

A INTERNACIONALIZAÇÃO DA CIÊNCIA DA INFORMAÇÃO BRASILEIRA:

uma análise temática a partir dos dados da base *Dimensions*

Fábio Castro Gouveia

 <https://orcid.org/0000-0002-0082-2392>.

✉ fgouveia@gmail.com.

🏢 Instituto Brasileiro de Informação em Ciência e Tecnologia
(Ibict) |

ROR:  <https://ror.org/006c42y96> |

🏢 Fundação Oswaldo Cruz (Fiocruz) |

ROR:  <https://ror.org/04jhswv08> | Rio de Janeiro, Brasil.

Eixo temático: Mapas e Redes na Ciência

Modalidade: Resumo expandido

DOI: 10.22477/ix.ebbc.399

Resumo: A internacionalização de um campo científico pode ser avaliada de diversas formas. Uma delas se dá pelas comunicações científicas indexadas em bases internacionais. Neste trabalho, visamos levantar a variação das temáticas em registros indexados na base *Dimensions* a partir de um recorte por autores vinculados a instituições brasileiras e delimitados pela base como sendo de *Library and information studies*. Os resultados indicam uma presença acima da média global e dentro da média internacional da produção indexada na base, no entanto com maior concentração em periódicos nacionais, e para o protagonismo dos estudos métricos da informação na internacionalização da Ciência da Informação brasileira.

Palavras-Chave: Internacionalização. *Dimensions*. *Library and information studies*. Análise de co-ocorrência. Termos em títulos e resumos.



1 INTRODUÇÃO

Podemos analisar a Ciência da Informação (CI), conforme concebida no contexto brasileiro, a partir da segmentação em seis correntes teóricas distintas propostas por Araújo (2009). Estas seriam a Teoria Sistêmica, a Teoria Crítica da Informação, as Teorias da Classificação e da Representação, os Estudos da Produção e Comunicação Científica, os Estudos de Usuários, e por último a Teoria Matemática, que inclui a Recuperação da Informação e Bibliometria, e é onde se insere este estudo. Assim, o escopo ao qual a Ciência da Informação se refere no Brasil é mais amplo do que aquele delimitado internacionalmente pela classificação *Library and Information Sciences*. Entretanto, não há como negar que há uma sobreposição da centralidade do campo da CI Brasileira quando comparada com o delimitador internacional.

A Digital Science, empresa detentora da base *Dimensions*, utiliza para classificar o conteúdo por ela indexado, o sistema de codificação de campos *Field of Research* (FoR). O FoR é parte do sistema da *Australian and New Zealand Standard Research Classification* de 2020 emitido pelo *Australian Bureau of Statistics*. Nele, o código 4610 é o que abrange a Biblioteconomia e Estudos de Informação (*Library and Information Studies*) (ABS, 2020).

De acordo com a *Australian Bureau of Statistics* (2020), este campo não inclui conteúdos de quimioinformática, bioinformática, organização e gestão de sistemas de informação, arquivo, repositório e estudos relacionados, e recuperação de informação e pesquisa na *web*. Apesar de ser mais restrito do que a definição do campo da Ciência da Informação no Brasil pela Capes (2017), entendemos que ele permite uma amostra do que se produz em comunicação científica internacionalmente em CI.

Uma das mais abrangentes bases de dados indexadoras de comunicações científicas atualmente disponível é a *Dimensions*¹. Segundo Singh *et al.* (2021), sua abrangência em termos de ISSN monitorados e indexados chega a quase 2 vezes ao da *Scopus* e mais de 6 vezes ao da *Web of Science (core collection)*. Por outro lado, a base não faz a indexação de palavras-chave de seus registros, nem atribui elas pela base, sendo qualquer análise temática limitada aos estudos de co-ocorrência de termos em títulos e resumos.

As análises de co-ocorrência são um recurso dos estudos cientométricos bastante utilizado (Oliveira, 2018; Sugimoto; Larivière, 2018). Aliado ao uso do *VOSviewer* (Van Eck; Waltman, 2010), com seu método de *network visualization* e *overlay visualization* por média de anos, podemos obter percepções sobre a evolução de temas em produções de pesquisadores afiliados à instituições brasileiras ao longo do tempo.

