



METODOLOGIA PARA LOCALIZAR E SEGUIR A DIÁSPORA CIENTÍFICA: abordagem multimétodos

Ana Maria Carneiro

 <https://orcid.org/0000-0001-6688-1881>.

✉ anamacs@unicamp.br.

🏢 Universidade Estadual de Campinas (Unicamp) |

ROR: <https://ror.org/04wffgt70> | Campinas, Brasil.

Flavia Meireles

 <https://orcid.org/0000-0001-7471-3263>.

✉ flavia.sayane@gmail.com.

🏢 Universidade Estadual de Campinas (Unicamp) |

ROR: <https://ror.org/04wffgt70> | Campinas, Brasil.

Renata Romolo

 <https://orcid.org/0000-0001-8520-8334>.

✉ romolobrito@gmail.com.

🏢 Universidade Estadual de Campinas (Unicamp) |

ROR: <https://ror.org/04wffgt70> | Campinas, Brasil.

César Antonio Pereira

 <https://orcid.org/0000-0001-7620-7874>.

✉ cesanpereira@gmail.com.

🏢 Universidade Estadual de Campinas (Unicamp) |

ROR: <https://ror.org/04wffgt70> | Campinas, Brasil.

Clara Harumi Tatsugawa

 <https://orcid.org/0009-0001-3960-2717>.

✉ harumitatsugawa100@gmail.com.

🏢 Universidade Estadual de Campinas (Unicamp) |

ROR: <https://ror.org/04wffgt70> | Campinas, Brasil.

Eixo temático: Produtividade e Colaboração Científica

Modalidade: Resumo expandido

DOI: 10.22477/ix.ebbc.286

Resumo: O objetivo deste trabalho é apresentar uma metodologia que vem sendo desenvolvida para mapear a colaboração de pesquisadores afiliados em instituições no Brasil com membros da diáspora brasileira no exterior. A pesquisa envolveu uma abordagem multimétodos, incluindo um estudo bibliométrico na base *OpenAlex*, aplicação de um algoritmo para localizar autores de nacionalidade brasileira usando o histórico completo de publicações dos autores na *OpenAlex* e a *Plataforma Lattes*, triangulação das informações e checagem manual. A metodologia foi aplicada em artigos em cooperação internacional no tema da COVID-19. Num conjunto de 82 artigos, foi possível encontrar 60 membros da diáspora científica brasileira em 40 dos artigos.

Palavras-Chave: Comunicação Científica. Estudos métricos da Informação. Altimetria. Bibliometria. Web Social.



1 INTRODUÇÃO

O objetivo deste trabalho é apresentar uma metodologia que vem sendo desenvolvida para mapear a colaboração de pesquisadores afiliados em instituições no Brasil com membros da diáspora científica brasileira no exterior. Diáspora, conceito originalmente utilizado para referenciar movimentos de migração forçada, é aqui acionado tendo como perspectiva pesquisadores expatriados que trabalham em Ciência, Tecnologia e Inovação (CT&I) e ensino superior.

Este conceito representa uma nova compreensão do fenômeno, em oposição ao conceito de fuga de cérebros (*brain drain*). Desde a década de 1990, estudos mostram que ambos os países, não apenas o anfitrião, podem ser beneficiados (Meyer; Brown, 1999). Além disso, nem a emigração nem o retorno têm sido entendidos como situações definitivas, devido à crescente mobilidade internacional e as consequentes redes formadas entre cientistas e outros profissionais nos vários países (*brain networking*) (Ciumasu, 2010). Nesta nova perspectiva, os sujeitos diasporados passam a ser compreendidos como um recurso valioso a ser explorado em proveito do desenvolvimento do país de origem. Para tal, faz-se necessária a criação de vínculos e redes, estabelecendo sinergias entre os países de origem e suas diásporas (Carneiro *et al.*, 2020).

