



ARTIGO

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

Lena Lúcia de Moraes ¹  <https://orcid.org/0000-0002-0778-9070>

Ivette Kafure ²  <https://orcid.org/0000-0002-5758-8226>

¹ Universidade de Brasília, Brasília, DF, Brasil / e-mail: lenamoraes@gmail.com

² Universidade de Brasília, Brasília, DF, Brasil / e-mail: ivettek@unb.br

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 pesquisador¹ que necessita fazer o levantamento bibliográfico dos assuntos: estudo de usuários e emoções envolvidas em episódios de interação com sistemas de informação, o 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

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

¹ Neste estudo o termo ser refere de forma ampla à estudantes, pesquisadores iniciantes ou experientes e às pessoas que buscam informações científicas para algum propósito.

problem situation of a researcher 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.



JITA: BB. Bibliometric methods.

1 INTRODUÇÃO

As Tecnologias da Informação e Comunicação (TICs) têm proporcionado muitos benefícios a todos. Acessar, em tempo real, informações sobre quase tudo que existe e poder estabelecer contato direto com as fontes de informações, representa uma drástica mudança na sociedade humana (BRAGA, 2016). Por outro lado, surgem os malefícios decorrentes da grande disponibilidade de informações e da constante interação com as TICs.

Para Wurman (2001), uma das manifestações de que nossos canais de percepção estão entrando em curto-circuito é o surgimento de distúrbios como a ansiedade da informação. A ansiedade é o resultado da distância cada vez maior entre o que compreendemos e o que pensamos que deveríamos compreender. E várias situações podem provocá-la: não compreender a informação; sentir-se assoberbado por seu volume; não saber se a informação existe, não saber onde encontrá-la, não saber usar as ferramentas de busca, dentre outras.

Conforme Braga (2016), para a psiquiatria, a ansiedade surge em consequência da superestimulação que não pode ser descarregada por meio da ação. No caso da ansiedade de informação, pode resultar tanto do excesso como da carência de informação. O que conta para a gênese da ansiedade é como nos sentimos perante a informação ou a falta dela.

A explosão de dados na Internet e nas organizações têm provocado um crescente interesse entre cientistas, formuladores de políticas, profissionais de diversos setores (advogados, jornalistas, entre outros.) no desenvolvimento de pesquisas que usam técnicas de Big Data, Ciência de Dados e Ciência de Redes. Para Miller (2013), esses termos se referem a tecnologias analíticas que existem há anos, mas que agora são aplicadas em grande escala, com rapidez e acessível a milhares de usuários. Nessas circunstâncias, várias coleções e bases de dados e ferramentas foram construídas e disponibilizadas na Web.

Nas universidades e centros de pesquisas, estudantes e pesquisadores frequentemente necessitam e buscam informações para o desenvolvimento de suas pesquisas. Com os recursos tecnológicos disponíveis atualmente, inicialmente, parece simples a busca e recuperação de informações científicas em periódicos e revistas, pois elas estão disponíveis e acessíveis na Internet. A depender da área e assunto, uma busca poderá retornar milhares de documentos ou um conjunto muito restrito.

É nesse contexto de avalanche de informações e constantes mudanças nas TICs que as pessoas buscam e selecionam as informações que atendam às suas necessidades. E isso, conforme observado por Wurman (2001) e Braga (2016), pode provocar curto-circuito em nossos canais de percepção.

Este estudo foi desenvolvido a partir da situação problema, na qual, um pesquisador ao propor um estudo ou projeto de pesquisa, necessita fazer o levantamento bibliográfico e mapear as principais abordagens teóricas e metodológicas, principais autores e fontes dos documentos relacionados aos assuntos abordados pela pesquisa. Existem diversas formas de realizar um levantamento bibliográfico e, cabe ao pesquisador avaliar e trilhar o caminho que seja mais adequado ao contexto da pesquisa. Neste trabalho, foram utilizadas ferramentas e tecnologias, dando assim uma roupagem tecnológica para o levantamento bibliográfico.

Objetivo do artigo é mostrar, por meio de um exemplo prático, como um pesquisador pode usar as técnicas e ferramentas da Bibliometria e Ciência de Redes para ajudá-lo no mapeamento do conhecimento de uma área ou assunto de pesquisa. Para atingir esse objetivo, são respondidas as questões: (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?

O estudo inicia com a definição dos termos a serem usados nas consultas e recuperação dos dados. Os termos foram definidos a partir do tema utilizado na situação problema: estudo de usuários e emoções envolvidas em episódios de interação com tecnologias da informação e comunicação. Ressalta-se que no presente estudo, foram utilizados os termos em Inglês pois, a maioria das bases e ferramentas de tratamento de dados disponibilizadas na Internet usam a língua inglesa e não contemplam a língua portuguesa. Os termos foram validados no *Google Books Ngram Viewer* e, em seguida, foram usados nas pesquisas nas bases de dados científicas. Os metadados dos 5427 documentos da *WOS* foram exportados, tratados e usados como insumos na construção das redes usadas para responder as perguntas apresentadas no objetivo.

Este artigo fornece uma breve introdução Ciência de Redes, Bibliometria e redes bibliométricas. Uma visão geral das ferramentas de software é apresentada na medida que as ferramentas são usadas no desenvolvimento prático do estudo. que foram usadas no desenvolvimento do estudo: *Google Books Ngram Viewer*, consultas em base de dados científicas e *VOSviewer*. Em seguida, o foco é na busca e tratamento dos meta-dados dos documentos usados no desenvolvimento do estudo. Posteriormente são construídas e apresentadas as redes de co-ocorrência de termos (palavras) e de co-citação de autores, fontes e referências. Finalmente, o artigo é concluído com uma discussão sobre o uso adequado e limitações de visualizações de redes bibliométricas.

