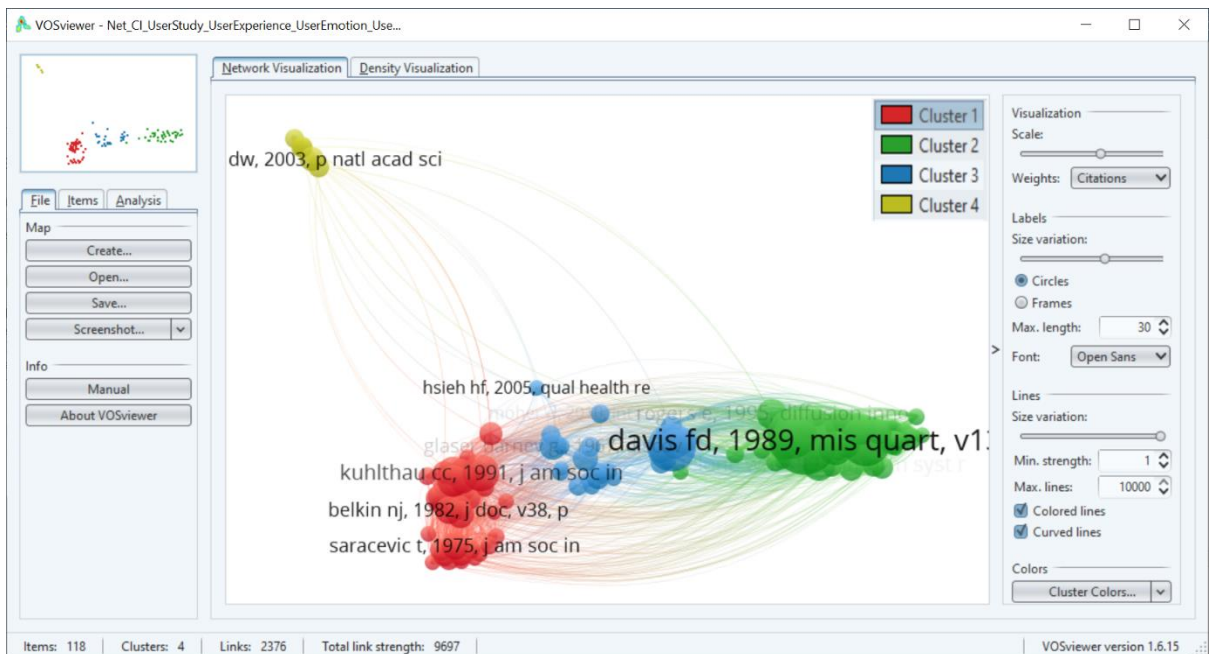
Figure 3 and Chart 4, according to the data retrieved from WOS, show that there are 6 main groups that structure the knowledge of user studies and how it is perceived by Information Science researchers. Here the researcher can visualize and analyze the knowledge structure of the area, according to the understanding of the community.

4.2.2 Co-Citation of References Network

The network of co-citation of references was built to answer the question: (iii) what are the references shared by the authors of the documents? The answer to this question is presented in Figure 4 and Table 5.

Of 177,718 references cited, 118 have at least 20 citations. Figure 4 shows the network of co-cycling references extracted from the 5427 WOS documents. In the network, the nodes and names of the largest references reflect their highest occurrence, the colors indicate the clustering and the lines present the interrelationship of the references.

Figure 5. 118 references co-citation network with at least 20 citations.






Source: the authors

Table 5 shows the main references in descending order of citation by cluster.

Table 5. Main references per cluster in descending citation order.

Cluster 1 – 45 items - kuhlthau cc, 1991, j am soc inform sci, v42, p361, doi 10.1002/(sici)1097-4571(199106)42:5<361::aid-asi6>3.0.co;2-#; wilson td, 1981, j doc, v37, p3, doi 10.1108/eb026702; wilson td, 1999, j doc, v55, p249, doi 10.1108/eum0000000007145; belkin nj, 1982, j doc, v38, p61, doi 10.1108/eb026722; dervin b, 1986, annu rev inform sci, v21, p3; saracevic t, 1975, j am soc inform sci,

