

Mapeamento da Ciência e a produção científica brasileira: uma metodologia do Observatório de Ciência, Tecnologia e Inovação

Adriana Badaró de Carvalho¹, Cesar Augusto Costa², Marcelo Augusto de Paiva dos Santos³, Ivone Alves de Oliveira Lopes⁴, Antônio da Silveira Brasil Junior⁵, Lucas Correia Carvalho⁶, Matheus Figueiredo Pimenta⁷, João Vitor Rodrigues Martins⁸

Resumo

Este artigo tem como objetivo descrever a abordagem metodológica utilizada pelo Observatório de Ciência, Tecnologia e Inovação (OCTI) do Centro de Gestão e Estudos Estratégicos (CGEE) para mapear a produção científica com participação nacional. Por meio de diferentes métodos, em especial aplicando o referencial de análise de redes, o OCTI tem investido na compreensão dos fluxos e dos novos padrões da produção científica no Brasil e no mundo. A partir da formação de *supernós* e da aplicação de

Abstract

This article aims to describe the methodological approach used by the Observatory of Science, Technology and Innovation (OCTI) of the Center for Management and Strategic Studies (CGEE) to map the national scientific production. Through different methods, in particular applying the network analysis framework, OCTI has invested in understanding the flows and new patterns of scientific production in Brazil and worldwide. From the formation of supernodes and the application

1 Mestre em Desenvolvimento Sustentável pela Universidade de Brasília (UnB). Líder do OCTI/CGEE.

2 Doutor em Astrofísica pelo Instituto Nacional de Pesquisas Espaciais (Inpe) e pós-doutor pela Universidade de São Paulo (USP) e pela Louisiana State University (LSU). Assessor técnico do CGEE e membro da equipe do OCTI.

3 Doutorando em Sociologia pela Universidade do Estado do Rio de Janeiro (UERJ). Assessor técnico do CGEE e membro da equipe do OCTI.

4 Pós-graduada em Gestão Pública pela Universidade Cândido Mendes e bacharel em Relações Internacionais pelo Centro Universitário IESB. É assessora técnica do CGEE e membro da equipe do OCTI.

5 Professor adjunto do Departamento de Sociologia e do Programa de Pós-graduação em Sociologia e Antropologia da Universidade Federal do Rio de Janeiro (PPGSA/UFRJ). Editor de Sociologia & Antropologia.

6 Doutor em Sociologia pela UFRJ e professor do Departamento de Sociologia e Metodologia das Ciências Sociais da Universidade Federal Fluminense (GSO/UFF).

7 Engenheiro de Software pela UnB. Assistente técnico do CGEE e membro da equipe do OCTI.

8 Bacharel em Ciência Política pela UnB. Assistente técnico CGEE e membro da equipe do OCTI.

métricas cientométricas, foi possível compreender a confluência entre mais de 150 áreas de pesquisa e as conexões semânticas entre as mais de 320 mil publicações disponíveis na *Web of Science*, no período de 2015 a 2020. Estas técnicas mostraram-se significativas para identificar os principais domínios temáticos, as tendências e as forças na participação brasileira ali mapeada.

of scientific metrics, it was possible to understand the confluence between more than 150 research areas and how semantics are included among the more than 320,000 publications on the Web of Science, in the period from 2015 to 2020. This technique is significant in identifying the main thematic domains, trends and strengths in Brazilian production.

Palavras-chave: Ciência. Produção científica brasileira. Análise de redes. Domínios temáticos. Mapeamento científico.

Keywords: Science. Brazilian scientific production. Network analysis. Thematic domains. Science maps.

1. Introdução

O Observatório de Ciência, Tecnologia e Inovação (OCTI), desenvolvido pelo Centro de Gestão e Estudos Estratégicos (CGEE), tem como objetivo monitorar o estado da arte, as tendências e os sinais emergentes relacionados ao ambiente de Ciência, Tecnologia e Inovação (CT&I), no Brasil e no mundo. O OCTI trabalha com a meta de identificar desafios e oportunidades para subsidiar tomadas de decisão nos âmbitos público e privado, na formulação e avaliação de programas e políticas na área de CT&I, fornecendo informações de qualidade, de maneira sistemática e contínua, para o Sistema Nacional de Ciência, Tecnologia e Inovação (SNCTI), por meio de serviços de inteligência estratégica e antecipatória.

Além dos boletins temáticos e anuais contendo os resultados dos trabalhos, publicados com a intenção de levar ao público-alvo uma amostra dos produtos e serviços oferecidos, o OCTI está investindo em uma *solução web* que, em breve, será lançada. O intuito é possibilitar ao usuário a navegação em redes e gráficos ilustrativos que trazem informações sobre a produção científica brasileira e mundial. Esse mapeamento busca identificar os principais domínios temáticos e as tendências da produção do conhecimento no Brasil e no mundo. Com os objetivos de propiciar um olhar mais amplo sobre a CT&I e a compreensão e difusão das pesquisas no SNCTI, esse ambiente eletrônico também disponibilizará representações e análises sobre Indicadores de CT&I que permitam orientar e monitorar políticas públicas no Brasil, além de avaliá-las em relação ao contexto internacional. Assim, o OCTI torna acessível um conjunto de indicadores que visam a avaliar o potencial e os gargalos das diferentes Regiões brasileiras e unidades da Federação, no que se refere aos desafios para o desenvolvimento regional e local, bem como

oferece informações a partir de várias bases de dados nacionais e internacionais de Indicadores de CT&I.