2 CONTEXTO TEÓRICO

A seguir, são apresentadas um breve contexto teórico sobre Ciência de Redes, Bibliometria e redes bibliométricas.

| 4

2.1 Ciência de Redes

Vivemos num mundo conectado. Somos membros de vários sistemas interligados e complexos. Olhando para o mundo natural ao nosso redor, temos vários exemplos de sistemas e ecossistemas complexos: cadeias alimentares biológicas, coleções de genes, neurônios do cérebro humano, dentre outros. E, os seres humanos em suas interações uns com os outros criam outros sistemas complexos de instituições humanas, tais como: Internet, redes sociais, marketing viral, universidades, instituições.

Dado o importante papel que os sistemas complexos desempenham na vida das pessoas, na ciência e na economia, seu entendimento, descrição matemática, previsão e, eventualmente, controle são os principais desafios intelectuais e científicos atuais. Por trás de cada sistema complexo existe uma rede intrincada que codifica as interações entre os componentes do sistema. Nas redes, os componentes do sistema são representados por nós (ou vértices) e as interações (ou conexões) são os links (ou arestas).

O Conselho Nacional de Pesquisa dos Estados Unidos define a Ciência de Redes como "*o estudo de representações em rede de fenômenos físicos, biológicos e sociais que levam a modelos preditivos desses fenômenos*" (NRC, 2005).

Embora o estudo das redes tenha uma longa história e com raízes na teoria dos grafos e na sociologia, o capítulo moderno da ciência de redes surgiu na primeira década do século XXI. Os avanços tecnológicos e a revolução da Internet possibilitaram a criação, compartilhamento e análise de dados de redes reais.

É um campo interdisciplinar e fundamenta-se em teorias e métodos de várias áreas, incluindo: teoria dos grafos da matemática, mineração de dados e visualização de informação

da ciência da computação, modelagem inferencial da estatística e estrutura social da sociologia (BARABÁSI, 2016).

Sampaio (2015) discute e defende o uso de análise de redes no apoio à avaliação e à gestão do conhecimento. Segundo Sampaio, o método de análise de redes aplicado à produção científica e tecnológica, que são produtos da ciência, apoia a construção do conhecimento acerca da ciência. Esse apoio ajuda identificar quem são os interlocutores, como se relacionam, o que discutem e a sua produção científica. A análise de redes utiliza a tecnologia para avaliar a inovação e colaboração na produção científica e tecnológica de entidades, grupos de pesquisa, e pesquisadores. Isso possibilita a difusão, apropriação, realocação ou reestruturação do conhecimento.

2.2 Bibliometria e Redes Bibliométricas

Bibliometria é um campo que usa técnicas matemáticas e estatísticas para estudo de padrões que surgem nas publicações e uso de documentos (DIODATO, 1994). Além disso, a bibliometria se preocupa com os substitutos de documentos e as relações que podem ser derivadas ou inferidas da produção, manipulação ou redistribuição de informações (NORTON, 2008).

Conforme Norton (2008, p. 75), a bibliometria pode ser usada para medir e descrever documentos e comportamentos dos usuários. O ato de descrição e medição pode revelar aspectos das unidades de informação, que podem ser explorados para outras aplicações ou interpretações. Ao medir e avaliar os recursos das unidades de informação pode ser possível inferir padrões de atividade intelectual ou interesse. Citações de citação ou co-citações podem revelar frentes de pesquisa, mesmo transformações disciplinares. A aplicação de técnicas de medição não garante que haverá resultados significativos, apenas que possam surgir novos indicadores de atividade ou áreas a serem revisadas.

Para Hjørland (2002), os estudos de citação e co-citação contribuem para a compreensão de um domínio, entendido como reflexo de uma comunidade discursiva e do seu papel na ciência. A bibliometria abre uma porta e oferece um caminho para examinar os componentes do enigma de informação e comunicação.

Na Ciência da Informação (CI), as redes de citação ou redes bibliométricas tem sido bastante estudada nos últimos anos (SAMPAIO, 2015). Uma rede bibliométrica consiste em nós e arestas, nas quais os nós podem ser publicações, periódicos, pesquisadores ou palavras-chave. As arestas indicam relações entre pares de nós e, nas redes ponderadas, a força da relação.

Os tipos de relações mais comumente estudados são relações de citação, relações de co-ocorrência de palavras-chave e relações de co-autoria. As relações de citação podem ser de citação direta, co-citação ou de acoplamento bibliográfico. Nas redes bibliométricas baseadas na co-autoria de pesquisadores, instituições de pesquisa ou países estão ligados uns aos outros com base no número de publicações que eles criaram em conjunto. Nas redes bibliométricas de co-ocorrências de palavras-chave, os nós são formados por palavras-chave extraídas do título, do resumo ou da lista de palavras-chave de uma publicação. O número de co-ocorrências de duas palavras-chave é o número de publicações nas quais ambas as palavras-chave ocorrem juntas no título, no resumo ou na lista de palavras-chave (GLÄNZEL, 2003) (ALVARADO, 2007).

Nas redes de citação, cocitação ou acoplamento bibliográfico não existe, na maior parte das vezes, uma relação social entre os nós, uma vez que o relacionamento se baseia no referenciamento à informação transmitida ou utilizada e não no pesquisador propriamente dito.

No caso da coautoria, assume-se uma relação social entre os autores, uma vez que se pressupõe o trabalho conjunto (SAMPAIO, 2015).

