

Desenvolvimento temático: contribuições metodológicas da bibliometria e cientometria para as revisões de literatura

Resumo

Pesquisas sobre o conhecimento científico, que têm a pesquisa científica como objeto do conhecimento, podem ser realizadas a partir de métodos e técnicas próprios da ciência da informação. Não obstante, esses métodos e técnicas recebem críticas por serem supostamente dicotômicos, descritivos e por operarem para compreensões superficiais. Questionamo-nos então: Seria possível propor um design de pesquisa que aborde a construção do conhecimento no qual se entrelacem a descrição, a compreensão e o contexto que conduz essa produção? Quais técnicas e métodos seriam utilizados? De que forma esse design de pesquisa poderia operar tomando por base perguntas de pesquisa? Na busca de contribuir para a mitigação dessas lacunas, o presente estudo tem como objetivo principal propor um design de pesquisa que possibilite um enlace das dimensões: descritiva, compreensiva e de contexto na investigação acerca das produções científicas. E como demais objetivos: apresentar e discutir técnicas e métodos subjacentes a esse design de pesquisa; explicar, a partir de possibilidades de pesquisa, como esses métodos e técnicas convergem no design de pesquisa proposto. No artigo, reúne-se uma série de métodos e técnicas, a fim de propor um design de pesquisa para a construção do conhecimento como um agente, que age em uma rede global e pode ser descrito, compreendido e analisado sob múltiplos olhares por pesquisadores e pesquisadoras. Esse desenho de pesquisa apresenta, como principais vantagens, uma forma de análise com maior profundidade, utiliza softwares livres, e nos informa o estado atual do campo de pesquisa, suas dinâmicas internas e externas e corrobora para a construção de novas perguntas de pesquisa.

Palavras-chave: metodologia; design de pesquisa; bibliometria; cientometria; teoria ator-rede.

Para citar este artigo:

ANTUNES, Jeferson; ARAIS, Estêvão Lima; PORTO, Bernadete de Souza. Desenvolvimento temático: contribuições metodológicas da bibliometria e cientometria para as revisões de literatura. *Revista Linhas*, Florianópolis, v. 24, n. 56, p. 228-252, set./dez. 2023.

DOI: 10.5965/1984723824562023228

<http://dx.doi.org/10.5965/1984723824562023228>

Jeferson Antunes

Universidade Federal do Ceará –
UFC – Fortaleza/CE – Brasil
jeferson.kalderash@gmail.com

Estêvão Lima Arrais

Universidade Estadual do Ceará –
UECE – Fortaleza/CE – Brasil
estevaolarrais@gmail.com

Bernadete de Souza Porto

Universidade Federal do Ceará –
UFC – Fortaleza/CE – Brasil
bernadete.porto@gmail.com

Thematic development: methodological contributions of bibliometry and scientometry to literature reviews

Desarrollo temático: aportes metodológicos de la bibliometría y de la cientimetría a las revisiones de literatura

Abstract

Research on scientific knowledge, which has scientific research as the object of knowledge, can be carried out using methods and techniques specific to information science. Nevertheless, these methods and techniques receive criticism for being supposedly dichotomous, descriptive and operating for superficial understandings. We then ask ourselves: Would it be possible to propose a research design that addresses the construction of knowledge in which the description, understanding and context that drive this production are intertwined? What techniques and methods would be used? How could this research design operate based on research questions? Contributing to mitigate these shortcomings, this study's main goal: presents a research design proposal that makes it possible to gather the description, understanding and context for investigating scientific productions. With the other objectives: a) present and discuss the techniques / methods for this research design; b) explain the possibilities of using methods/techniques that are associated with the proposed research design. In the article, we bring together a series of methods and techniques, proposing a research design that comprises the scientific construction of knowledge as an agent of knowledge, which acts in a global network, which can be described, understood and analyzed under multiple perspectives by researchers and researchers. This research design presents, as its main advantages, a form of analysis with greater depth, through the use of free software exemplified in the body of the text, which informs us of the current state of the research field, helping us in the construction of new research questions.

Keywords: methodology; research design; bibliometrics; scientometry; actor-network theory.

Resumen

Las investigaciones sobre el conocimiento científico, que tienen como objeto del conocimiento la investigación científica, pueden realizarse con base en métodos y técnicas propias de la ciencia de la información. Sin embargo, estos métodos y técnicas reciben críticas por ser supuestamente dicotómicos, descriptivos y por operar para comprensiones superficiales. Entonces nos preguntamos: ¿Sería posible proponer un diseño de investigación que aborde la construcción del conocimiento en el que se entrelacen la descripción, la comprensión y el contexto que conduce a esa producción? ¿Qué técnicas y métodos se utilizarían? ¿En qué manera podría operar ese diseño de investigación tomando como base preguntas de investigación? En un intento por contribuir a la mitigación de estas brechas, el objetivo principal del presente estudio es proponer un diseño de investigación que permita un enlace de las dimensiones: descriptiva, comprensiva y de contexto en la investigación acerca de las producciones científicas. Y como otros objetivos: presentar y discutir técnicas y métodos subyacentes a ese diseño de investigación; explicar, con base en las posibilidades de investigación, en qué manera convergen estos métodos y técnicas en el diseño de investigación propuesto. En el artículo, se reúne una serie de métodos y técnicas con el fin de proponer un diseño de investigación para la construcción del conocimiento como un agente, que actúa en una red global y que puede ser descrito, comprendido y analizado desde múltiples perspectivas por investigadores e investigadoras. Este diseño de investigación presenta como principales ventajas una forma de análisis más profunda, utiliza softwares libres, y nos informa del estado actual del campo de investigación, sus dinámicas internas y externas, y corrobora para la construcción de nuevas preguntas de investigación.

Palabras clave: metodología; diseño de investigación; bibliometría; cientimetría; teoría actor-red.

Introdução

As pesquisas acerca da construção do conhecimento científico são pouco usuais fora do escopo da ciência da informação. Seus resultados e informações são relevantes por contribuírem para que pesquisadoras e pesquisadores conheçam dados, análises e dinâmicas de pesquisa acerca de um campo de estudos.

Diferente dos estudos que abordam o estado da arte, estado do conhecimento ou estado da questão, essa construção teórica tem a pesquisa científica como objeto do conhecimento. Artigos científicos, livros, editoriais, resumos e conferências são dados recuperáveis, que podem ser analisados para responder perguntas de pesquisa que abordem a dinâmica da construção do conhecimento científico.

Seja através de indicadores de produção científica, estudo de citações e/ou a investigação epistemológica do campo, os estudos bibliométricos e cientométricos têm contribuído não apenas na avaliação dessa produção científica, mas também com a compreensão do horizonte ao qual as dinâmicas de construção do conhecimento encaminham seus processos (Araújo, 2006).

Esses métodos e técnicas, no entanto, acabam por ser dicotomizados, considerados mais descritivos ou que buscam compreender pontos específicos das discussões sem, no entanto, abordar contextos nos quais essas produções científicas se entrelaçam (Romancini, 2017). Essa lacuna está ligada às perguntas de pesquisa e ao design de pesquisa – entendido como a trajetória construída a partir da tomada de decisão acerca da coleta, análise e interpretação dos dados (Creswell, 2012). Tal caminho, é muitas vezes, conduzido em função dos dados e não das perguntas de pesquisa a serem investigadas.