<p>v26, p321, doi 10.1002/asi.4630260604; schamber l, 1990, inform process manag, v26, p755, doi 10.1016/0306</p>
<p> Cluster 2 – 39 items - davis fd, 1989, mis quart, v13, p319, doi 10.2307/249008; davis fd, 1989, manage sci, v35, p982, doi 10.1287/mnsc.35.8.982; venkatesh v, 2003, mis quart, v27, p425; fishbein m., 1975, belief attitude inte; venkatesh v, 2000, manage sci, v46, p186, doi 10.1287/mnsc.46.2.186.11926; taylor s, 1995, inform syst res, v6, p144, doi 10.1287/isre.6.2.144; ajzen i, 1991, organ behav hum dec, v50, p179, doi 10.1016/0749-5978(91)90020-t; delone wh, 1992, inform syst res, v3, p60,</p>
<p> Cluster 3 – 29 items - hevner ar, 2004, mis quart, v28, p75; glaser barney g., 1967, discovery grounded t; peffers k, 2007, j manage inform syst, v24, p45, doi 10.2753/mis0742-1222240302; march st, 1995, decis support syst, v15, p251, doi 10.1016/0167-9236(94)00041-2; rogers e. m., 2003, diffusion innovation; gregor s, 2013, mis quart, v37, p337, doi 10.25300/misq/2013/37.2.01; walls jg, 1992, inform syst res, v3, p36, doi 10.1287/isre.3.1.36; simon h.a., 1996, sci artificial; gregor s, 2006, mis quart, v30, p611;</p>
<p> Cluster 4 – 5 items - lemos mc, 2012, nat clim change, v2, p789, doi [10.1038/nclimate1614 10.1038/nclimate1614]; cash dw, 2003, p natl acad sci usa, v100, p8086, doi 10.1073/pnas.1231332100; dilling l, 2011, global environ chang, v21, p680, doi 10.1016/j.gloenvcha.2010.11.006; mcnie ec, 2007, environ sci policy, v10, p17, doi 10.1016/j.envsci.2006.10.004; lemos mc, 2005, global environ chang, v15, p57, doi 10.1016/j.gloenvcha.2004.09.004</p>

Source: the authors.

Figure 4 and Table 5, according to the data retrieved from WOS, show that there are 4 main reference groups of user studies in Information Science. Here the researcher can view and select the most important references in the knowledge area.

4.2.3 Co-Citation Sources Network

The source co-citation network was built to answer the question: (iv) what are the main sources of the documents? The answer to this question is presented in Figure 5 and Table 6.

Of the 62825 sources cited, 196 have at least 100 citations. For these 784 cited sources, the total strength of the co-citation links was calculated and shown in the network (Figure 5).

Figure 6. Network of the 196 sources with at least 100 citations.

Source: the authors

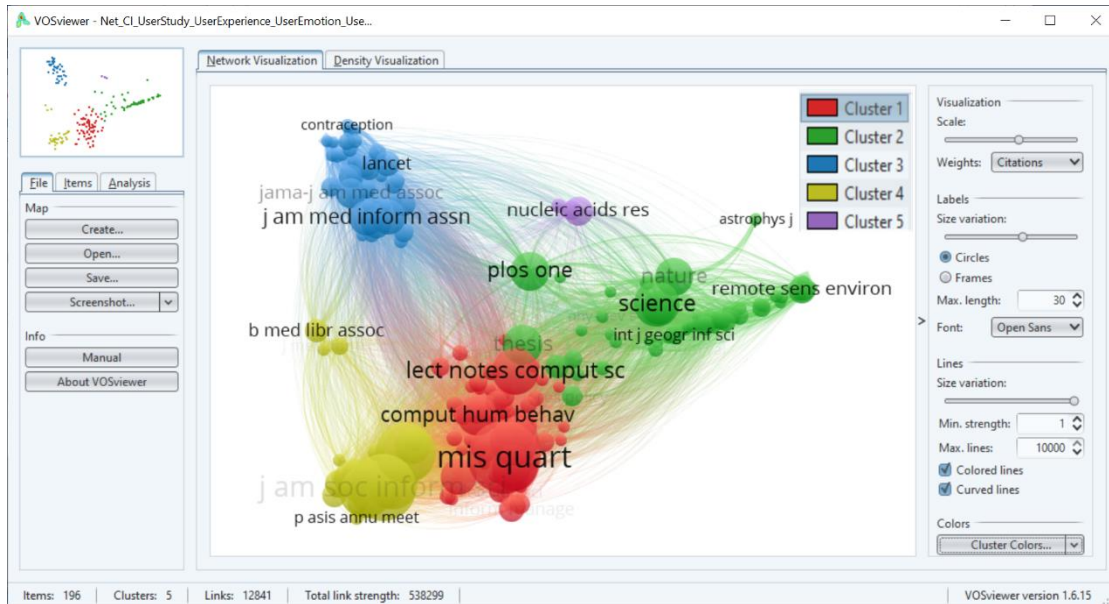


Table 6. Main sources per cluster in descending order of link strength.