A visualização de redes bibliométricas é estudada desde os primórdios da pesquisa bibliométrica e diversas técnicas e ferramentas avançadas foram desenvolvidas. As abordagens de visualização mais populares são: baseada na distância; baseada em grafos; baseada na linha do tempo. Existem diversas ferramentas de software para criar, analisar e visualizar redes. Algumas dessas ferramentas são de análise de rede (*Pajek*, *Gephi*, entre outras.) ou específicas para a visualização de redes bibliométricas (*CiteSpace*, *Sci2*, *VOSviewer* etc.) (L3P, 2016).

3 METODOLOGIA, COLETA, TRATAMENTO DOS DADOS

Essa pesquisa é essencialmente uma pesquisa aplicada, pois usa conhecimentos já desenvolvidos para solucionar problemas práticos de busca, recuperação e seleção de informações em bases de dados bibliográficas.

Dada a situação problema e visando responder as questões da pesquisa foram realizadas ações direcionadas à obtenção dos resultados. Para delimitar, compreender o domínio, identificar os pesquisadores e os periódicos mais importantes sobre o assunto em estudo foram realizados: (i) levantamento de ocorrências de termos no *Google Books Ngram Viewer*, (ii) pesquisa nas bases de dados científicas; e, (iii) criação e análise das redes de co-ocorrência de palavras-chave e de co-citação (autores, referências e fontes) construídos a partir dos metadados dos documentos obtidos da *WOS*.

3.1 Coleta e Tratamento dos Dados

Atualmente, diante de uma necessidade de informação, o primeiro pensamento que surge na mente das pessoas é buscar no mais popular mecanismo de busca da Internet: *Google*² (NETMARKETSHARE, 2018).

A *Google Inc.* possui diversas ferramentas de indexação e recuperação de informações. O *Google Books Ngram Viewer*³ é uma dessas ferramentas e permite a consulta de frases ou termos no conjunto de livros e documentos digitalizados. As consultas podem ser por período e em idiomas específicos. O resultado da pesquisa é mostrado num gráfico de frequência de ocorrência dos termos por período de tempo. É possível acessar os termos nas obras por meio dos links disponibilizados nos períodos. Além disso, os dados podem ser baixados para que o pesquisador desenvolva seus experimentos (GOOGLE, 2013).

A primeira dúvida que surge ao se deparar com um sistema de busca na Internet é determinar quais os termos usar para que a busca retorne as informações que mais se adequam às necessidades. A partir do tema da situação problema e do levantamento bibliográfico inicial, foi determinado um conjunto de termos (em Inglês) usados nas ferramentas de busca: *information science*, *user study*, *user experience*, *user emotion* e *user affect*. Para validar a pertinência desses termos foi realizada uma consulta no *Google Books Ngram Viewer*. A Figura 1 mostra como o uso desses termos variam ao longo do tempo.

Na área acadêmica e científica, além de consultar as informações contidas nos livros, é importante buscar informações nas coleções, periódicos e bases de dados bibliográficos das

² www.google.com

³ <https://books.google.com/ngrams>

diversas áreas da Ciência. Uma base de dados bibliográficos é uma coleção digital que contém os registros da literatura publicada, com informações sobre o que foi publicado, quem publicou e onde se publicou (RUAS; PEREIRA, 2014).

Para compreender como se desenvolvem as pesquisas dos assuntos relacionados ao exemplo adotado na situação problema, foram realizadas consultas nas bases de dados científicas⁴: *Web Of Science (WOS), Scopus, PubMed e Scielo*.

Os termos, definidos anteriormente, foram combinados em expressões lógicas por meio dos conectivos lógicos “e” (and) e “ou” (or) para direcionar o foco dos resultados para as áreas de interesse. Utilizou-se a opção avançada dos mecanismos de busca, adaptando as expressões lógicas de acordo com as configurações exigidas pelas ferramentas de busca das bases de dados. Não foram aplicados filtros ou restrições de ano, língua, país e tipo de documento. As consultas retornaram um grande número de documentos de diversas categorias e áreas de conhecimento (Quadro 1).

Quadro 1. Resultado das buscas nas bases *WOS, Scopus, PubMed e Scielo*.

Base	Expressão lógica de consulta usada na busca avançada dos sites das bases de dados científicas	Número total de artigos	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	Não disponibilizado.
SciELO. ORG	<i>(Ciência da Informação) AND ((Estudo de Usuários) OR (Experiência do Usuário) OR (Emoção do Usuário) OR (Afeto do Usuário))</i>	84	<i>Ciências Sociais Aplicadas (63), Ciências da Saúde (19), Ciências Humanas (3), Ciências Agrárias (1)</i>

Fonte: As autoras

Ao realizar pesquisas nas bases de dados, a depender da área e assunto, uma busca pode retornar milhares de documentos, ou um conjunto muito restrito. Nos resultados apresentados no Quadro 1, pode-se verificar a alta variabilidade no número de documentos retornados pelas ferramentas de busca. Essa variabilidade pode ocorrer por erros nas informações ou outros fatores, tais como: tipo de armazenamento e indexação dos dados e diferentes abordagens tecnológicas e algorítmicas adotadas nos mecanismos de busca.

⁴ As consultas foram realizadas por meio do Portal de Periódicos da CAPES (<http://www.periodicos.capes.gov.br>), com acesso remoto via CAFE. Acesso em: 15 jun. 2020.

No presente estudo, a partir da análise das principais áreas de pesquisa e do número de documentos por área (Quadro 1), optou-se por utilizar os documentos da WOS, pois a WOS classifica as informações numa área específica relacionada à CI - *Information Science Library Science* e possui uma quantidade significativa de documentos.