A construção do conhecimento científico necessita, portanto, ser conduzida a partir da investigação da realidade, pautada em métodos e técnicas que sistematizam essa investigação. Nesse sentido, nos questionemos: Seria possível propor um design de pesquisa que aborde a construção do conhecimento no qual se entrelacem a descrição, a compreensão e o contexto que conduz essa produção? Quais técnicas e métodos seriam utilizados? De que forma esse design de pesquisa poderia operar tomando por base perguntas de pesquisa?

Em vistas a contribuir com essas questões, o presente estudo tem por objetivo principal: propor um design de pesquisa que possibilite um enlace das dimensões descritiva, compreensiva e de contexto na investigação acerca das produções científicas. Delinearam-se enquanto objetivos específicos: apresentar e discutir técnicas e métodos subjacentes a esse design de pesquisa; e explicar – a partir de possibilidades de pesquisa – como esses métodos e técnicas convergem no design de pesquisa proposto.

Para alcançar esses objetivos, o estudo descreve métodos e técnicas, buscando explicar como eles se articulam, explorando perguntas de pesquisas a partir desses aportes, partindo de três dimensões: a descrição, a compreensão e o contexto necessário ao presente design de pesquisa.

O presente estudo nasce das experiências das autorias, na busca de opções às revisões de literatura – uma vez que são métodos exigidos nos programas de pós-graduação durante o processo de escrita científica –, mas acabam por construir, por vezes, conteúdos repetitivos no âmbito das linhas de pesquisa quando se comparam teses e dissertações com referencial teórico similar. Se justifica, portanto, por ofertar novas possibilidades de reflexão teórica – para além de teses e dissertações – em que a compreensão da produção científica como objeto de pesquisa contribui para entender também a dinâmica dessas produções.

Metodologia

Para o presente estudo metodológico, parte-se da bibliografia (Creswell, 2012) em vistas a compreender os elementos epistemológicos e gnosiológicos que alicerçam o conjunto do método e das técnicas acerca da metodologia proposta. Com o intuito de explicar esse conjunto, foi elaborada uma recuperação de dados a partir da base centralizada *Web of Science* (WoS), com o termo “Education AND Science”, entre 2016 e 2020.

Os dados recuperados são analisados através de técnicas da bibliometria (Araújo, 2016) e da cientometria (Silva; Bianchi, 2001), métodos quantitativos que admitem estatística descritiva e/ou inferencial no processo de análise (Agresti; Finlay, 2012). Os dados bibliométricos analisados são: quantidade de publicações, quantidade de citações,

internacionalização, idioma e áreas do conhecimento envolvidas. As técnicas cientométricas aplicadas são: mapas da ciência, redes de cocitação, nuvem de palavras, redes de palavras-chave, evolução temática e diagrama de Sankey.

No decorrer da pesquisa metodológica é explicada cada uma das técnicas, demonstrando o processo de análise e apresentando outras possibilidades de pesquisa e possíveis lacunas que o estudo possui. Nesse sentido, em vistas a não tornar o texto repetitivo, toma-se essa decisão para desenvolver as explicações como um processo, com o intuito de tornar a leitura fluída.

Referencial teórico

Para o entendimento do processo metodológico perquirido apresentam-se, nessa seção, conceitos basilares para a construção do conhecimento. A bibliometria, uma abordagem quantitativa, estrutura técnica para a obtenção de indicadores e para a avaliação da produção científica. Essa abordagem utiliza a estatística para descrever aspectos de dados como citações, *h-index*¹, *impact fator*² (IF), produção por autoria e periódicos envolvidos na publicação para demonstrar padrões acerca da produção científica (Araújo, 2016; Silva; Hayashi; Hayashi, 2011; Filippo, 2002).

A cientometria, por sua vez, também visa mensurar o progresso científico, mas com enfoque no estudo das interrelações da atividade, produtividade e progresso da ciência (Silva; Bianchi, 2001). A bibliometria busca descrever a produção, enquanto a cientometria procura compreender esse mesmo conjunto de dados através de suas relações internas e externas.

Os dados recuperados, normalmente de bases de dados centralizadas³, são analisados por uma série de métodos e técnicas como a estatística descritiva e/ou a

¹ O *h-index* é calculado de acordo com o número de citações dos artigos analisados. É utilizado tanto para pesquisadores e pesquisadoras, quanto para periódicos; um índice h_2 , por exemplo, significa que ao menos dois artigos/publicações possuem duas citações.

² O *impact fator* (IF) ou fator de impacto, é calculado a partir dos dados dos dois últimos anos do periódico em publicação. Trata-se da soma de todas as citações recebidas nos dois últimos anos, dividida pelo número de artigos publicados no periódico científico nesses dois anos. As pesquisas que são sistematizadas em bases de dados a partir dessa métrica são categorizadas como pesquisas de alto impacto.

³ Bases de dados centralizadas são portais que sistematizam o conteúdo de diversos periódicos para acesso. Elas utilizam métricas próprias para a escolha dos periódicos indexados. A exemplo temos: SciELO,

inferência estatística. A estatística descritiva, utilizada em pesquisas bibliométricas, tem como função resumir a informação em um conjunto de dados – costumeiramente fazendo uso de métricas através de medidas como moda, média e mediana; enquanto a inferência estatística, aplicada a estudos cientométricos, fornece previsões com base na amostra analisada e utiliza-se de testes de hipótese a partir das perguntas de pesquisa (Agresti; Finlay, 2012).

A bibliometria e a cientometria têm sido utilizadas para avaliar a produção do conhecimento, descrito e compreendido em suas áreas de pesquisa. Não obstante, entender essa produção, também através das inter-relações entre os atores humanos e não humanos, promove um desenvolvimento epistemológico importante, em que a teoria ator-rede pode contribuir (Latour, 2001). Tomando como base a ideia de que o conhecimento é um produto social, articulado em rede, em que interagem a sociedade, as organizações, agentes e máquinas, estes se influenciam mutuamente para ajudar-nos a entender o fenômeno social da construção do conhecimento (Cavalcante *et al.*, 2017).

Ao se partir desse pressuposto epistemológico, a bibliometria e a cientometria auxiliam na análise dos nós das redes de construção do conhecimento, através de suas conexões, associações, interconexões e articulações (Cavalcante *et al.*, 2017; Latour, 2001). Mas não permitem a percepção da construção individual e unidirecional, dada pela descrição e a compreensão, possível ao se investigar um maior contexto de análise.

Perguntas de pesquisa como: Quais são as relações existentes entre palavras-chave, autorias e instituições em um determinado campo do conhecimento? Qual novo arranjo teórico se constrói nas pesquisas de uma determinada área do conhecimento? De que forma autorias constituem redes de construção do conhecimento que privilegiam determinadas teorias? Podem ser investigadas não apenas através da pesquisa cientométrica e bibliométrica, mas exigem um novo percurso metodológico, que se amplia ao se investigar o contexto dessas produções científicas, para que se possam analisar seus processos a partir das conexões, associações, interconexões e articulações.

Resultados e discussão

Para iniciar uma pesquisa que envolve as técnicas que serão discutidas a seguir, o passo basilar é uma sistemática recuperação de dados. Essa construção requer perguntas de pesquisa bem elaboradas que culminem em objetivos claros, a serem investigados por pesquisadores e pesquisadoras. A título de exemplo, ilustra-se todo o processo descrito recuperando dados em uma base centralizada.