	Cluster 1 – 67 sources - mis quart; inform syst res; manage sci; common acm; comput hum behav; lect notes comput sc; comput educ; decis support syst; inform manage-amster; j manage inform syst; j pers soc psychol; inform manage; int j hum-comput st; decision sci; journal of management information systems; j marketing res; psychol bull; j assoc inf syst; acad manage rev;
	Cluster 2 – 56 sources - science; plos one; nature; remote sens environ; p natl acad sci usa; thesis; int j remote sens; global environ chang; remote sens-basel; res policy; environ sci policy; climatic change; ecol soc; conserv biol; environ modell softw; int j geogr inf sci; sci technol hum val; b am meteorol soc; j environ manage; computer; geophys res lett;
	Cluster 3 – 38 sources - j am med inform assn; j med internet res; int j med inform; jama-j am med assoc; psychiat serv; cochrane db syst rev; brit med j; ann intern med; soc sci med; new engl j med; health affair; lancet; j biomed inform; bmj-brit med j; j gen intern med; patient educ couns; am j prev med; am j public health; j telemed telecare; med care; stud health technol;
	Cluster 4 – 32 sources - j am soc inf sci tec; j am soc inform sci; inform process manag; j doc; scientometrics; libr inform sci res; annu rev inform sci; j acad libr; coll res libr; inform res; libr quart; b med libr assoc; libr trends; j inf sci; j assoc inf sci tech; online inform rev; aslib proc; j inform sci; electron libr; j med libr assoc; p asis annu meet;
	Cluster 5 – 3 sources - nucleic acids res; bioinformatic; bmc bioinformatics;

Fonte: as autoras.

Figure 5 and Chart 6, according to the data recovered from WOS, show that there are 5 main groups of sources for publication of user studies in Information Science. Here the researcher can view and select the sources he or she finds most interesting for his or her research.

5 FINAL CONSIDERATIONS

This work showed, through an empirical analysis of 5427 documents recovered from WOS, how a researcher can use the techniques and tools of Bibliometrics and Network Science to perform the bibliographic survey of research.

In general, the present study allowed a greater familiarity with the terminology associated with user studies in Information Science, as well as an overview of the knowledge area.

The use of Bibliometrics and network analysis (words and co-quotation), together with the visual exploration characteristics made available in the VOS viewer tool, facilitate the combination of elements for understanding and evaluating the mode of communication, information exchange and knowledge in the studied area. With these networks, the researcher has an overview of the area of knowledge, as well as, can explore specific contexts according to their need for information. It is believed that these procedures facilitate and direct the delimitation of research, thus minimizing information anxiety.

The results obtained show that the summarization, visualization and analysis of networks allow the combination of elements for understanding information and knowledge of the studied area. They enable the exploration of general and specific aspects and, thus, point out possible paths and approaches to define the scope and delimitation of the research. In the context of the search, recovery and analysis of scientific documents, the tools used were effective in identifying who the interlocutors are, what they discuss and their scientific production.

REFERENCES

ALVARADO, Rubén Urbizagástegui. **A bibliometria: história, legitimação e estrutura.** In: TOUTAIN, Lídia Maria Batista Brandão (org.). Para entender a ciência da informação. Salvador: EDUFBA, 2007. pp. 185-217.

ARAÚJO, Carlos Alberto Ávila. **O que é Ciência da Informação.** Belo Horizonte: KMA, 2018.

BARABÁSI, Albert-László. **Network science.** Northeastern University, Boston. 2016. ISBN: 9781107076266. Disponível em: <http://networksciencebook.com>. Acesso em: 05 maio 2018.

BORKO, Harold. Information Science: what it is? **American Documentation.** 1968. p. 3-5.

BRAGA, Ryon. O excesso de informação - a neurose do século XXI. **Revista Aprender Virtual - O mundo da educação**, v. 23, 2016. Disponível em: <https://bit.ly/2Z43Koz>. Acesso em: 27 fev. 2018.

CUNHA, Murilo Bastos da; AMARAL, Sueli Angelica do Amaral; DANTAS, Edmundo Brandão. **Manual de estudo de usuários da informação.** São Paulo: Atlas, 2015. 448 p.