Os metadados dos documentos e as referências citadas foram exportados do site da WOS para a máquina local. Foram selecionadas as opções de registro completo e referências citadas e de texto sem formatação. A WOS permite a exportação de 500 registros por vez, assim, o processo de exportação foi repetido para grupos de 500 registros (1 a 500, 501 a 1000 ... 5001 a 5427), resultando em 10 arquivos textos.

Num editor de texto, os 10 arquivos textos foram juntados (cópia e cola) num único arquivo texto (.txt) simples. Ressalta-se que é necessário a retirada das linhas do cabeçalho (2 primeiras linhas) e do fim (última linha) dos arquivos intermediários.

Para construir e visualizar as redes bibliométricas, optou-se por utilizar o *VOSviewer*⁵. É uma ferramenta gráfica que apresenta uma boa visualização das redes. Foi desenvolvida especificamente para a construção e visualização de redes bibliométricas. Essas redes podem, por exemplo, incluir periódicos, pesquisadores ou publicações individuais, e podem ser construídas com base em citação, acoplamento bibliográfico, co-citação ou relações de co-autoria (Quadro 2). Além disso, oferece funcionalidades de mineração de texto que podem ser usadas para construir e visualizar redes de co-ocorrência de termos extraídos de um corpo de literatura científica (VAN ECK e WALTMAN, 2010, 2011, 2014).

Quadro 2. Tipos de redes bibliométricas que podem ser criadas no *VOSviewer*.

Tipo de Análise	Unidade de Análise	Descrição da Relação
Co-autoria	Autores, organizações ou países.	A relação dos itens é determinada com base na co-autoria de documentos.
Co-ocorrência	Todas palavras-chave, palavras-chave do autor, palavras-chave plus.	A relação dos itens é determinada com base no número de documentos nos quais eles ocorrem juntos.
Citação	Documentos, fontes, autores, organizações ou países.	A relação dos itens é determinada com base no número de vezes que eles se citam nos documentos.
Acoplamento bibliográfico	Documentos, fontes, autores, organizações ou países.	A relação dos itens é determinada com base no número de referências que eles compartilham nos documentos.
Co-citação	Referências citadas, fontes citadas, autores citados (primeiro autor)	A relação dos itens é determinada com base no número de vezes que são citados em conjunto nos documentos.

Fonte: as autoras

O *VOSviewer* constrói as redes a partir da matriz de co-ocorrência dos dados. Co-ocorrência é um conceito que se refere à presença comum, frequência de ocorrência e proximidade de palavras-chave semelhantes nos documentos. As redes de co-ocorrência são geralmente usadas para fornecer uma visualização gráfica das possíveis relações entre conceitos ou entidades representadas em documentos eletrônicos. O processo de construção de uma rede consiste de três etapas: primeiro, uma matriz de similaridade é calculada com base na matriz de

⁵ Disponível gratuitamente em: <http://www.vosviewer.com/>

co-ocorrência; segundo, a rede é construída aplicando a técnica de mapeamento VOS à matriz de similaridade; e na terceira etapa, a rede é traduzida, girada e refletida.

Nas visualizações das redes, a distância entre dois nós indica aproximadamente o parentesco dos nós. Uma distância menor, geralmente, indica uma relação mais forte. Por padrão, classifica os nós da rede em clusters e usa cores para indicar o cluster ao qual um nó foi atribuído. Um cluster é um conjunto de nós intimamente relacionados (VAN ECK e WALTMAN, 2010).

No *VOSviewer*, o arquivo texto com os metadados dos 5427 documentos foi carregado para gerar as redes. Dos tipos de redes que podem ser criadas no *VOSviewer* (Quadro 1), foram construídas e analisadas as redes de co-ocorrência de palavras-chave e as redes de co-citação de autores, referências e fontes citadas.

Nas primeiras tentativas de criação da rede co-ocorrência de palavras, analisando a lista de palavras extraídas, percebeu-se que era necessário padronizar as palavras e usar a opção de *Thesaurus* do *VOSviewer*. Em seguida, foram removidas as palavras obrigatórias na consulta (*information, science e nformation Science*) porque apresentaram alta frequência de ocorrência e poderia dificultar a verificação e visualização das demais informações da rede.

As redes geradas e os resultados são apresentados na seção seguinte.