Procedeu-se a recuperação de dados da Principal Coleção da WoS⁴, em agosto de 2020, com utilização do *Science Citation Index Expanded*, *Social Sciences Citation Index* e *Arts & Humanities Citation Index* que compõem o *Journal Citations Report (JCR)*. A WoS indexa periódicos com grande número de citações, categorizados por área, e informa, para cada artigo, os documentos citados e os documentos que o citaram. Seu banco possui mais de 18.000 periódicos indexados. É conhecida por sua organização, atualização e comprometimento com a produção científica (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, 2018; Falagas et al., 2008).

Na pesquisa por tópicos da WoS buscou-se por “*Education AND Science*”⁵, retornando 55.066 resultados. Filtrando apenas por artigos (tipo de documento), 45.804 resultados permaneceram do total inicial. Ao adicionar o filtro campo de pesquisa “*Education Educational Research*” retornaram 15.359 resultados. Quando se estabelece o filtro temporal de 2016 até 2020, temos 5.581 resultados, refinando os dados apenas para Brasil, atinge-se 68 resultados.

A presente sequência de refinamento dos dados a serem recuperados é de suma importância, uma vez que ela garante a replicabilidade (validade externa) da pesquisa. Em nosso exemplo, tem-se como objeto de estudos: os dados de pesquisas de alto impacto sobre artigos com temas que envolvam educação e ciência, no campo de

⁴ O acesso a WoS pode ser feito, até a presente data, através do Portal de Periódicos da CAPES, conectando-se em um ponto de internet da Rede Nacional de Ensino e Pesquisa – RNP – que provê internet a todas as Universidades Federais e a algumas outras instituições – ou por acesso remoto pela Comunidade Acadêmica Federada – Acesso CAFe, disponível no Portal de Periódicos da CAPES em todos os níveis da rede pública de ensino superior.

⁵ Em todas as recuperações de dados em bases centralizadas é importante que a semântica utilizada na pesquisa, em que se incluem os operadores booleanos (AND, OR, NOT, uso de aspas e aspas duplas, parênteses, entre outros) seja respeitada e entendida, ao custo de prejudicar o processo de refinamento dos dados. Todas as bases possuem documentos indicativos que ensinam como pesquisar na base de dados.

educação e pesquisa educacional, entre 2016 e 2020, realizadas por pesquisadoras e pesquisadores brasileiros.

Os dados bibliométricos para esse conjunto de dados são: Tipo de documento – artigo, quantidade de documento – 68 artigos, *h-index* 5, quantidade de citações 107, quantidade de citações sem autocitação 102, média de citações por item 1,57, média de citações sem autocitação por item 1,5, número de publicações por ano, em que se obtêm: 2016 (15), 2017 (13), 2018 (13), 2019 (16), 2020 (11) e; número de citações por ano, a saber: 2016 (3), 2017 (7), 2018 (16), 2019 (38), 2020 (43) (Gráfico 1). Os dados iniciais, cedidos pela própria WoS⁶, descrevem de maneira superficial a dinâmica de publicação do objeto de estudos.

Gráfico 1 - Distribuição anual de publicações e citações⁷



Fonte: Elaborado pela autoria com dados da WoS, 2022.

Uma análise comparativa pode ser utilizada para demonstrar a representatividade no mundo desse tipo de publicação científica, ou ainda um estudo comparativo com as publicações de outros países. Não obstante, cabe refletir sobre um problema de pesquisa comum a esse tipo de estudo, representado pelos dados do ano de 2020.

O corrente ano (2020) ainda não foi finalizado, portanto, os dados de publicações não estão consolidados. O uso desses dados causa ruído na análise de publicações e citações e, prejudica a validade externa da pesquisa. Para uma melhor recuperação de dados, deve-se trabalhar apenas com dados consolidados, ou seja, aqueles recuperados

⁶ Esses dados podem ser encontrados na guia criar relatório de citações, guias similares existem na maioria das bases de dados. As bases podem fornecer gráficos, tabelas e diagramas prontos para o uso.

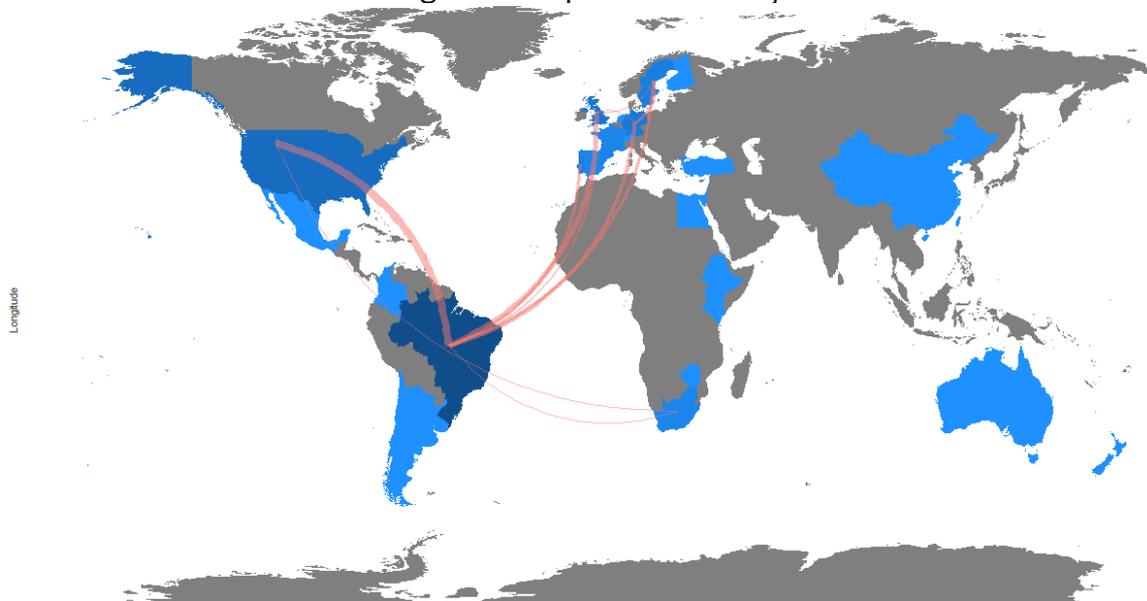
⁷ Gráficos, tabelas e diagramas são importantes para ilustrar e resumir o conjunto de dados, que devem ser descritos – para não se tornarem instrumentos capacitistas de exclusão –, e auxiliam a compreensão e posterior análise dos dados.

apenas de anos completos. Se você está realizando uma pesquisa em 2023, deve recuperar dados, no máximo, até 2022.

De posse dos dados, pode-se optar por uma abordagem compreensiva, cientométrica, que articule citações e publicações, empregando como técnicas de pesquisa a correlação linear e a regressão linear – úteis para responder algumas possíveis hipóteses de pesquisa. Nesse sentido, uma ferramenta que é amplamente utilizada na ciência da informação e na ciência de dados, é o *Google trends*, base de dados que reúne os dados estatísticos de pesquisas que são realizadas no motor de busca do Google. Correlacionar os dados de citações, publicações e buscas na internet pode sugerir relação do campo de estudos com demandas sociais, articulação com fenômenos, agendas políticas e conformação com grupos de interesse.