Nesse contexto, os objetivos deste artigo são apresentar e discutir as metodologias utilizadas, aprimoradas e desenvolvidas pelo OCTI para a realização dos estudos sobre o mapeamento da ciência. São abordadas as ferramentas metodológicas de análise de redes e suas aplicações para a identificação e caracterização da produção científica feita com colaboração brasileira, assim como justificativas teóricas e metodológicas que apoiam a escolha das técnicas de análise de dados utilizadas. Por fim, o OCTI exemplifica o conteúdo das discussões com a apresentação de alguns resultados de observação da ciência brasileira no período de 2015 a 2020.

2. Mapear a ciência: desafios teóricos e metodológicos

Os mapeamentos sobre a ciência estão cada vez mais presentes nos espaços destinados à comunicação nessa área. Com a ampliação das grandes bases indexadoras (como *Web of Science*, *Google Scholar* e *Scopus*) e do sistema científico em diferentes contextos nacionais, o desafio de acompanhar os esforços em CT&I tornou-se ainda mais central para a gestão pública. Esse monitoramento, muitas vezes, depende da extensão e dos recursos de obtenção de dados oferecidos pelas grandes corporações de bases científicas, desenhando, assim, importantes limitações para esse tipo de investimento.

Contudo, os aparatos da ciência – tais como a estruturação e o aperfeiçoamento de universidades; o fornecimento de canais para a divulgação e a comunicação entre pares; a possibilidade de revisão editorial; e o próprio financiamento essencial para o rumo de diferentes pesquisas – não são construídos à luz de uma única régua de desenvolvimento. Conforme expresso pelo *Manifesto de Leiden*, quase nunca os números frios expressam a dinâmica das áreas de conhecimento e de pesquisa, tampouco ajudam a compreender a diversidade de atuações e divulgações (HICKS *et al.*, 2015).

Diferentes agendas científicas encontram dificuldades para alcançar a visibilidade competitiva das grandes plataformas, considerando seus vieses de objetivos, missões e capacidade de coleta e disponibilização de dados. Além disso, os esforços sistêmicos a fim de contabilizar e, eventualmente, avaliar as contribuições da ciência também geram consequências no contexto acadêmico-científico. São os seus *efeitos sistêmicos*, para citar o próprio *Manifesto de Leiden*. Em outras palavras, os indicadores cientométricos – que buscam medir a produção, difusão

e aplicação dos conhecimentos científicos e tecnológicos - não passam incólumes dentro dos sistemas científicos. Pelo contrário: produzem desigualdades, endossam diagnósticos precipitados e influenciam estratégias, tanto políticas quanto científicas (HICKS *et al.*, 2015).

Nessa direção, torna-se imprescindível construir novas rotas para a análise do conhecimento produzido, capazes de abrir mão de pretensões rápidas e prematuras e de assumir a transparência como critério central para desenvolvimento dos indicadores cientométricos. Mediante o extraordinário volume de dados já disponíveis para análise, tal tarefa deve ser vista como um desafio necessário para conhecer melhor as diferentes ciências feitas com participação brasileira, sem taxá-las como ciências puramente nacionais. Nem sempre os desenhos territoriais são capazes de compreender os arranjos promovidos pelo contexto científico, ainda mais quando a cooperação tem o potencial de gerar benefícios a todas as partes envolvidas.

3. Metodologia de redes: um breve histórico

Em termos de construções metodológicas, é possível sintetizar as discussões presentes em Kirschbaum (2019) sobre como os avanços de métodos recentes destacam o caráter dinâmico das relações em rede. Ao traçar o histórico da metodologia de redes, o autor apresenta termos e conceitos como “*betweenness*” e “papel estrutural” como aspectos associados à capacidade de ação individual e coletiva em diferentes organismos sociais.

Para o autor, a inserção de abordagens mais relacionais busca corrigir as tendências estruturalistas, introduzindo atributos de noção mais relativa e individual. Dessa forma, analistas buscam levar em consideração aspectos como a agência humana, a cognição social individual e os processos que levam a formações de relações diversas expostas nas redes vistas (KIRSCHBUAM, 2019).

Além disso, os analistas de redes passaram a incorporar a ideia da importância da diversidade de relações estabelecidas entre os mesmos atores. Exemplos de tal mudança metodológica, visando a incorporar a complexidade de atributos individuais, são as análises de redes «multiplex», aquelas com vários tipos de relações (KIRSCHBUAM, 2019).

O trabalho de Chen (2017) justifica o uso de conceitos e ferramentas metodológicas da análise de redes para caracterizar mapeamentos científicos. Nesse sentido, surgem ferramentas - como análise de *clusters*, centralidade de *betweenness*, análise de palavras-chave e explosão de citações - centrais para identificar pontos de potencial disruptivo na produção científica (CHEN, 2017, p.9).