4 RESULTADOS E DISCUSSÃO

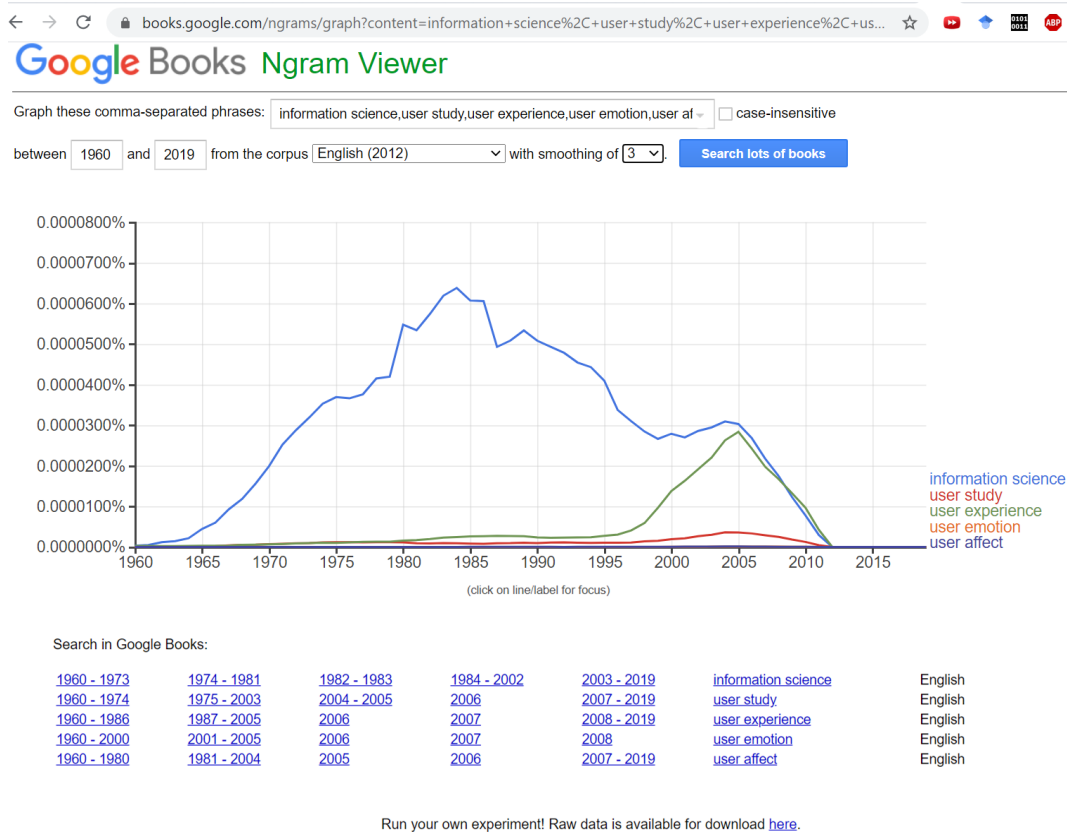
O resultado da consulta no *Google Books Ngram Viewer* (Figura 1) mostra que o termo ciência da informação (*information science*) surge nos livros a partir de 1960, com o surgimento da Ciência da Informação (CI) (BORKO, 1968) e com a evolução e expansão do uso de computadores.

No início, a CI centrava-se nos sistemas de armazenamento e recuperação da informação (centrada no sistema). Como observa Araújo (2018), a CI vem mudando e, subáreas foram consolidando conforme as alterações vivenciadas pela humanidade, sobretudo (mas não só) com o desenvolvimento das tecnologias. As tecnologias solucionaram uma série de problemas, mas trouxe outros problemas relativos às questões humanas (sociais, culturais, políticas, econômicas, jurídicas).

No fim da década de 70 começaram a surgir os sistemas de informação centrados nos usuários. Cunha e outros (2015, p.49), abordam os aspectos da mudança do foco nos estudos de usuários, que antes era centrado no sistema e, a partir do fim da década de 70, passa a ser centrado no usuário. As situações e contextos de uso dos sistemas, a cognição, a interatividade e interesses dos usuários são priorizados.

E, a partir de 2000, a explosão informacional provocou o crescimento de estudos voltados para a experiência do usuário, considerando aspectos cognitivos, afetivos e emocionais dos usuários (NORMAN, 2004), (HASSENZAHN, 2008).

Figura 1. Resultado da consulta no *Google Books Ngram Viewer* com os termos: *information science, user study, user experience, user emotion, user affect*.



Fonte: <https://books.google.com/ngrams>. Acesso em: 15 jun. 2020.

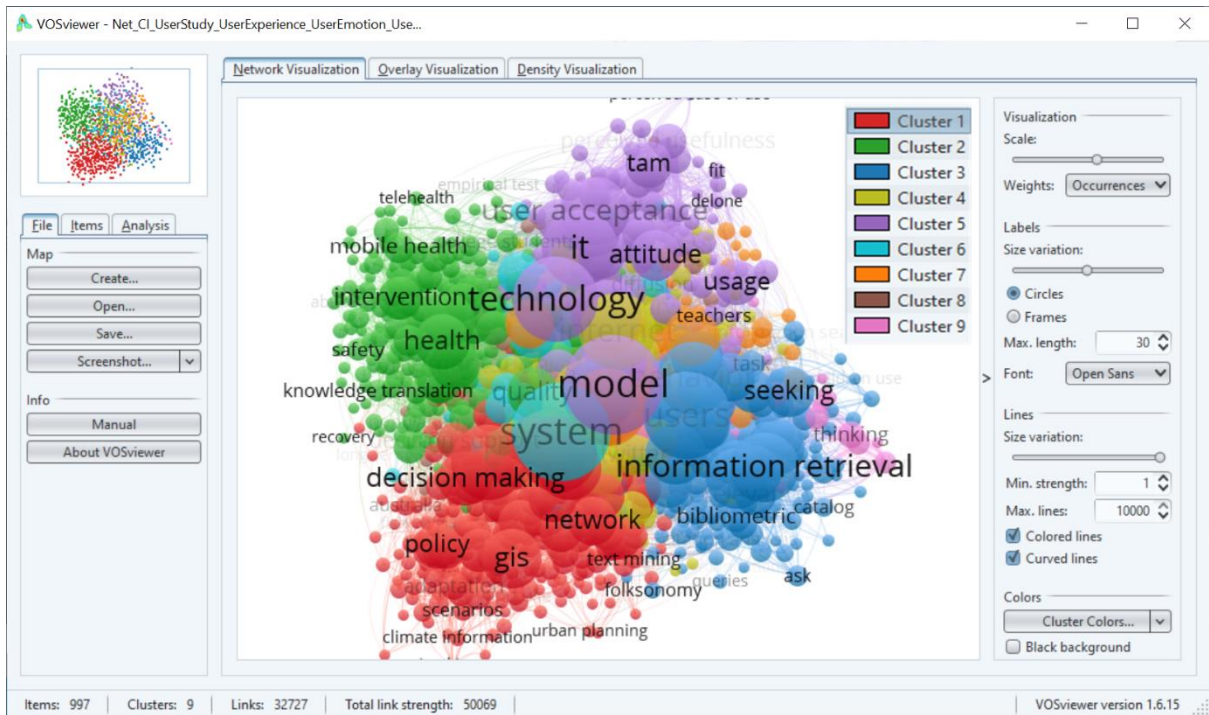
4.1 Rede de Co-ocorrência de Palavras-chave

A rede de co-ocorrência de palavras-chave foi construída para responder a questão: (i) quais são as palavras-chave que mais ocorrem e como estão relacionadas? A resposta desta questão é apresentada na Figura 3 e no Quadro 3.