Assim, o objetivo central deste estudo foi analisar a partir de uma coleta na *Dimensions* os principais temas e onde ocorre o escoamento das comunicações científicas de autores com vínculos com instituições brasileiras em uma área delimitada pela base como *Library and Information Studies*. As questões a serem respondidas são: quais os principais temas abordados pelas comunicações científicas destes pesquisadores em publicações internacionais? Quais têm sido as principais publicações onde estas comunicações têm sido efetuadas?

¹ Disponível em: <http://www.dimensions.ai>. Acesso em: 10 mar. 2024.

2 METODOLOGIA

A coleta foi efetuada na base Dimensions, onde filtramos os dados por *FIELDS OF RESEARCH* (AN-ZSRC, 2020) com o código 4610 Library and Information Studies. Em seguida, os dados foram restringidos pelo campo *LOCATION – RESEARCH ORGANIZATION -> COUNTRY/TERRITORY = Brazil*. Ao tentarmos remover os *SOURCE TITLES* de publicações nacionais diretamente na base, verificamos que isto seria inviável dada a limitação para a estratégia de busca. Foi então feita a exportação dos dados em formato XLS para edição manual.

O tratamento foi então efetuado em planilha eletrônica com verificação dos títulos das fontes para remoção das que eram brasileiras, e posterior preparação de um arquivo de corpus e score para o *VOSviewer*. Também foi totalizado o montante de registros por publicação (título de fonte) para gerar uma tabela com as mais frequentes. Os arquivos de corpus e score foram abertos no *VOSviewer* v. 1.6.20 para geração de grafos de co-ocorrência de termos a partir dos títulos e resumos nos formatos de rede e de sobreposição (*Network Visualization* e *Overlay Visualization*).

3 RESULTADOS E DISCUSSÃO

Foram encontrados 517.448 registros na base para o código 4610 *Library and Information Studies*, o que representa 0,36% de todos os registros indexados na *Dimensions*. A partir do recorte para a afiliação brasileira tivemos 6.579 registros, o que curiosamente representa 0,36% dos registros de produções brasileiras indexadas na base. Além disso, neste recorte o Brasil figurava em 10º lugar, o que chama a atenção visto que o país não figura nem entre os 11 países com maior produção indexada na *Dimensions*, não sendo possível assim verificar sua posição.

Nos passos seguintes, retiramos o *Cheminform*, um boletim de resumos que consta da base e que entendemos como inadequado para este tipo de estudo por ser um agregador, resultando em 6.433 registros. Ao retirar as revistas nacionais chegamos à lista final de 2.553 registros a serem analisados, 39,69% do total de registros sem o *Cheminform*.

A Tabela 1 apresenta as principais publicações onde as comunicações científicas brasileiras da área 4610 se apresentam. Nota-se que a principal publicação onde a Ciência da Informação brasileira tem atuado internacionalmente é a *Scientometrics*. Podemos considerar então, que os Estudos Métricos da Informação lideram, ou ao menos concentram principalmente em uma publicação, a internacionalização da CI brasileira. Há também listada a *Journal of Informetrics* que também tem como foco os estudos métricos.

Vemos também que títulos ligados a computação se fazem presentes, desde a forma de séries de livros, até a de anais de conferências. É importante salientar a presença de repositórios de *preprints* na amostra, um conjunto de dados normalmente não incluído nas bases, o que é uma exceção no caso da *Dimensions*.

Também é digno de nota a presença de revistas em espanhol como a *Biblios* e a *Palabra Clave*, o mega journal *PLOS ONE*, e a *Sustainability*, uma revista não diretamente ligada à área da CI.



Tabela 1- Publicações com 20 ou mais comunicações científicas brasileiras.