Tais ferramentas são utilizadas pelo OCTI para realizar o mapeamento da ciência brasileira nos últimos seis anos, por meio de uma extensiva extração de dados sobre a produção científica com a participação de pesquisadores vinculados a instituições nacionais.

4. Rede Brasil: o mapeamento da ciência brasileira a partir da *Web of Science (WoS)*

A partir das justificativas e escolhas metodológicas apresentadas no bloco anterior, o OCTI expõe os resultados de sua incursão na base *Web of Science (WoS)*, para fornecer subsídios sobre as publicações científicas de instituições nacionais no período compreendido entre janeiro de 2015 e maio de 2020. Para a extração dos dados na WoS, foram utilizados os seguintes critérios de seleção:

- i. Publicações com pelo menos um(a) autor(a) vinculado(a) a instituições brasileiras. Foi utilizado o campo CU (*Country*), referente ao país das instituições vinculadas; e a estratégia de busca foi CU= (*brazil* OR *brasil*);
- ii. Trabalhos encaixados na categoria “*articles*” da WoS; e
- iii. Produções compreendidas no período descrito anteriormente.

Atendendo a esses critérios, foram coletados 320.861 artigos, nos dias 26 e 27 de maio de 2020.

Para a aplicação da metodologia escolhida para este estudo, foram utilizados os metadados dos artigos extraídos que contêm informações como: o código WoS, autores, título, resumo, ano de publicação, áreas de pesquisa a que a revista está associada (segundo classificação da base), entre outros.

O Gráfico 1 mostra a distribuição dos artigos dentro do período. O ano de 2020 é parcial (até maio) e os artigos que não determinam o ano, intitulados *None*, são, em geral, classificados na categoria “*article; early access*” ou inserções recentes na base.

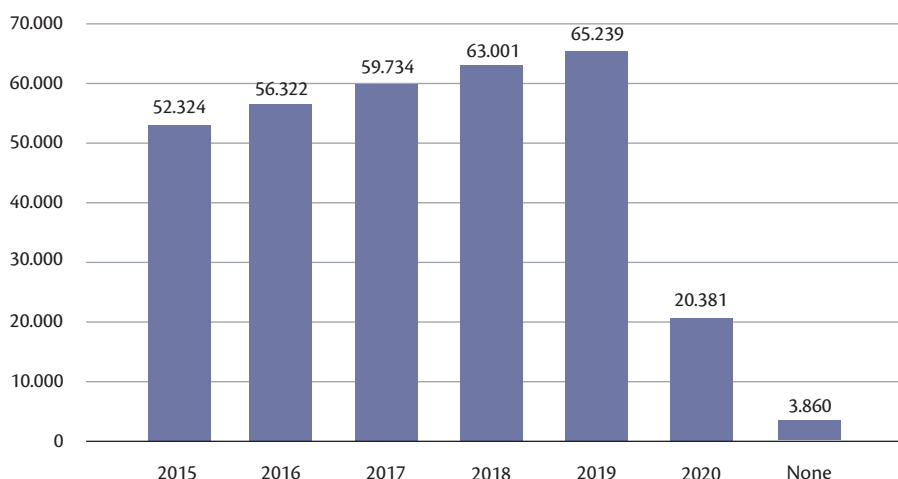


Gráfico 1. Distribuição anual da produção científica brasileira (2015 - 2020)

Fonte: Web of Science.

Observa-se, pelo Gráfico 1, o crescimento natural, ano a ano, do número de publicações com a participação de, pelo menos, um autor vinculado a uma instituição brasileira. Os artigos são distribuídos em cerca de 20 mil revistas. O Gráfico 2 mostra a distribuição de artigos nas dez revistas com maior número de publicações.

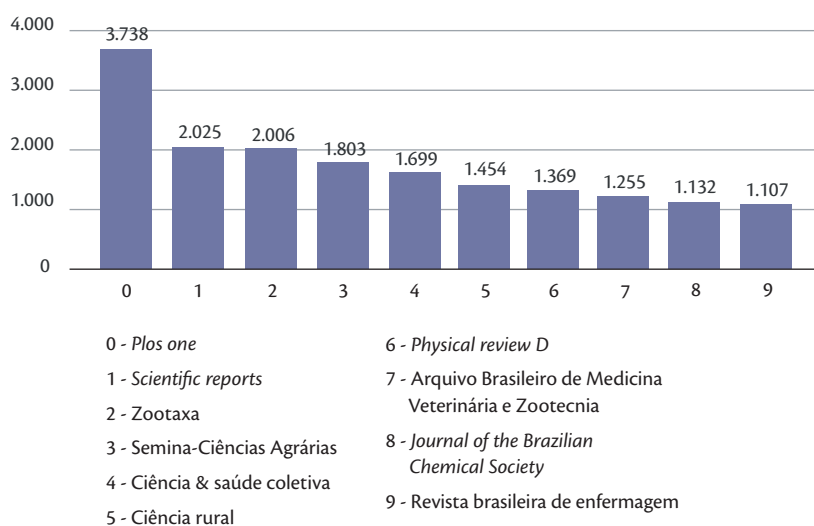


Gráfico 2. As dez revistas com maior número de publicações com participação brasileira

Fonte: Web of Science.