Usando técnicas de mineração de texto e de processamento de linguagem natural (Apache *OpenNLP*)⁶, o VOSviewer extrai palavras-chave dos metadados (título, resumo e lista de palavras-chave) das publicações científicas. Uma palavra-chave é definida como uma sequência de substantivos e adjetivos (terminando com um substantivo).

⁶ <http://opennlp.apache.org>

Figura 2. Rede das 997 palavras-chave (All Keywords) com no mínimo 5 ocorrências.









Fonte: as autoras

A Figura 2 mostra a rede de co-ocorrências das 997 palavras-chave (*All Keywords*), com no mínimo 5 ocorrências, extraídas do título, do resumo e da lista de palavras-chave dos 5427 documentos da *WOS*. Na rede, os nós e palavras maiores refletem sua maior ocorrência, as cores indicam os agrupamentos em cluster e as linhas apresentam o inter-relacionamento das palavras-chave.

O Quadro 3 apresenta as principais palavras-chave em ordem decrescente de força de link por agrupamento (cluster). A partir dos clusters, pode-se supor o inter-relacionamento temático que caracterize áreas ou aplicações específicas de estudos de usuários. Na visualização da rede, pode-se verificar que os nós dos agrupamentos agrupados e próximos, caracterizando o intercâmbio e diversidade dos estudos de usuários.

Quadro 3. Principais palavras-chaves por cluster em ordem decrescente de força de link.

	Cluster 1 – 249 palavras - <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 palavras - <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 palavras - <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>

 <p>Cluster 4 – 113 palavras - <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></p>
 <p>Cluster 5 – 99 palavras - <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></p>
 <p>Cluster 6 – 67 palavras - <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></p>
 <p>Cluster 7 – 64 palavras - <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></p>
 <p>Cluster 8 – 33 palavras - <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></p>
 <p>Cluster 9 – 22 palavras - <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></p>

Fonte: as autoras.

Os dados da Figura 2 e Quadro 3 evidenciam que na Ciência de Informação, de acordo com os dados recuperados da WOS, existem 9 áreas principais de estudos de usuários, sendo cada uma delas voltadas para aspectos específicos. Nota-se a predominância dos estudos em tecnologia da informação, sistemas, Internet e Web.

4.2 Redes de Co-citação

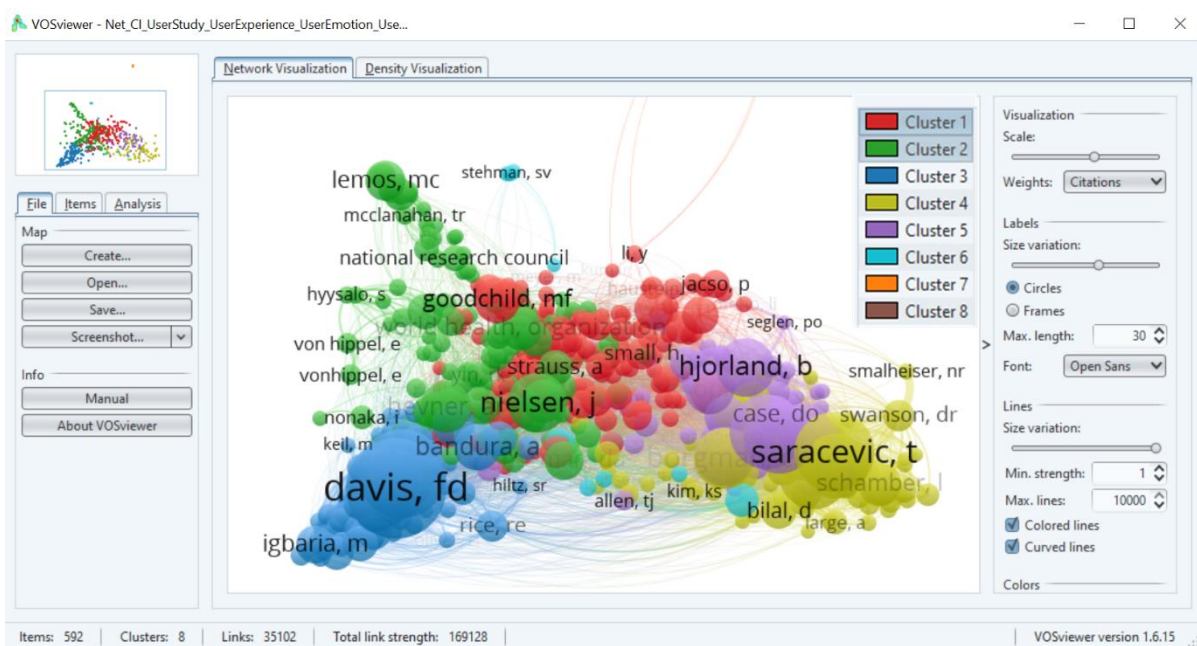
Os estudos de análise de co-citação, baseados na frequência com que dois autores ou documentos são citados de forma conjunta na produção científica de uma área, evidenciam como a estrutura de conhecimento de uma área é percebida pelos pesquisadores. Tem como princípio o fato de que, quando dois documentos ou autores são citados juntos em um trabalho posterior, existe, na perspectiva do autor citante, uma proximidade de assunto entre os citados. Assim, quanto maior a frequência de co-citação, mais próxima a relação entre esses autores citados.