Título da Fonte	QTD
Scientometrics	223
Lecture Notes in Computer Science (*)	92
Scandinavian Simulation Society	60
Biblios Journal of Librarianship and Information Science	46
Research Square (#)	38
PLOS ONE	31
Communications in Computer and Information Science	29
Palavra Clave (La Plata)	29
SSRN Electronic Journal (#)	23
Data	22
Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (\$)	22
Sustainability	21
Journal of Informetrics	20
Journal of the Association for Information Science and Technology	20
medRxiv (#)	20

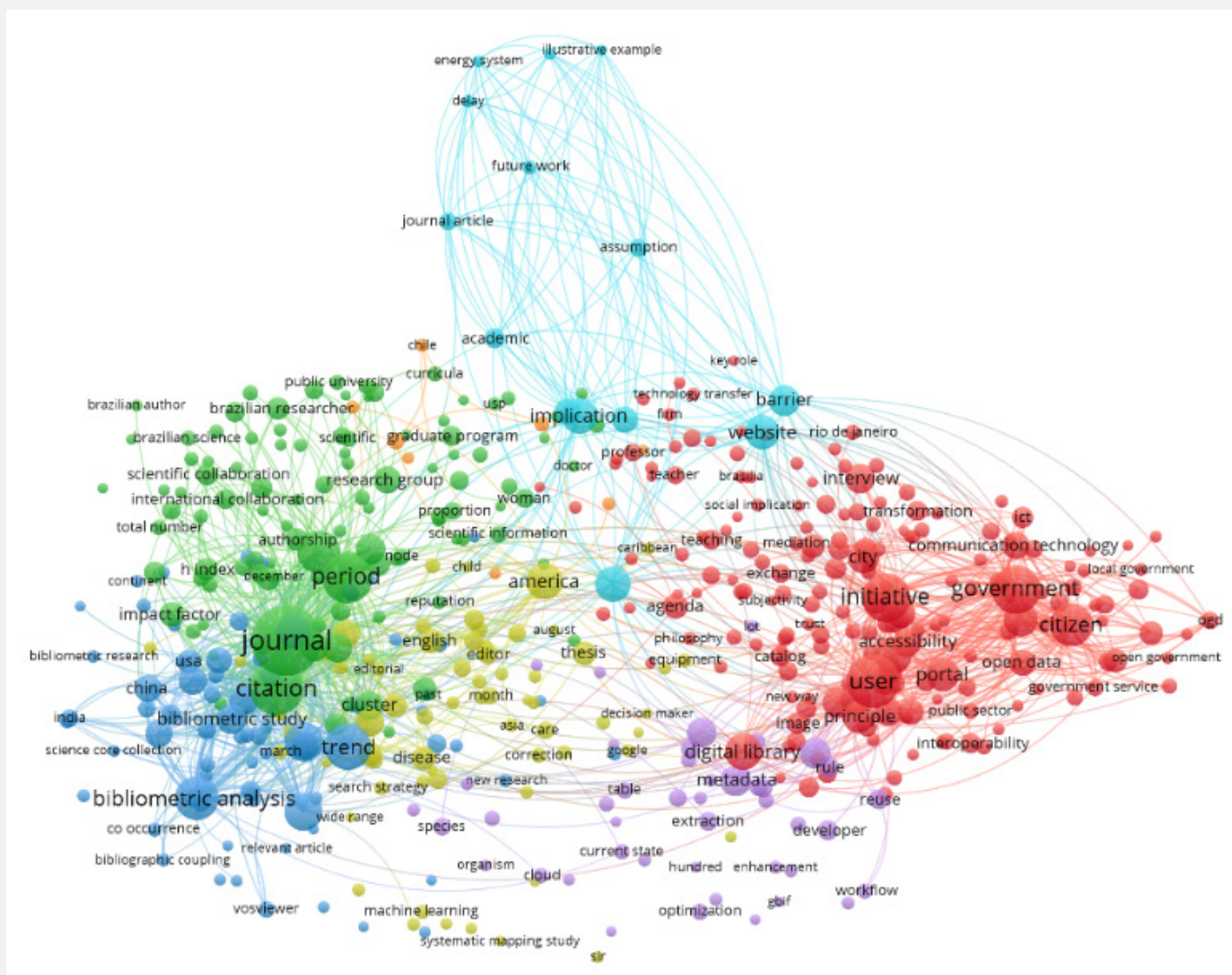
Fonte: Dados da pesquisa obtidos na base *Dimensions* (2024).

Legenda: (*) Anais de conferência, (#) Repositório de *Preprints* e (\$) Série de livros.