Quanto aos idiomas das publicações, estes são a porta de entrada para a compreensão do nível de internacionalização da discussão. Idiomas das publicações, países envolvidos nas publicações, relação de coprodução científica entre esses países e periódicos das publicações, são dados importantes para pesquisadores e pesquisadoras que desejam realizar essa discussão.

Figura 1 - Mapa de colaboração⁸



Fonte: Elaborado pela autoria com dados da WoS, 2022.

⁸ Figura gerada a partir do software livre R Studio, com uso do pacote *bibliometrix*, a partir dos dados de pesquisa.

Em nosso estudo de exemplo, 53 dos artigos estão em inglês, 14 em português e um em espanhol. Parte-se do princípio de que todos os artigos, por refinamento, têm envolvimento com o Brasil; desse modo, os países que realizaram algum nível de parceria são (Figura 1): Estados Unidos (9), Inglaterra (6), Alemanha (4), Portugal (3), Suécia (3), Países Baixos (2), África do Sul (2), França (2), Espanha (2), Argentina (1), Chile (1), Colômbia (1), Egito (1), Etiópia (1), Finlândia (1), Israel (1), Quênia (1), México (1), Nova Zelândia (1), República Popular da China (1), Suíça (1), Taiwan (1) e Turquia (1).

Por sua vez, as parcerias firmadas de forma interinstitucional são realizadas entre Universidade Federal do Rio de Janeiro (6), Fundação Oswaldo Cruz (5), Universidade de São Paulo (5), Universidade Federal do Rio Grande do Sul (5), Centro de Educação Tecnológica do Rio de Janeiro (4), Universidade de Brasília (4), Universidade Federal da Bahia (4), Universidade Federal de Minas Gerais (4), Universidade Estadual de Feira de Santana (3), Universidade federal de São Carlos (3), Universidade Estadual do Espírito Santo (3), *University of Bremen* (3), Instituto Federal do Rio Grande do Sul (2), *Malmö University* (2), Universidade de Lisboa (2), Universidade Estadual Paulista (2), Universidade Federal de Itajuba (2), Universidade Federal de Pelotas (2), Universidade Federal do Rio Grande do Norte (2), Universidade Federal do Rio Grande (2), *University of Cape Town* (2), *University of York* (2), *Addis Ababa University* (1), *African institute for Biomedicine and Thecnology* (1), *Americam Universita of Cairo* (1).

Com esses dados é possível estabelecer uma rede de parcerias, que auxilia a compreender a dinâmica da construção do conhecimento e o processo de internacionalização, a partir de atores e das instituições. Esses dados podem ser cruzados com base nos nomes das pesquisadoras e dos pesquisadores, para sobrepor a rede levando em conta grupos de pesquisas cadastrados no Brasil⁹ e em outros países, em vistas a compreender o contexto dessas conexões. Na base de dados constam também as agências financiadoras e os números de subsídios para essas pesquisas, que podem ser planos de fundo do contexto discutido com o intuito de potencializar as considerações político e/ou econômicas ante a construção do conhecimento.

⁹ Esses dados podem ser recuperados a partir da Plataforma Lattes em associação ao Diretório dos Grupos de Pesquisa no Brasil – DGP, ambas mantidas pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Ainda em nível descritivo, as áreas de conhecimento envolvidas na produção científica, apresentam dados importantes quanto à possibilidade interdisciplinar de construção do conhecimento. A WoS categoriza todos os documentos indexados em 224 áreas do conhecimento, também utilizadas pelo JCR e sistematizadas por IF divididos em quartis; outras bases de dados têm métricas similares.

Figura 2 - Áreas do conhecimento envolvidas na produção científica¹⁰



Fonte: Elaborado pela autoria com dados da WoS, 2022.

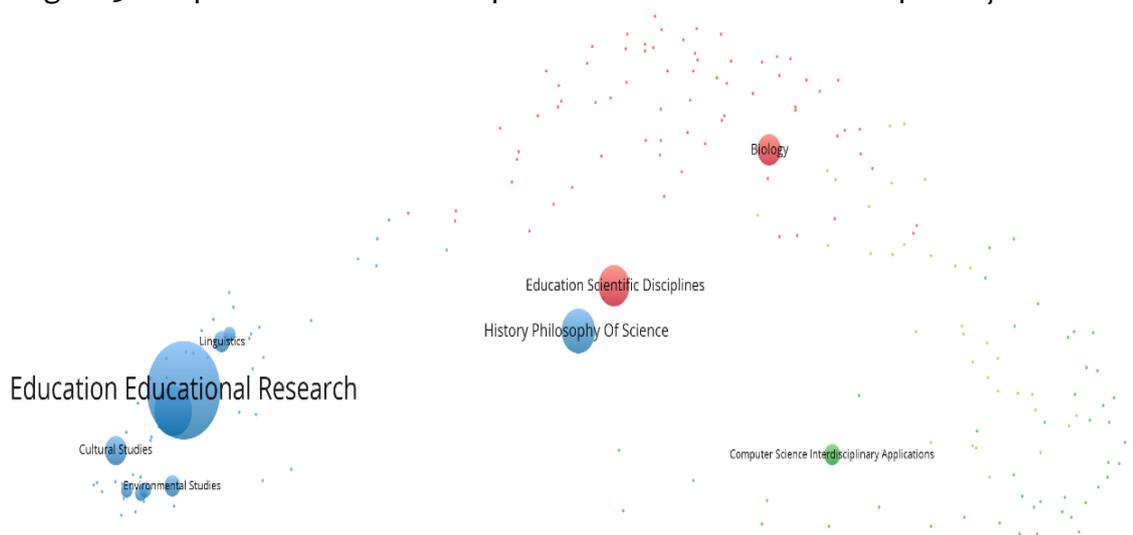
O refinamento realizado escolheu focalizar apenas artigos das áreas *Education Educational Research*, outras 14 áreas do conhecimento estão interligadas (Figura 2): *Social science interdisciplinary* (14), *History philosophy of science* (11), *Education scientific disciplines* (9), *Biology* (5), *Cultural studies* (4), *Green sustainable science technology* (2), *Language Linguistic* (2), *Linguistics* (2), *Computer science interdisciplinary applications* (2), *Environment Studies* (2), *History of social science* (1), *Business* (1), *Gerontology* (1), *Management* (1).

O ponto inicial para a compreensão dessas relações está no conteúdo dos artigos, que podem ser complementados com os dados contextuais anteriores, em suas dimensões teórica e social. Esses dados são base também para outros tipos de análise.

¹⁰ Figura gerada, de forma automática, pela plataforma WoS.

As áreas do conhecimento envolvidas nas pesquisas podem ser analisadas através de suas relações, proximidade e distanciamento, possibilitadas pelo uso dos mapas da ciência. Essa técnica permite sobrepor os dados recuperados – de maneira a produzir um diagrama de visualização espacial (mapa) – potencialmente útil na investigação da produção científica a partir da compreensão das dinâmicas internas do campo, sendo potencialmente precisos na descrição desses relacionamentos (Rafols; Porter; Leydesdorff, 2010).

Figura 3 - Mapa da Ciência criado a partir das áreas envolvidas na produção científica¹¹



Fonte: Elaborado pela autoria com dados da WoS, 2022.