No VOSviewer, foram geradas as redes de co-citação de autores, de referências e de fontes. Para construir as redes foram selecionados autores, referências e fontes com no mínimo 10 ocorrências.

4.2.1 Rede de Co-citação de Autores

A rede de co-citação de autores foi construída para responder a questão: (ii) quem são os autores mais citados na área de estudo? A resposta desta questão é apresentada na Figura 3 e no Quadro 4.

Figura 3. Rede de co-citação dos 592 autores com no mínimo 20 citações.











Fonte: as autoras

A Figura 3 mostra a rede de co-citação dos 592 autores (*first author only*), com no mínimo 20 ocorrências, extraídos dos 5427 documentos da WOS. Na rede, os nós e nome dos autores maiores refletem sua maior ocorrência, as cores indicam os agrupamentos em cluster e as linhas apresentam o inter-relacionamento dos autores.

O Quadro 4 apresenta os principais autores em ordem decrescente de força de link citação por agrupamento (cluster).

Quadro 4. Principais autores por cluster em ordem decrescente força de link de citação.

	Cluster 1 – 185 autores - <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 autores - <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 autores - <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 autores - <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 autores - <i>dervin, b; wilson, td; kuhlthau, cc; bates, mj; ellis, d; vakkari, p; hjarland, 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 autores - <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 autores - <i>zaidan, aa; zaidan, bb</i>
 Cluster 8 – 1 autor - <i>faniel, im</i>

Fonte: as autoras.

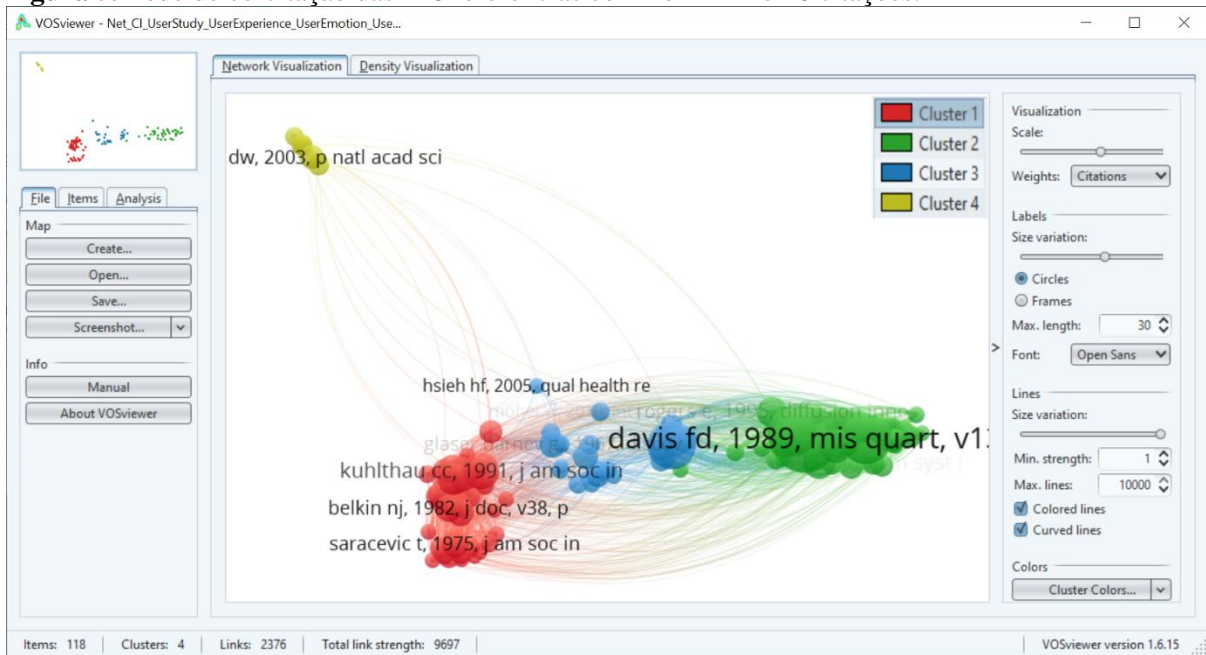
A Figura 3 e o Quadro 4, de acordo com os dados recuperados da WOS, mostram que existem 6 grupos principais que estruturam o conhecimento de estudos de usuários e como é percebida pelos pesquisadores da Ciência da Informação. Aqui o pesquisador pode visualizar e analisar a estrutura de conhecimento da área, segundo a compreensão da comunidade citante.

4.2.2 Rede de Co-citação de Referências

A rede de co-citação de referências foi construída para responder a questão: (iii) quais são as referências compartilhadas pelos autores dos documentos? A resposta desta questão é apresentada na Figura 4 e no Quadro 5.

De 177.718 referências citadas, 118 possuem no mínimo 20 citações. A Figura 4 mostra a rede de co-citação de referências extraídas dos 5427 documentos da WOS. Na rede, os nós e nome das referências maiores refletem sua maior ocorrência, as cores indicam os agrupamentos em cluster e as linhas apresentam o inter-relacionamento das referências.

Figura 5. Rede de co-citação das 118 referências com no mínimo 20 citações.



Fonte: as autoras

O Quadro 5 apresenta as principais referências em ordem decrescente de citação por agrupamento (cluster).

Quadro 5. Principais referências por cluster em ordem decrescente de citação.