Para um olhar temático, os títulos e resumos foram trabalhados para compor um par de arquivos de análise de texto no formato *VOSviewer* onde os anos de publicação foram oferecidos como *score*. Os grafos estão apresentados na Figura 1 (em *Network Visualization*) onde podemos observar os agrupamentos formados, e Figura 2 (em *Overlay Visualization*) onde podemos ver a colorização pelo ano médio de ocorrência dos termos. Ambos os grafos foram, inicialmente, filtrados para o mínimo de 10 ocorrências e limitados aos 500 nós com maior *score* de relevância de acordo com o *software*.

Os termos mais centrais foram "*america*" e "*dataset*", e os agrupamentos mais claramente divididos entre grupos à direita (vermelho e roxo) focados em tecnologias da comunicação e uso de dados e o da esquerda (verde e azul) focados em temas das metrias da informação. Nos grupos da esquerda podemos ver que há uma segmentação em estudos mais focados no Brasil, na colaboração científica, fator de impacto, produção científica e estudos de citação (verde) e outro com o foco em análises bibliométricas que focam em co-ocorrências (azul). O agrupamento azul claro tem pouca adesão e parece ter termos mais genéricos que num corte para um menor número de nós levaria a sua exclusão, à exceção dos mais próximos aos outros agrupamentos. Esta distribuição de temas reforça o observado na Tabela 1 quanto à centralidade dos estudos métricos da informação na internacionalização da Ciência da Informação brasileira.

Figura 1 - Grafo de Rede dos 500 principais termos em títulos e resumos de comunicações científicas brasileiras em revistas internacionais na *Dimensions*



Fonte: Dados da pesquisa (2024).

Já na Figura 2, onde o grafo nos permite percepções sobre a linha temporal das temáticas abordadas, vemos que o que seria o equivalente ao agrupamento azul da Figura 1 apresenta grande contemporaneidade, sendo os temas mais próximos da média anual de 2020. Por outro lado, os estudos focados no Brasil parecem ter ficado mais no passado com ano médio em torno de 2014 para *"brazilian author"*, *"brazilian publication"*, *"brazilian science"* e *"brazilian researcher"*.

O termo “*dataset*” que tem centralidade na Figura 1 se situa com média do ano de 2020 na Figura 2 em meio a outros nós com médias menos recentes. O grafo interativo² representado nas Figuras 1 e 2 pode ser acessado para livre manipulação e avaliação.

O Brasil está *pari-passu* do ponto de vista quantitativo com a produção de comunicações científicas indexadas na *Dimensions* para área código 4610, mesmo considerando que as produções em revistas nacionais são da ordem de 60,31%. Este resultado, aliado a posição de 10º lugar entre os países com registros para o código 4610 mostra que a Ciência da Informação brasileira está, do ponto de vista da indexação ao menos, bem representada em uma base internacional como a *Dimensions*. Por outro lado, é nítida a concentração dessa internacionalização na temática dos Estudos Métricos da Informação, conforme pode ser visto pelo protagonismo da revista *Scientometrics*. Por mais que isto seja algo ao qual os cientometristas brasileiros possam se orgulhar, há que se considerar que esta é uma temática que tem naturalmente um bom potencial de internacionalização.

6

de refletir uma boa parte do que se entende por Ciência da Informação no Brasil, não consegue representá-la como um todo. A escolha da base *Dimensions* traz consigo seu viés de indexação, apesar de podermos também considerá-la uma boa escolha para levantamentos de produtos de comunicação científica nacionais e latino-americanos. A exclusão de periódicos nacionais pode ser melhorada no desenvolvimento da pesquisa com cruzamento com bases em que constem a indicação do país, apesar de ser necessária a verificação dos que não estiverem representados em outras bases, visto que, conforme demonstrado por Singh *et al.* (2021), a *Dimensions* é uma base muito abrangente.

A abrangência da *Dimensions* é uma de suas virtudes, mesmo que como toda a base seu viés represente uma limitação. Para uma expansão futura deste estudo, pensamos que seria interessante um aprofundamento a partir de outros códigos como o 4302 - *Heritage, archive and museum studies*, 4605 - *Data management and data science* e 4609 - *Information systems*, porém sendo necessário um trabalho de filtragem e verificação muito mais intenso, o que exigiria um investimento em uma equipe de pesquisa.