Foram extraídas informações textuais dos 320.861 artigos, a fim de agrupá-los de acordo com o contexto semântico dos seus resumos. Para reunir os artigos em *clusters*, isto é, agrupamentos temáticos, foram utilizadas técnicas de processamento de linguagem natural (PLN) [ou do Inglês, *natural language processing* (NLP)] no tratamento dos dados. A principal técnica envolve o cálculo do tf-idf [do Inglês, *term frequency – inverse document frequency*]. Tf-idf é uma medida estatística popular na maioria das ferramentas de busca na internet e permite determinar a similaridade entre documentos a partir da ocorrência de seus termos. O CGEE possui ferramentas customizadas para aplicar esta técnica a grandes volumes de documentos. Assim, para calcular a similaridade entre os artigos, foram utilizados os campos: título, *abstract* e palavras-chave fornecidas pelos autores.

Após aplicado o cálculo do tf-idf, um valor é atribuído a cada termo contido em cada artigo, gerando a chamada matriz termo-documento, com a qual é possível calcular o chamado cosseno de similaridade. Este cosseno quantifica quão semelhante um texto é de outro, criando, assim, uma rede de similaridade semântica. Os valores da semelhança entre dois artigos variam entre 0 e 1, sendo 0 para artigos totalmente distintos e 1 para artigos idênticos.

Esta rede de relações entre documentos (artigos) possibilita a utilização de metodologias de análise de redes⁹ para categorizar o conjunto de publicações. Os artigos são agrupados pelo algoritmo Louvain, que lhes atribui diferentes classes de modularidade, método este que tem “como objetivo separar os vértices de uma rede em módulos, com base nos tipos de relacionamentos existentes”. É uma medida que ressalta a relação dos nós (vértices) com seus vizinhos, considerando o quanto um nó tende a aparecer em determinado grupo (SAMPAIO, 2014, p.11). Os artigos (nós ou vértices) de uma mesma classe de modularidade reúnem-se para formar um agrupamento temático, dada a sua grande proximidade semântica.

No presente estudo, a quantidade de nós dentro de cada agrupamento temático da rede construída depende: das relações semânticas entre os artigos; e da aplicação do limiar inferior (ou corte) no valor da similaridade semântica a ser considerada. Para tanto, foi utilizado um algoritmo que considera o grau médio ponderado das relações de similaridade entre os artigos, obtendo-se, desta forma um limiar inferior de aproximadamente 0,3.

Por meio da aplicação deste limiar e, em seguida, do cálculo dos agrupamentos, foram obtidos cerca de 26 mil *clusters* temáticos, sendo que a grande maioria é representada por agrupamentos unitários, ou seja, por artigos isolados. Essa é uma característica comum em análises de redes semânticas com fontes científicas.

⁹ Uma rede é um grafo no qual há um conjunto de vértices (ou nós) e um conjunto de arestas (ou arcos) que conectam esses vértices. As arestas estabelecem algum tipo de relação entre dois vértices, de acordo com o problema modelado. Além disso, o grafo pode ser direcionado ou não (METZ et al., 2007).

A Figura 1 apresenta a rede de similaridade semântica formada a partir dos procedimentos metodológicos descritos anteriormente. Cada cor na rede representa um *cluster* temático e cada círculo, um artigo científico extraído. Outros artigos, em tonalidade cinza e no interior da rede, também possuem agrupamentos, mas, para facilitar a compreensão da imagem, foram coloridos apenas os maiores *clusters* e, em cinza, ficaram os menores. Outros artigos, mais periféricos, não agruparam com outras produções.

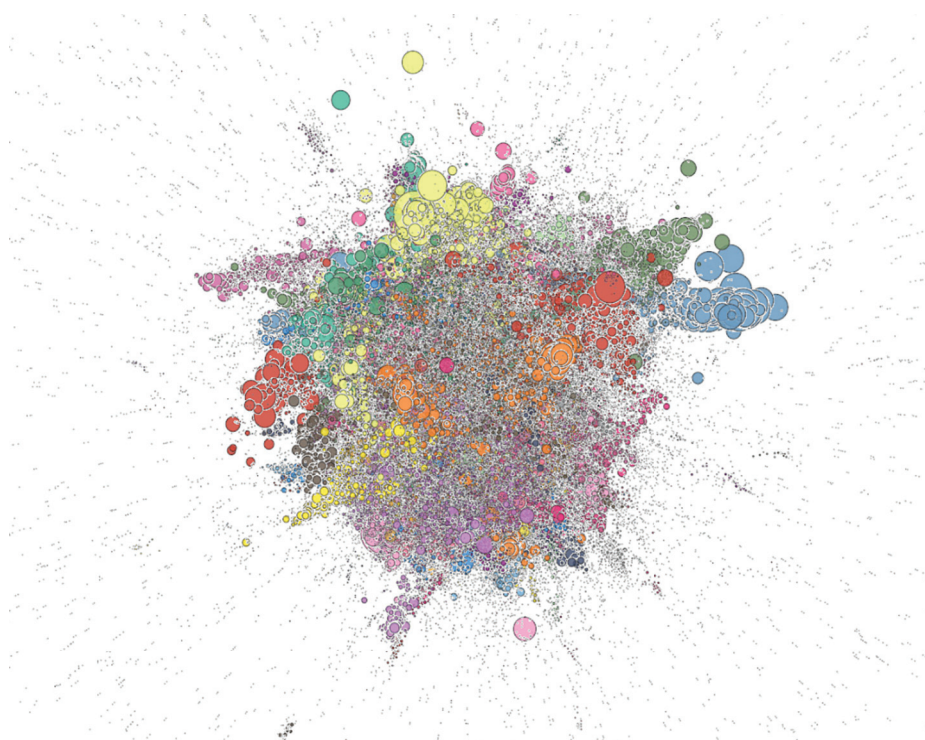


Figura 1. Rede de similaridade semântica da produção científica com participação brasileira

Fonte: *Web of Science*.