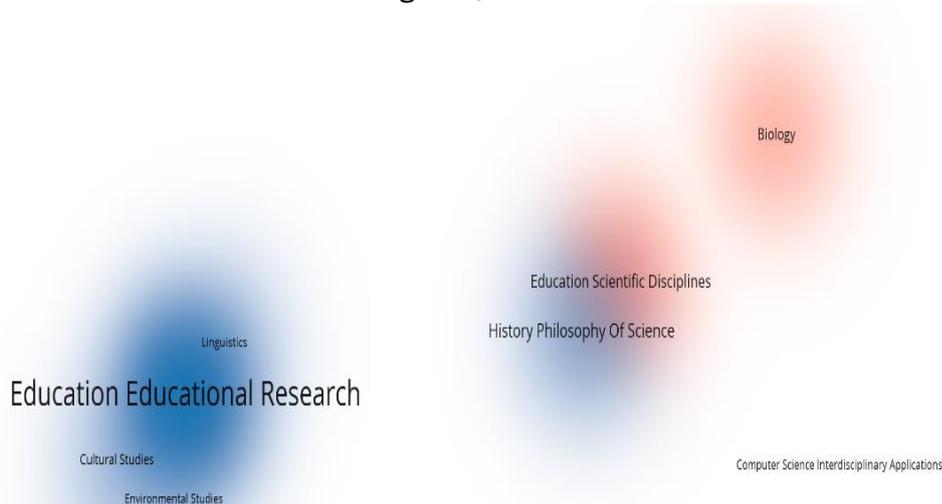
No mapa da ciência criado a partir dos dados recuperados (Figura 3) são apresentadas três medidas visuais: a força das áreas (tamanho do círculo), a distância entre as áreas (distância entre os círculos) e a cor das áreas (agrupamentos ou *clusters*). O nível inicial é descritivo, no qual se apresentam os *clusters* das áreas de conhecimento, em que descrevem sua afinidade, proximidade e possível relação interdisciplinar. Utilizando-se a teoria dos grafos, existe a possibilidade de compreender essa rede do conhecimento a partir do enlace de dados, para perceber as interações em diferentes níveis (Matheus; Silva, 2006; Netto, 2003).

¹¹ Esse mapa da ciência foi desenvolvido a partir dos dados recuperados na WoS ao utilizar o algoritmo (software livre) WC10 e o software livre VOSviewer, ambos disponíveis no portal do professor Loet Leydesdorff. Ver: <https://www.leydesdorff.net/overlaytoolkit/>.

Em nosso exemplo (Figura 3), o mapa da ciência apresenta três clusters: Vermelho - *Biology, Education Scientific Disciplines*; Verde - *Computer Science Interdisciplinary Applications*; Azul - *Business, Cultural Studies, Education Educational Research, Environmental Studies, Gerontology, History Of Social Sciences, History Philosophy Of Science, Linguistics, Management, Social Sciences Interdisciplinary*.

No todo, *Education Educational Research* é a principal área do conhecimento, centro do cluster azul, com grande proximidade da área *Social Sciences Interdisciplinary*; em volta dessa relação orbitam, equidistantes, *Linguistic*, mais próxima a *Gerontology*. *Cultural Studies, Environmental Studies; History of Social Sciences, Business e Management* também fazem parte do cluster, no entanto, são mais próximas entre si do que do centro. *History Philosophy of Science* também faz parte do cluster azul, sendo elemento de ligação com o cluster vermelho a partir do campo *Education Scientific Disciplines*, que contém ainda o campo *Biology*. Por fim, o cluster verde, *Computer Science Interdisciplinary Applications*, é distante do todo, pouco representativo e com baixa interação.

Figura 4 - Densidade dos clusters



Fonte: Elaborado pela autoria com dados da WoS, 2022.

Figura 5 - Densidade dos itens



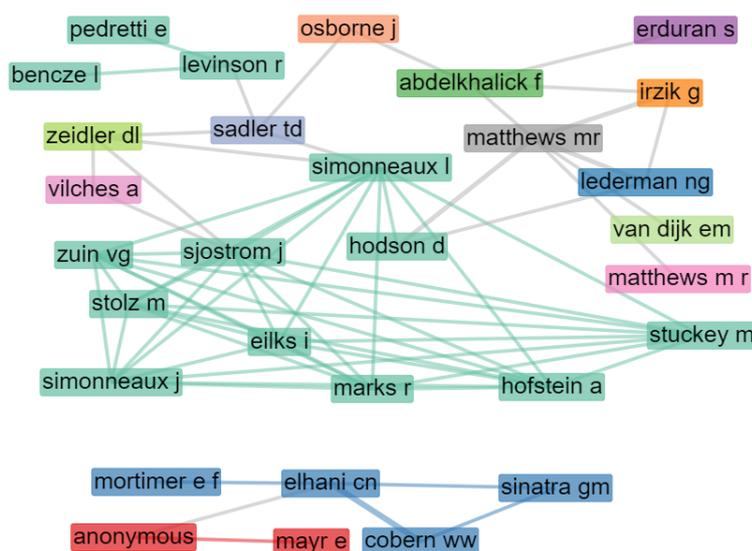
Fonte: Elaborado pela autoria com dados da WoS, 2022.

Em outro nível de compreensão dos dados, pode-se optar por analisar a densidade de cada *cluster* (Figura 4) e a densidade dos itens (Figura 5) a partir da teoria dos grafos (Matheus; Silva, 2006; Netto, 2003). Em nosso exemplo, o campo *Education Educational Research* apresenta uma maior densidade como item, seguido de *Social Sciences Interdisciplinary* (Figura 5); isso indica que são os itens com maior produção científica. O *cluster* azul apresenta uma maior densidade e centraliza a produção científica, e os campos *History Philosophy of Science* e *Education Scientific Disciplines* são elementos de ligação entre os *clusters* azul e vermelho (Figura 4).

Esse mapa da ciência apresenta como oportunidade de investigação, para além da descrição, o aprofundamento das relações que podem responder importantes perguntas de pesquisa. Analisar as relações internas dos *clusters*, as relações entre eles, a motivação de distanciamento entre *clusters*, são possibilidades a partir da leitura e sistematização dos documentos recuperados. Além disso, Rafols, Portere e Leydesdorff (2010) apresentam três outros prontos: a) compreender o periódico no qual esses documentos estão publicados como unidade de classificação, b) com finalidade propositiva para uma agenda de política de pesquisa; e c) contribuir com hipóteses de pesquisa que dizem respeito à avaliação ou inferência da produção científica.

Para se compreender os relacionamentos entre autorias, a rede de cocitações pode ser utilizada. Ela indica a formação de áreas de afinidade entre autorias através de citações, e demonstra quais autorias têm mais impacto, bem como, a formação de *clusters* que podem ser utilizados como índice de colaboração (Simatupang; Sridharan, 2004).

Figura 6 - Rede de cocitações¹²



Fonte: Elaborado pela autoria com dados da WoS, 2022.

Tomando por base nosso conteúdo de exemplo, a rede de cocitação formada (Figura 6) apresenta duas redes distintas. A primeira, maior e com mais enlace de autorias, tem centralidade em Simonneaux I. A segunda, mais restrita, tem centralidade em Elhani Ci. Esse tipo de rede pode ser analisado através de teoria dos grafos (Matheus; Silva, 2006; Netto, 2003), análise de redes sociais (Lemieux; Ouimet, 2012) ou por meio da descrição dos assuntos abordados nos artigos dessas autorias, que revela as motivações que levam ao seu estado atual. Esses dados podem ser sobrepostos a outras redes – como grupos de pesquisa, instituições, financiamento – em vista à emergência de contextos, nos quais as autorias estão inseridas, remetendo a realidade por detrás do campo científico.