	Cluster 1 – 45 itens - 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, v26, p321, doi 10.1002/asi.4630260604; schamber l, 1990, inform process manag, v26, p755, doi 10.1016/0306
	Cluster 2 – 39 itens - 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,
	Cluster 3 – 29 itens - 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;
	Cluster 4 – 5 itens - 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

Fonte: as autoras.

A Figura 4 e o Quadro 5, de acordo com os dados recuperados da WOS, evidenciam que

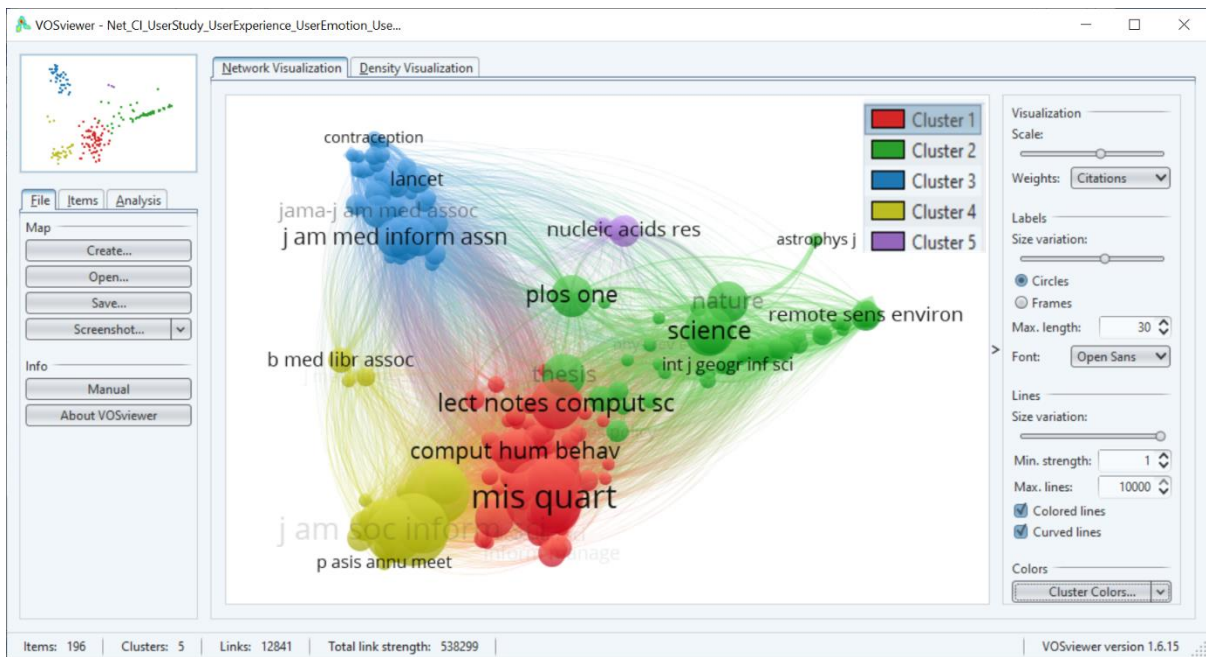
existem 4 grupos principais de referências de estudos de usuários na Ciência da Informação. Aqui o pesquisador pode visualizar e selecionar as referências mais importantes da área do conhecimento.

4.2.3 Rede de Co-citação de Fontes

A rede de co-citação de fontes foi construída para responder a questão: (iv) quais são as principais fontes dos documentos? A resposta desta questão é apresentada na Figura 5 e no Quadro 6.

De 62825 fontes citadas, 196 possuem no mínimo 100 citações. Para essas 784 fontes citadas, a força total dos links de co-citação foi calculada e mostrada na rede (Figura 5).



Figura 6. Rede de co-citação das 196 fontes com no mínimo 100 citações.






| 16

Fonte: as autoras

Quadro 6. Principais fontes por cluster em ordem decrescente da força de link.

	<p>Cluster 1 – 67 fontes - mis quart; inform syst res; manage sci; commun 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;</p>
	<p>Cluster 2 – 56 fontes - 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;</p>

 Cluster 3 – 38 fontes - 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 fontes - 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 fontes - nucleic acids res; bioinformatic; bmc bioinformatics;

Fonte: as autoras.

A Figura 5 e o Quadro 6, de acordo com os dados recuperados da *WOS*, evidenciam que existem 5 grupos principais de fontes de publicação de estudos de usuários na Ciência da Informação. Aqui o pesquisador pode visualizar e selecionar as fontes que julgar mais interessante para sua pesquisa.

5 CONSIDERAÇÕES FINAIS

Este trabalho mostrou, por meio de uma análise empírica de 5427 documentos recuperados da *WOS*, como um pesquisador pode usar as técnicas e ferramentas de Bibliometria e Ciência de Redes para realizar o levantamento bibliográfico de pesquisas.

De modo geral, o presente estudo possibilitou uma maior familiaridade com a terminologia associada aos estudos de usuários na Ciência da Informação, bem como uma visão geral da área de conhecimento.

O uso da Bibliometria e da análise de redes (palavras e co-citação), juntamente com as características de exploração visual disponibilizadas na ferramenta *VOSviewer*, facilitam a combinação de elementos para entendimento e avaliação do modo de comunicação, troca de informações e conhecimentos na área estudada. Com essas redes, o pesquisador tem uma visão geral da área do conhecimento, bem como, pode explorar contextos específicos conforme sua necessidade de informação. Acredita-se que esses procedimentos facilitam e direcionam a delimitação da pesquisa, minimizando assim a ansiedade da informação.

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.

REFERÊNCIAS

- 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=z2jvRlqhdlwC>. 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.



Artigo submetido ao sistema de similaridade

Submetido em: 27/02/2020 – Aceito em: 28/05/2020 – Publicado em: 21/06/2020