A aplicação da metodologia de redes permite verificar a fruição do conhecimento em torno de objetivos científicos compartilhados, em diferentes perspectivas. O aspecto relacional introduz a interação contextual dos vocabulários presentes nos resumos e nos títulos dos artigos extraídos. Esse efeito imprime uma interpretação mais circular da ciência com participação nacional, apontando também para os seus principais vetores. A vinculação da ciência brasileira, por exemplo, com participação de diferentes países, ressalta os objetos mais recorrentes nessas parcerias internacionais, sugerindo importantes *drivers* para a sua performance.

O benefício da **clusterização** (formação de agrupamentos) reside na capacidade de extrair informações antes não disponíveis em grandes volumes de dados não trabalhados. O reconhecimento de padrões é, assim, o objetivo final desse método, cuja execução conta com um conjunto extenso de técnicas. As contribuições da clusterização têm destaque na detecção de comunidades – *clusters* - com propriedades textuais compartilhadas e grande conexão interna. Tal estrutura reforça um grande papel na visualização de fenômenos de **formação de grupos**, isto é, de fenômenos que revelam afinidades relacionais antes não expressas por metadados individuais.

O conceito de *caminho médio* é estratégico nesse recurso metodológico: a partição da rede, com seus respectivos nós, em diferentes densidades de conexões permite identificar rotas de conexão que definem as principais comunidades. Dito de outra forma, o centro e a margem dessas comunidades dependem do caminho médio que seus nós desempenham para produzir a sua rede final de interações. Nós com grande conexão tendem a conectar-se bem em alguns agrupamentos e nós com baixa adesão imprimem buracos relacionais na teia de conexões analisada. A presença de buracos na estrutura de rede define o alcance dos grupos e a conformação da própria rede final.

Após o cálculo realizado para a geração dos agrupamentos temáticos na rede de similaridade semântica da produção brasileira (no período de janeiro de 2015 a maio de 2020), aqueles *clusters* com maiores quantidades de artigos receberam rótulos conforme seus principais termos em comum. Para isso, foi realizada uma interpretação complementar dos dados levantados de cada agrupamento, visando a detectar seus principais assuntos abordados. Esses agrupamentos, em conjunto ou isoladamente, revelam domínios temáticos com um vasto número de publicações interrelacionadas, por semelhança no uso compartilhado de termos.

Nesse sentido, foram identificados os primeiros 30 agrupamentos temáticos, em termos de quantidade de documentos, e mais cinco *clusters* considerados estratégicos¹⁰, cobrindo um total de 238.629 artigos, conforme a classificação da Tabela 1.

¹⁰ A escolha das áreas estratégicas a serem analisadas tomou como base, além dos critérios mencionados, a portaria nº 1.122, de 19 de março de 2020, que define as prioridades do então Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC) – atual Ministério da Ciência, Tecnologia e Inovações (MCTI) - para projetos de pesquisa, desenvolvimento de tecnologias e inovações no período 2020 a 2023 (BRASIL, 2020).

Tabela 1. Total de artigos científicos segundo os 35 agrupamentos temáticos selecionados

Agrupamentos temáticos	Total de artigos
Educação	16.672
Biodiversidade	14.418
Nanopartículas	12.469
Pecuária e aquicultura	10.991
Agricultura e irrigação	10.556
Saúde pública	10.336
Física teórica/gravitação	10.319
Fisiologia e esporte	9.157
Solos e lavouras	8.732
Inovação e sustentabilidade	8.594
Otimização e sistemas de energia	8.567
Compósitos, fibras e filmes	8.206
Atenção primária à saúde	8.133
Microbiologia farmacêutica	6.899
Implantes	6.540
Câncer	6.249
Geriatria	6.190
Doenças parasitárias	5.733
Física matemática	5.534
Tratamento de água e saneamento	5.489
Energias renováveis	5.133
Direitos e democracia	5.093
Internet das Coisas e cidades	4.675
Doenças neurológicas	4.422
Doenças autoimunes	4.398
Materiais metálicos	4.330
Cosmologia	4.200
Odontologia e materiais dentários	4.118
Doenças cardiovasculares e renais	4.070
Doenças infecciosas (HIV e tuberculose)	3.618
Transplantes	3.441
Germinação	3.415
Taxonomia e conservação	2.859
Violência de gênero	2.686
Arboviroses	2.387

Fonte: Elaboração própria.