¹² Essa rede foi criada utilizando o software livre R Studio, com uso do pacote *Bibliometrix*, configurando *field* como *authors*, *network layout* como *MultiDimensional Scaling*, *clustering algorithm* como *edge betweenness*, demonstrando até 50 nós de rede, com *min edges* 3, no formato *box*.

Para além da teoria dos grafos (Matheus; Silva, 2006; Netto, 2003), existe a possibilidade de análise através de medidas como centralidade de grau (*degree centrality*), a centralidade de proximidade (*closeness centrality*) e a centralidade de intermediariedade (*betweenness centrality*) (Lemieux; Ouimet, 2012). A centralidade de grau infere que o ator no centro da rede é o ator principal; a centralidade de proximidade estuda a distância entre dois atores, para mensurar o afastamento (*farness*) em vez da proximidade (*closeness*), para o qual se estabelece a hipótese de que quanto mais afastado, maior é a autonomia daquele ator (Lemieux; Ouimet, 2012). A centralidade de intermediariedade estuda o papel de coordenação e controle, a partir da hipótese de que sua relevância se dá a partir da quantidade de informação – ou nós – que perpassam o elemento de análise (Lemieux; Ouimet, 2012).

Utilizando os dados de nosso exemplo, a rede de palavras (Figura 8) apresenta cinco *clusters*, dos quais dois deles não fazem conexão com a rede principal. Em uma análise por grau de centralidade de intermediariedade, o principal elemento é *epistemology*, que realiza a ligação com os *clusters* roxo e verde, respectivamente, aos elementos *sociology* e *nature of science*, cuja influência do tema pode ser representada dentro do conjunto de dados analisado.

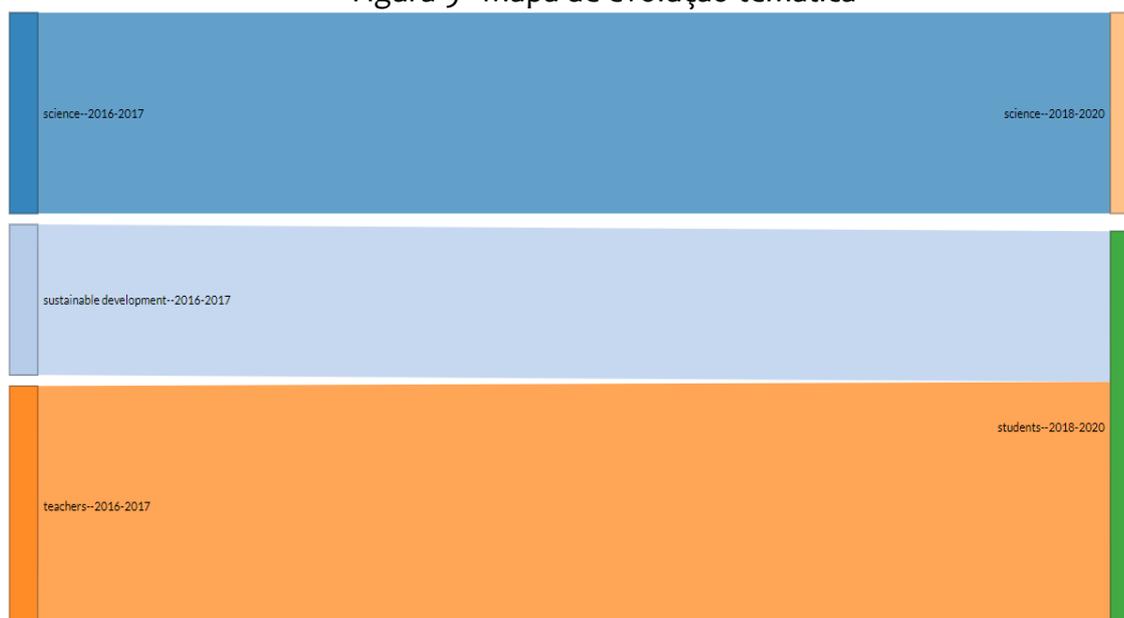
Cada um dos três *clusters* que se inter-relacionam podem ser analisados separadamente ou em conjunto, em nível descritivo, para entender os elementos mais frequentes; em nível compreensivo, ao investigar a conexão com textos e autorias; e em nível contextual, para averiguar a interligação entre autorias, instituições, linguagem e temas (a título de exemplo), com o objetivo de compreender como os atores não humanos se entrelaçam aos atores humanos.

Figura 8 - Rede de palavras-chave¹⁵

Fonte: elaborado pela autoria com dados da WoS, 2022.

Para compreender os temas abordados pelas autorias, o mapa de evolução temática revela a relação histórica entre conceitos e termos empregados e identifica os assuntos mais importantes e os mais recentes. É útil para comparar/analisar, ao longo da trajetória do campo, importantes mudanças teóricas (Aria; Cuccurullo, 2017).

¹⁵ Rede de palavras-chave criada a partir do software livre R Studio, com emprego do pacote *Bibliometrix*, aplicando os parâmetros: *field* como *author's Keywords*, *network layout* como *Fruchterman & Reingold*, *normalization* como *association*, *clustering algorithm* como *edge betweenness*, *number of nodes* como 44, *min edges* 1.

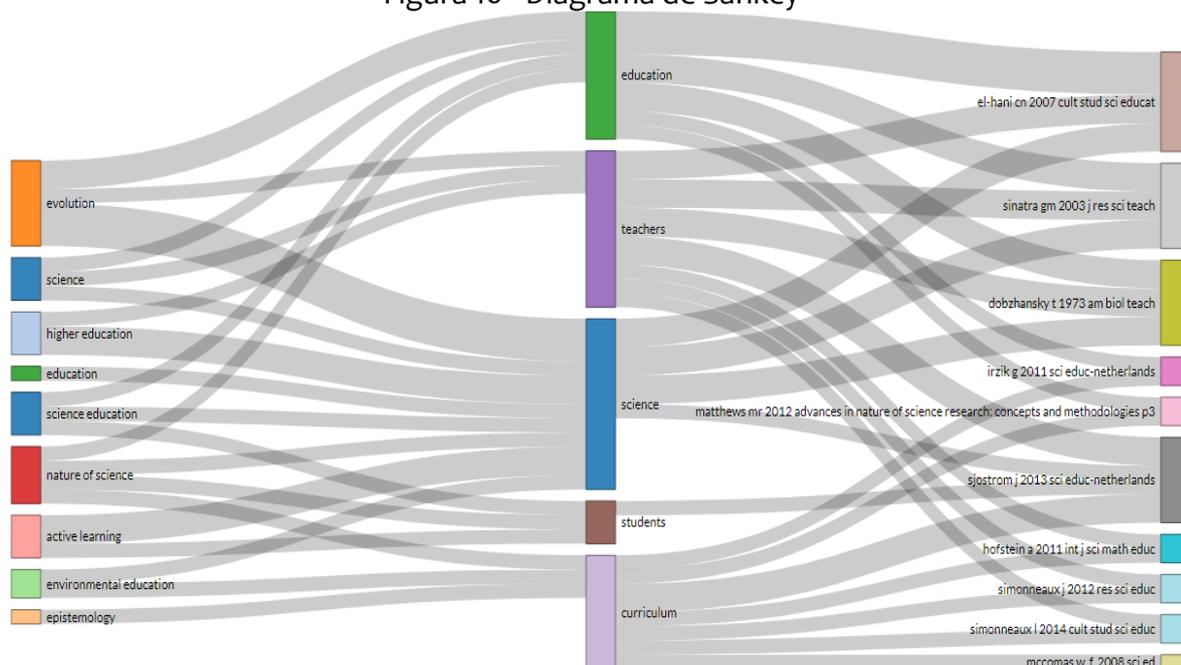
Figura 9 - Mapa de evolução temática¹⁶

Fonte: Elaborado pela autoria com dados da WoS, 2022.