DIODATO, Virgil Pasquale. **Dictionary of bibliometrics.** New York: Haworth Press, 1994. ISBN: 1-56024-852-1.

GLÄNZEL, Wolfgang. **Bibliometrics as a research field**: a course on theory and application of bibliometric indicators. 2003. Disponível em: <https://bit.ly/37SFkCN>. Acesso em: 5 maio 2018.

GOOGLE, 2013. **Google Books Ngram Viewer**. In: Google Inc. Google, 2013. Disponível em: <https://books.google.com/ngrams/info>. Acesso em: 5 maio 2018.

HASSENZAHN, Marc: User experience (ux): towards an experiential perspective on product quality. In: PROCEEDINGS OF THE 20TH CONFERENCE ON L'INTERACTION HOMME-MACHINE, IHM', 8., p.11–15. **Proceedings of...** New York, NY: ACM, 2008., ISBN 978-1-60558-285-6. Disponível em: <http://doi.acm.org/10.1145/1512714.1512717>. Acesso em: 5 maio 2018.

HJØRLAND, Birger. Domain analysis in information science: eleven approaches-traditional as well as innovative. **Journal of Documentation**, v. 58, n. 4, p. 422-462, 2002.

LABORATÓRIO DE POLÍTICAS PÚBLICAS PARTICIPATIVAS (L3P) – **100 Ferramentas para Análise de Redes Sociais**. Disponível em: <https://bit.ly/2B4IreF>. Acesso em: 5 maio 2018.

MILLER, Holmes. E. Big-data in cloud computing: a taxonomy of risks. **Information Research**, v.18, n. 1, 2013. Disponível em: <http://www.informationr.net/ir/18-1/paper571.html#.WwiM2KQvzcc>.

NETMARKETSHARE. **Desktop search engine market share**. 2018. Market Share Statistics for Internet Technologies. <https://www.netmarketshare.com>. Acesso em: 15 fev. 2018.

NORMAN, Donald: **Emotional design**: why we love (or hate) everyday things. Basic Books, 2004, ISBN 9780465051359. Disponível em: <https://books.google.com.br/books?id=z2jvRlqhdIwC>. Acesso em: 5 maio 2018.

NORTON, Melanie J. **Introductory concepts in Information Science**. ASIS Monograph Series, 2008. ISBN 0-573-87087-0.

NRC, National Research Council. **Network Science**. Committee on Network Science for Future Army Applications. 2005. Disponível em: <https://www.nap.edu/catalog/11516/network-science>. Acesso em: 5 maio 2018.

RUAS, Terry Lima e PEREIRA, Luciana. Como construir indicadores de Ciência, Tecnologia e Inovação usando Web of Science, Derwent World Patent Index, Bibexcel e Pajek? **Perspectivas em Ciência da Informação**. v.19. n.3. p.52-81. jul/set. 2014. Disponível em: <https://bit.ly/3ewN0wU>. Acesso em: 5 maio 2018.

SAMPAIO, Ricardo Barros. **As estruturas globais e regionais do campo de pesquisa, desenvolvimento e inovação das doenças negligenciadas leishmaniose e tuberculose sob a ótica das redes complexas**. 2015. Tese (Doutorado em Ciência da Informação) - Faculdade de Ciência da Informação - Universidade de Brasília, 2015. Disponível em: <http://repositorio.unb.br/handle/10482/19126>. Acesso em: 28 maio 2018.

VAN ECK, Nees Jan; WALTMAN, Ludo. Software survey: VOSviewer, a computer program for bibliometric mapping. 2010. **Scientometrics**, n.84, 523–538.

VAN ECK, Nees Jan; WALTMAN, Ludo. Text mining and visualization using VOSviewer. **CoRR**, v. 84, n. 2, p. 523–538, 2011. Disponível em: <http://arxiv.org/abs/1109.2058>. Acesso em: 5 maio 2018.

VAN ECK, Nees Jan; WALTMAN, Ludo. Visualizing bibliometric networks. *In*: DING, Y.; ROUSSEAU, R.; WOLFRAM, D. (ed.). **Measuring scholarly impact: methods and practice**. Springer, 2014. p. 285–320.

WURMAN, Richard Saul. **Ansiedade de informação 2**. São Paulo: Editora Cultura, 2001.

Article submitted to the similarity system



Submitted: 27/02/2020 – Accepted: 28/05/2020 – Published: 21/06/2020