A análise feita a partir do recorte descrito anteriormente possibilita, entre outros resultados, identificar os principais agrupamentos temáticos com a participação de pesquisadores em instituições nacionais, seus objetos de pesquisa mais recorrentes, suas redes de cooperação e alguns dos seus principais temas emergentes.

5. 35 agrupamentos temáticos: o perfil da ciência feita com participação brasileira

Os agrupamentos temáticos da rede de artigos revelam os principais focos científicos da produção feita com brasileiros. O conjunto vasto de domínios temáticos¹¹ presentes nesses agrupamentos aproxima diferentes áreas de pesquisa da categorização prévia da WoS e sinaliza esforços coletivos dos pesquisadores em instituições brasileiras. Para melhor compreendê-los, o OCTI aplicou uma medida que simplifica sua visualização, comprimindo todos os artigos em torno de **supernós**, isto é, de nós únicos que representam cada agrupamento identificado.

Os 35 agrupamentos temáticos selecionados (Figura 2) foram dispostos a partir de suas conexões, aproximando domínios com maior afinidade temática. Esse método permite compreender dinâmicas da produção, para além das categorizações prévias de suas áreas de pesquisa. Assim, é possível verificar sinergias das pesquisas selecionadas.

Os agrupamentos temáticos, expostos enquanto **supernós**, são definidos pela sua mutualidade de conexões internas e são limitados em sua vizinhança pela presença de buracos relacionais. Essa dinâmica sugere a presença de grupos semânticos com grande importância temática. É por essa técnica que são definidos os temas nos quais há maior dedicação de pesquisadores alocados em instituições brasileiras. Esse desenho metodológico, aplicado pelo OCTI, transforma os dados disponíveis pela WoS, apresentando grande poder de caracterização da ciência brasileira.

Na rede formada pelos 35 agrupamentos listados na Tabela 1, destaca-se o domínio temático de *Educação*, correspondente ao maior conjunto de trabalhos convergentes nos últimos cinco anos, que apresenta internamente pesquisas com diferentes objetivos e desafios no cenário educacional brasileiro. *Educação* também apresenta aproximações com outros agrupamentos temáticos, como *Direito e democracia*, *Inovação e sustentabilidade*; além de outros, como

¹¹ O termo “domínio temático” utilizado neste estudo significa o desenho de publicações inter-relacionadas por semelhança de termos que sugerem a definição de um campo de atuação conjunto em temas próximos ou convergentes. Um domínio temático pode compreender artigos que compartilham algo em comum, assim como um conjunto de relações que estes estabelecem.

Atenção primária em saúde; e Violência de gênero, que conectam fortemente a pesquisa em saúde com o campo mais geral das ciências humanas e sociais.



Figura 2. Rede dos 35 agrupamentos temáticos, formada a partir da rede de similaridade semântica da produção científica com participação brasileira

Fonte: Web of Science.

Outras comunidades de agrupamentos identificados podem ser descritas como:

- i. ligadas à pesquisa médica, biomédica, da saúde e das ciências da vida, como doenças infecciosas, transplantes, implantes, doenças parasitárias, saúde pública, câncer, doenças cardiovasculares e renais, arboviroses, doenças autoimunes, geriatria, fisiologia e esporte, assim como aqueles ligados a temas como farmácia e odontologia.
- ii. nucleadas por diferentes campos da física, como física teórica e gravitação, física matemática e cosmologia.
- iii. reunidas em torno das ciências ambientais, veterinária e ciências do solo, como biodiversidade, taxonomia e conservação, solos e lavouras. Esses agrupamentos encontram-se entrelaçados com as duas comunidades a seguir:

- a. de destaque a grandes domínios na agropecuária, como agricultura e irrigação e pecuária e aquicultura; e
- b. relacionadas à ciência dos materiais, nanociências, energia e água/saneamento (nanopartículas, compósitos, fibras e filmes, energias renováveis, materiais metálicos, tratamento de água e saneamento).

Vale destacar que o agrupamento **biodiversidade** parece ocupar um lugar central de intermediação entre os três últimos itens.