Em nosso exemplo (Figura 9), o diagrama de evolução temática aponta para quatro importantes temas dentro do conjunto de dados explorado: o tema ciência, que entre os anos de 2016 e 2017 permanece relevante e sem substituições ou inserções até 2018 e 2020; já os temas desenvolvimento sustentável e professores, que estavam separados entre 2016 e 2017, se imbricam em uma nova temática, entre os anos de 2018 e 2020, na relação com estudantes. Em vista a compreensão dessa evolução temática, cabe a leitura e categorização dos documentos, a partir das palavras-chave, para um real entendimento de como se dá essa relação.

Em uma análise contextual das palavras-chave, temas, conceitos, fluxo de informações e termos, o Diagrama de Sankey, como técnica, possibilita uma série de interconexões e uma análise mais complexa dos dados. A função desses diagramas é de visualização de fluxo em várias redes e processos, ilustrando seus relacionamentos e transformações, construídos a partir do peso de entrada e saída em uma rede de acordo com seus nós (Riehmman; Hanfler; Froehlich, 2005).

¹⁶ Esse mapa foi criado a partir do software livre R Studio, com uso do pacote *Bibliometrix*, tendo como parâmetros: *field* como *keyword plus*, *number of words* como 500, *min cluster frequency* como 2, *weight index* como *stability index*, *min weighth index* entre 0,02 e 0,12.

Figura 10 - Diagrama de Sankey¹⁷

Fonte: Elaborado pela autoria com dados da WoS, 2022.

Esse tipo de diagrama pode ser utilizado para relacionar três elementos distintos entre os dados recuperados (Figura 10), em uma abordagem que busca responder questões específicas de pesquisa. Autores, afiliações institucionais, países, palavras-chave, palavras-chave adicionais, título, resumo, periódico científico, referências e periódicos citados são os elementos que podem ser contextualizados na discussão acerca dos dados analisados.

Ao observar o Diagrama de Sankey (Figura 10), tem-se na coluna da esquerda palavras-chave, na coluna central palavras-chave adicionais e, na coluna da direita, referências utilizadas pelos artigos que fazem parte do conjunto de dados. Essas interações podem ser base para contribuir, por exemplo, com a pergunta de pesquisa: Como se relacionam os temas elaborados pelas autorias dos estudos em face à apreensão de novos temas por parte de periódicos científicos quando interagem com as referências apresentadas pelas autorias?

¹⁷ Esse diagrama foi criado a partir do software livre R Studio, com emprego do pacote *bibliometrix*, aplicando os parâmetros: *middle field* com *keyword plus*, *middle field number of itens* com 5, *left field* com *keywords*, *left field number of itens* com 10, *Right field* com *References*, *Right field number of itens* com 10.

Em nosso exemplo, a utilização do Diagrama de Sankey (Figura 10) requer o estudo minucioso das referências citadas pelas autorias, ampliando os estudos recuperados da base de dados. Dessa forma, é possível compreender a relação entre os temas e os temas adicionais, e de que forma esses três níveis de informação se conectam, articulam e interagem na construção do conhecimento.

A elaboração de Diagramas de Sankey deve ser utilizada, de forma específica, a partir de elementos que contribuam com as perguntas de pesquisa apresentadas. Uma vez que exigem um maior nível de abstração e compreensão, o que deságua na análise desses dados e leva em consideração elementos específicos entre os dados recuperados – em um esforço de interlocução da teoria sobre a teoria.

Uma questão limitante quanto ao volume de dados a ser analisado utilizando as técnicas do Diagrama de Sankey, mapa de evolução temática e rede de palavras-chave refere-se ao tamanho da amostra. Em nosso exemplo, temos uma amostra de 68 artigos, não obstante, essa pode não ser a realidade da maioria das pesquisas.

Os softwares empregados nos auxiliam a configurar mapas e diagramas, mas existe uma preocupação teórica com perda e ruído de dados. Para mitigar esse possível problema de pesquisa, a amostragem por saturação teórica – como técnica –, ajuda a delimitar as escolhas dos dados. Tem-se por princípio que o refinamento dos dados, que são acrescentados nos diagramas e mapas, deve ser realizado até o ponto em que não exista repetição dentro das categorias as quais se pretende estudar (Fontanella *et al.*, 2011).

Conclusão

O presente estudo toma por base a bibliometria – como conjunto de técnicas descritivas; a cientometria – enquanto técnicas para se compreender o objeto de pesquisa e; a teoria ator-rede – que conduz o processo de pesquisa a partir de uma epistemologia do contexto entre os atores humanos e não humanos. Discutiu-se de forma exemplificada o design de pesquisa para o desenvolvimento temático como metodologia de pesquisa.

O esforço inicial, para tanto, está na recuperação de dados em bases centralizadas do conhecimento científico, cujo processo deve se dar de modo rigoroso, sistemático e

descrito minuciosamente para garantir maior validade externa das pesquisas que se apropriem da pesquisa de desenvolvimento temático. Os dados bibliométricos, descritivos, são úteis, mesmo que a título de informação, e podem contribuir com uma visão geral do objeto a ser investigado. Esses dados formam a base da pesquisa, fornecendo vários indicadores de comparação/análise da produção científica, que não se esgotam na discussão apresentada.

As abordagens pautadas na cientometria – de caráter compreensivo – discutem as variáveis a partir de sua integração, relação e correlação, que contribuem com testes de hipótese na construção de conhecimento. Construir esse conhecimento – tendo por base a bibliometria e a cientometria – também requer uma investigação histórica a partir do contexto das atividades humanas e não humanas, que formam redes de construção do conhecimento, visando o aprofundamento das contribuições aos distintos campos da ciência. Dessa forma, a escolha das técnicas descritas, exemplificadas e discutidas no contexto do presente artigo, deve ser feita a partir das perguntas de pesquisa que inquietam pesquisadores e pesquisadoras.

Em campos emergentes da pesquisa existe uma maior necessidade descritiva, ao passo que, em campos de múltiplos conceitos em disputa existe a necessidade de compreender essa disputa na atualidade do campo e, em campos da ciência nos quais existe relativo consenso, análises contextuais podem integrar a construção do conhecimento com atenção aos detalhes acerca das relações.

O presente estudo metodológico não tem a intenção de substituir qualquer método, técnica ou estrutura de pesquisa, mas, antes, apresentar um design de pesquisa a partir de uma racionalidade que não é limitada às especificidades e técnicas, e sim, está ancorado na tomada de decisão de pesquisadores e pesquisadoras perante seu objeto do conhecimento.