A abordagem deste estudo pode dialogar com outras avaliações sobre a internacionalização da Ciência da Informação, quer seja na forma de estudos institucionais (Rodrigues *et al.*, 2021), fontes (Mugnaini; Leite; Leta, 2011) ou revistas (Alencar; Oliveira, 2017). Acreditamos ser este um dos caminhos para a expansão desta pesquisa.

Por último, consideramos que o presente estudo traz contribuições para o entendimento dos principais temas e publicações onde tem ocorrido a internacionalização da Ciência da Informação Brasileira. Incentivos para a ampliação das temáticas representadas pela área seriam a recomendação a partir deste estudo.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico, Processos 315521/2020-1 e 315689/2023-4.

REFERÊNCIAS

ALENCAR, Maria Simone de Menezes; OLIVEIRA, Eloisa da Conceição Príncipe de. A internacionalização das coleções da scielo citation index na área de ciência da informação e biblioteconomia. **Em Questão**, v. 23, n., 2017. DOI: <http://dx.doi.org/10.19132/1808-5245230.142-158>. Disponível em: <https://seer.ufrgs.br//index.php/EmQuestao/article/view/68074>. Acesso em: 10 mar. 2024.

ARAÚJO, Carlos Alberto Ávila. Correntes teóricas da ciência da informação. **Ciência da Informação**, v. 38, n. 3, p.192-204, set./dez., 2009. Disponível em: <http://www.scielo.br/pdf/ci/v38n3/v38n3a13>. Acesso em: 01 fev. 2024.

AUSTRALIAN BUREAU OF STATISTICS (ABS). **Australian and New Zealand standard research classification (ANZSRC)**: a statistical classification used for the measurement and analysis of R&D in Australia and New Zealand, 2020. *Online*. Disponível em: <https://www.abs.gov.au/statistics/classifications/australian-and-new-zealand-standard-research-classification-anzsrc/latest-release>. Acesso em: 01 fev.



2024.

MINISTÉRIO DA EDUCAÇÃO. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Relatório da Avaliação Quadrienal 2017:** Comunicação e Informação. Coordenação de área: Mauricio Lissovsky; Coordenação adjunta: Marisa Bräscher; Coordenação adjunta de mestrado profissional: Gisela Eggert-Steindel. 2017. Disponível em: <https://www.gov.br/capes/pt-br/centrais-de-conteudo/20122017-comunicacao-quadrienal-pdf>. Acesso em: 10 mar. 2024.

MUGNAINI, Rogério; LEITE, Paula; LETA, Jacqueline. Fontes de informação para análise de internacionalização da produção científica brasileira. **Ponto de Acesso**, [S. l.], v. 5, n. 3, p. 87-102, 2012. Disponível em: <https://periodicos.ufba.br/index.php/revistaici/article/view/5684>. Acesso em: 10 mar. 2024.

OLIVEIRA, Ely Francina Tannuri de. **Estudos métricos da informação no Brasil:** indicadores de produção, colaboração, impacto e visibilidade. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2018.

RODRIGUES, Rosângela Schwarz *et al.* Internacionalização de artigos científicos: estudo dos autores de uma universidade brasileira. **Em Questão**, Porto Alegre, v. 27, n. 4, p. 215-242, 2021. DOI: <http://dx.doi.org/10.19132/1808-5245274.215-242>. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/108042>. Acesso em: 3 mar. 2024.

SINGH, Viver Kumar *et al.* The journal coverage of Web of Science, Scopus and Dimensions: a comparative analysis. **Scientometrics**, v. 126, n. 6, p. 5113–5142, jun. 2021. Disponível em: <https://arxiv.org/abs/2011.00223>. Acesso em: 10 mar. 2024.

SUGIMOTO, Cassidy. R.; LARIVIÈRE, Vincent. **Measuring research:** what everyone needs to know. New York: Oxford University Press, 2018.

VAN ECK, Nees Jan; WALTMAN, Ludo. Software survey: VOSviewer, a computer program for bibliometric mapping. **Scientometrics**, v. 84, n. 2, p. 523–538, ago. 2010. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/20585380>. Acesso em: 10 mar. 2024.