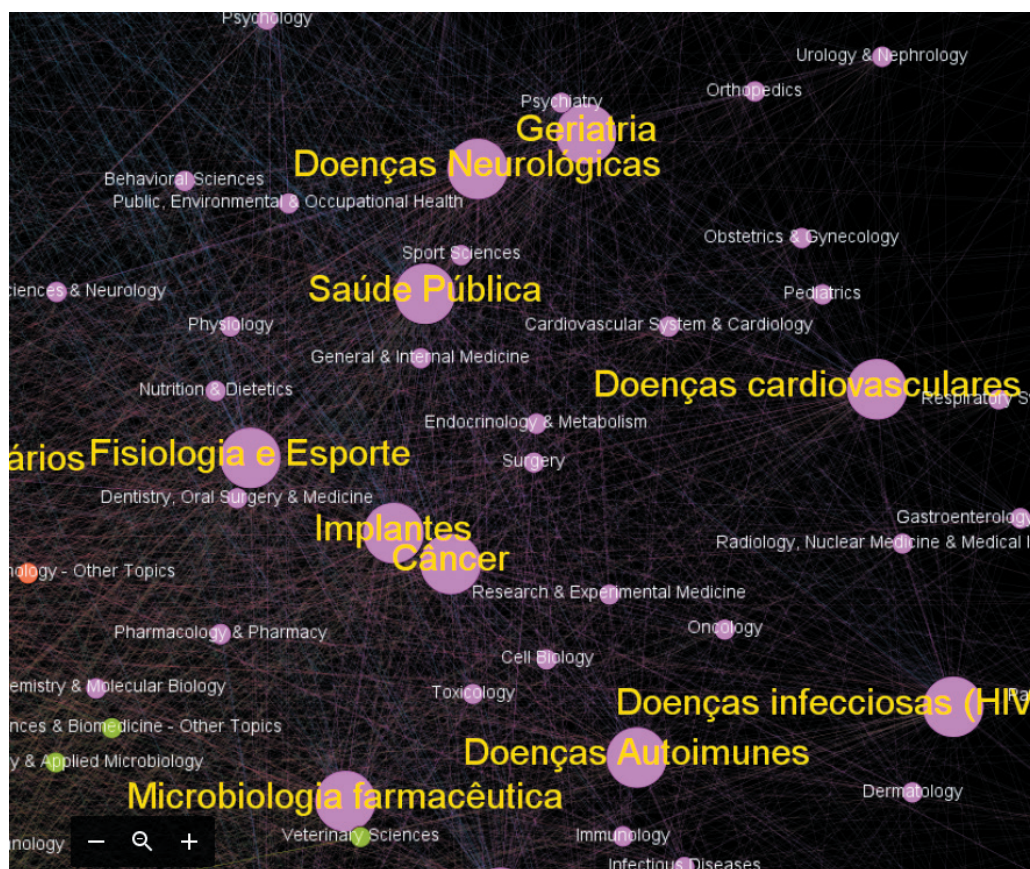


Figura 3. Recorte da rede de 35 agrupamentos temáticos da produção científica com participação brasileira, segundo domínios temáticos em saúde

Fonte: Web of Science.

Os agrupamentos temáticos são justificados pelos esforços científicos das áreas de pesquisa próximas e correlatas. A Figura 3 revela, por exemplo, que há um grande conjunto de domínios temáticos em saúde, como aqueles dedicados a pesquisar doenças autoimunes, doenças infecciosas, microbiologia farmacêutica, saúde pública, doenças cardiovasculares e doenças neurológicas. Ao redor desses temas, com grande característica de especialização, diferentes áreas de pesquisa compartilham léxicos mútuos: toxicologia, dermatologia, oncologia, cirurgia, farmacêutica e radiologia, entre muitas outras.

Em análise de redes, há distintas formas de medir a importância de um determinado elemento, dependendo do cálculo e da pergunta de pesquisa em jogo (BORGATTI, 2005). Para tanto, são utilizadas algumas medidas estruturais de centralidade.

O agrupamento *Saúde pública*, além de ser o sexto em volume de artigos, também é o mais central, de acordo com a métrica de centralidade de autovetor, que calcula a importância dos nós conforme com a qualidade das conexões (o nó conectado com outros bem conectados possuirá pontuações mais altas).

Atenção primária em saúde é outro agrupamento com alta centralidade de autovetor. Pode-se inferir, preliminarmente, que isto deve-se à forte conexão tanto com o nó de maior volume (*Educação*) quanto com o mais central (*Saúde pública*).

Ao utilizar a medida de centralidade de intermediação ou *betweenness*, que calcula os nós que estão na maior quantidade de caminhos curtos entre todos da rede – ou seja, aqueles que operam como verdadeiros atalhos – identifica-se o lugar estratégico de *Biodiversidade* e, em menor medida, de *Pecuária e aquicultura*. Em síntese, são agrupamentos temáticos que possuem conexões próximas com todas as regiões da rede, sugerindo a existência de um vetor de multi ou interdisciplinaridade em seu interior.

O desenho de redes, portanto, permite a identificação dessas sinergias entre áreas e da presença de domínios temáticos, antes invisibilizada pelos dados simples da WoS. Esses resultados revelam as forças atuantes da produção aqui analisada, abrindo caminhos metodológicos para identificar outros fatores que expandem a pesquisa em instituições brasileiras.

6. Considerações finais

Este artigo teve como finalidades apresentar e descrever uma metodologia para mapeamento temático a partir de grandes volumes de dados de artigos científicos, com o objetivo, ainda, de aprimorar a observação sobre a ciência brasileira, a partir de um melhor entendimento a respeito da sua constituição, dinâmica e suas relações inter e multidisciplinares.

Apesar da aplicação prática deste estudo no âmbito do OCTI, com o propósito de fornecer evidências que apoiem a tomada de decisões na área de ciência, tecnologia e inovação, para este artigo, a premissa foi descrever os conceitos e métodos utilizados para a construção de uma visão panorâmica e sistêmica da produção científica brasileira nos últimos cinco anos.