O desenvolvimento temático, como design de pesquisa, pode, portanto, ser conceituado como um enlace de métodos e técnicas que visam contribuir para a aquisição de perguntas de pesquisa a partir das relações existentes na construção teórica. É passível da conexão entre descrição, compreensão e contextualização dos elementos que envolvem a construção científica, retornando do universo de dados

recuperados às leituras específicas da teoria, reconectadas a partir de uma nova construção teórica.

Um dos desafios da pesquisa, que não limitou o estudo, mas exigiu pesquisa adicional e preparação, foi a escolha de exemplificar ponto a ponto as possibilidades de pesquisa em cada uma das técnicas discutidas. Caso contrário, entende-se que seria exigido um maior nível de abstração do processo de escrita científica e redução do potencial de aproximação com leitoras e leitores. Outro desafio, de ordem técnica, foi a escolha pela utilização apenas de software livre, que pode ser utilizado em múltiplos sistemas operacionais, para tornar mais acessível o processo de construção do conhecimento, sem restringir as possibilidades de investigação.

Para estudos futuros, a utilização do design de pesquisa apresentado para se investigar objetos do conhecimento é o ponto óbvio e, com o desenvolvimento desses pontos, uma revisão do design será possível. No que tange à divulgação científica, torna-se necessário a difusão dos elementos técnicos e do processo de pesquisa que se alicerça no desenvolvimento temático por meio de vídeos explicativos e podcasts. Esses meios visam a difusão e o entendimento entre pesquisadores, pesquisadoras, estudantes, professores e professoras de diversas áreas do conhecimento.

Referências

AGRESTI, Alan; FINLAY, Barbara. **Métodos estatísticos para as ciências sociais**. 4. ed. Porto Alegre: Penso, 2012.

ARAÚJO, Carlos Alberto. Bibliometria: evolução histórica e questões atuais. **Em Questão**, Porto Alegre, v. 12, n. 1, p. 11-32, jan./jun. 2006.

ARIA, Massimo; CUCCURULLO, Corrado. Bibliometrix: an R-tool for comprehensive science mapping analysis. **Journal of Informetrics**, Taipei, v. 11, n. 4, p. 959-975, nov. 2017.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR – CAPES. Portal de Periódicos. **Web of Science**. [Brasília, DF: Capes], c2018. Disponível em: https://www.periodicos.capes.gov.br/?option=com_pcollection&mn=70&smn=79&cid=81. Acesso em: 22 Set. 2018.

CAVALCANTE, Ricardo Bezerra; ESTEVES, Cristiano José da Silva; PIRES, Mariana Calisto de Assis; VASCONCELOS, Daniela Dias; FREITAS, Mônica de Melo; MACEDO, Antonio Sávio

de. A teoria ator-rede como referencial teórico-metodológico em pesquisas em saúde e enfermagem. **Texto contexto - enferm.**, Florianópolis, v. 26, n. 4, e0910017, 2017. Disponível em: <https://doi.org/10.1590/0104-07072017000910017>. Acesso em: 20 abr. 2022.

CRESWELL, John W. **Educational research: planning, conducting, and evaluating quantitative and qualitative research**. 4th. ed. Boston: Pearson, 2012.

FALAGAS, Matthew E.; PITSOUNI, Eleni I.; MALIETZIS, George A.; PAPPAS, Georgious. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses: scientific databases, pros and cons. **The FASEB Journal**, Rockville, v. 22, n. 2, fev. 2008.

FILIPPO, Daniela; FERNÁNDEZ, María Teresa. Bibliometría: importancia de los indicadores bibliométricos. In: EL ESTADO DE LA CIENCIA: principales indicadores de ciencia y tecnología Iberoamericanos/ interamericanos. Buenos Aires: Artes Gráfica Integradas, 2002. p. 1-10.

FONTANELLA, Bruno Jose Barcellos; LUCHESI, Bruna Moretti; SAIDEL, Maria Giovana Borges; RICAS, Janete; TURATO, Egberto Ribeiro; MELO, Débora Gusmão. Amostragem em pesquisas qualitativas: proposta de procedimentos para constatar saturação teórica. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 27, n. 2, p. 389-394, fev. 2011.

LATOUR, Bruno. Um coletivo de humanos e não humanos: no labirinto de Dédalo. In: LATOUR, Bruno. **A esperança de Pandora**. Bauru: Edusc; 2001. p. 201-237.

LEMIEUX, Vicent; OUIOMET, Mathieu. **Análise estrutural das redes sociais**. 2. ed. Lisboa: Instituto Piaget, 2012.

MATHEUS, Renato Fabiano; SILVA, Antonio Braz de Oliveira. Análise de redes sociais como método para a ciência da informação. **DataGramZero - Revista de Ciência da Informação**, Porto Alegre, v. 7, n. 2, p. 72-93, abr. 2006.

NETTO, Paulo Oswaldo Boaventura. **Grafos: teoria, modelos, algoritmos**. São Paulo: Edgard Blücher, 2003.

RAFOLS, Ismael; PORTER, Alan ; LEYDESDORFF, Loet. Science overlay maps: a new tool for research policy and library management. **Journal of the American Society for Information Science & Technology**, Syracuse, v. 61, n. 9, p. 1871-1887, ago. 2010.

RIEHMANN, Patrick; HANFLER, Manfred; FROEHLICH, Bernd. Interactive sankey diagrams. In: IEEE SYMPOSIUM ON INFORMATION VISUALIZATION 2005. Minneapolis: INFOVIS 2005.

ROMANCINI, Richard. A bibliometria na (e além da) avaliação. **Em questão**, Porto Alegre, v. 23, n. 3, p. 300-305, set./dez. 2017.

SANTOS, Raimundo Nonato de Macedo dos; KOBASHI, Nair Yumiko. Bibliometria, cientometria, infometria: conceitos e aplicações: tendências da pesquisa brasileira em ciência da informação. **Pesq. bras. Ci. Inf.**, Brasília, v. 2, n. 1, p. 155-172, jan./dez. 2009.

SILVA, José Aparecido da; BIANCHI, Maria de Lourdes Pires. Cientometria: a métrica da ciência. **Paidéia**, Ribeirão Preto, v. 11, n. 21, p. 5-10, 2001.

SILVA, Márcia Regina da; HAYASHI, Carlos Roberto Massao; HAYASHI, Maria Cristina Piumbato Innocentini. Análise bibliométrica e cientométrica: desafios para especialistas que atuam no campo. **InCID: Revista de Ciência da Informação e Documentação**, São Paulo, v. 2, n. 1, p. 110-129, jun. 2011.

SIMATUPANG, Togar M.; SRIDHARAN, Ramaswami. The collaboration index: a measure for supply chain collaboration. **International Journal of Physical Distribution & Logistics Management**, Verona, v. 35, n. 1, p. 44-62, jul. 2004.

Recebido em: 12/11/2021
Revisões requeridas: 20/01/2023
Aprovado em: 14/02/2023

Universidade do Estado de Santa Catarina – UDESC
Programa de Pós-Graduação em Educação – PPGE
Revista Linhas
Volume 24 - Número 56 - Ano 2023
revistalinhas@gmail.com