Sobre a metodologia desenvolvida e aplicada, pode-se concluir que a utilização de métodos de análise de redes extrapola o caráter mais fechado da indexação das pesquisas por áreas, permitindo a utilização de uma pluralidade de métricas, tanto síncronas quanto em escalas longitudinais. Um dos próximos passos desse mapeamento é a construção de interpretações mais temporais, permitindo, dessa forma, compreender o desdobramento das pesquisas, em uma visão temática, ao longo de um determinado período. A evolução ou retração desses agrupamentos, bem como a maior intensidade de conexões semânticas entre diferentes comunidades, parecem servir como chave para entender o sentido da ciência para os próximos anos.

A respeito dos resultados obtidos, o mapeamento revelou o papel estrutural da biodiversidade para diferentes aplicações temáticas, visando a atender desafios mais heterogêneos no cenário social brasileiro e mundial. Este também parece ser o caso dos artigos indexados na área de pesquisa *Engenharias*, da WoS. Essas pesquisas foram absorvidas por diferentes domínios temáticos, revelando a importância dessa área para o fomento de diferentes missões científicas. Também salta aos olhos a importância das pesquisas em *saúde* e *educação*, que ocupam lugares estratégicos no mapeamento aqui proposto.

Por fim, esses são apenas os primeiros levantamentos realizados, ainda centrados somente na base WoS. Com a ampliação do escopo de estudo do OCTI, a expectativa é a de que novos arranjos científicos sejam revelados, tanto em um aprofundamento local, como se espera no mapeamento da Scielo, quanto em uma abordagem mais internacionalizada, que busca compreender como diferentes países se aproximam em diversos contextos de produção científica.

Referências

BORGATTI, Stephen P. Centrality and network flow. **Social Networks**, v. 27, n.º 1, p. 55–71, 1 Jan 2005. Disponível em: https://aaronclauset.github.io/courses/5352/readings/Borgatti_2005_CentralityAndNetworkFlow.pdf

BRASIL. Ministério da Ciência, Tecnologia, Inovações e Comunicações. **PORTARIA Nº 1.122, de 19 de março de 2020**. Define as prioridades, no âmbito do Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC), no que se refere a projetos de pesquisa, de desenvolvimento de tecnologias e inovações, para o período 2020 a 2023. Disponível em: <https://www.in.gov.br/en/web/dou/-/portaria-n-1.122-de-19-de-marco-de-2020-249437397>

CHEN, Chaomei. Science mapping: a systematic review of the literature. **Journal of Data and Information Science - JDIS**, v. 2, n.º 2, p. 1–40, 2017. Disponível em: https://www.researchgate.net/profile/Chaomei-Chen/publication/313991204_Science_Mapping_A_Systematic_Review_of_the_Literature/links/5d12527ea6fdcc2462a625b7/Science-Mapping-A-Systematic-Review-of-the-Literature.pdf. Acesso em: 17 jan. 2021.

HAAS, P. Introduction: epistemic communities and international policy coordination. **International, Organization**, v. 46, n.º 1, Knowledge, Power, and International Policy Coordination, winter 1992, p.1-35. 1992.

HICKS *et al.* The Leiden Manifesto for research metrics. **Nature**, v. 520, p. 429-431, 2015. Disponível em: https://www.nature.com/news/polopoly_fs/1.17351!/menu/main/topColumns/topLeftColumn/pdf/520429a.pdf. Acesso em: 17 jan. 2021.

KIRSCHBAUM, Charles. Network analysis: emergence, criticism and recent trends. **RAUSP Manag. J.**, São Paulo, v. 54, n.º 4, p. 533-547, Dec, 2019. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2531-04882019000400533&lng=en&nrm=iso. Access on: 11 Feb. 2021.

KUHN, Thomas S. **A Estrutura das revoluções científicas**. 5. ed. São Paulo: Perspectivas, 1970. Disponível em: https://edisciplinas.usp.br/pluginfile.php/4103727/mod_resource/content/1/Kuhn-Estrutura-das-revolucoes-cientificas%201989.pdf

MACIAS-CHAPULA, C. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ci. Inf.**, Brasília, v. 27, n.º 2, p. 134-140, maio/ago. 1998. Disponível em: <https://www.scielo.br/pdf/ci/v27n2/macias.pdf>

METZ, J.; CALVO, R.; SENO, E.R.; ROMERO, R.A.; LIANG, Z. **Redes complexas**: conceitos e aplicações. Relatórios Técnicos do ICMC-USP. São Carlos. 2007. Disponível em: https://web.icmc.usp.br/SCATUSU/RT/BIBLIOTECA_113_RT_290.pdf

MUSSO, P. A Filosofia da rede. In: PARENTE, A. (Org.) **Tramas da rede**: novas dimensões filosóficas, estéticas e políticas da comunicação. Porto Alegre: Sulina, p. 17-38, 2004.

NEWMAN, J. The structure of scientific collaboration networks, **PNAS**, v. 98, n.º 2, p. 404-409, jan. 16, 2001. <https://doi.org/10.1073/pnas.98.2.404>

NEWMAN, M. The structure and function of complex networks. **Siam Review**, v. 45, n.º 2 p. 167–256. 2003. Disponível em: <https://epubs.siam.org/doi/pdf/10.1137/S003614450342480>

SAMPAIO, R.; Silva, L.; Esteves, E. Proposta de metodologia para análise de redes sociais aplicadas a sites de saúde. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 8, n.º 1, 2014. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/download/463/1115>