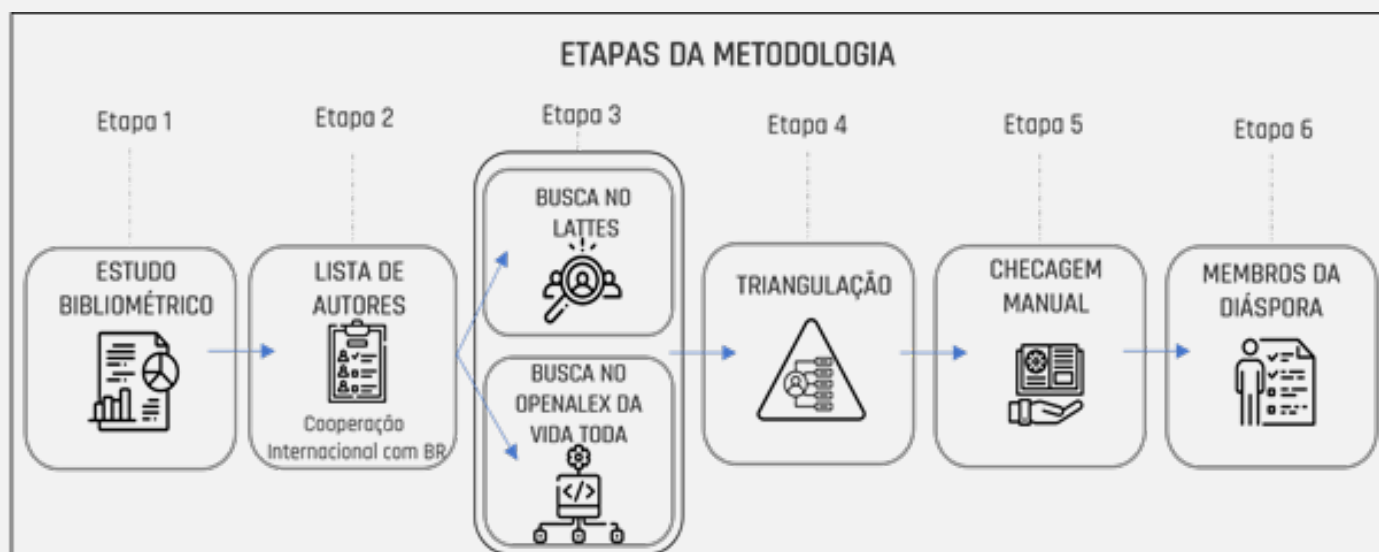
A bibliometria tem sido usada como uma ferramenta para rastrear as colaborações científicas entre instituições, a partir da informação de filiação institucional dos autores, o que inclui dados sobre o nome da instituição, unidade e endereço (Moed; Halevi, 2014). Essas informações de filiação permitem rastrear a migração científica de um país para outro e a análise das trajetórias individuais de forma longitudinal (Cañibano; Woolley, 2015). Laudel (2003), um dos pesquisadores pioneiros do estudo de fuga de cérebros no nível micro, adotou como ponto de partida da carreira e da história de mobilidade do cientista suas primeiras publicações, entendidas como decorrentes do doutorado. Este formato, entretanto, não é totalmente adequado para analisar a diáspora, pois segundo Moed e Halevi (2014), o país refere-se à localização geográfica onde o autor estava trabalhando durante o tempo no qual desenvolveu o artigo, e não a sua nacionalidade ou país de nascimento.

Nesse sentido, está em curso uma pesquisa que visa analisar a contribuição da diáspora científica brasileira na cooperação internacional na área de COVID-19, incluindo a mudança em seus padrões. Um dos desafios é identificar a mobilidade internacional dos cientistas brasileiros, sua fixação em outros países e a cooperação destes com cientistas no Brasil. A metodologia envolve a identificação da nacionalidade dos coautores de trabalhos científicos, entendendo a coautoria de trabalhos como proxy de cooperação (Maher; Van Noorden, 2021). A hipótese é que parte do bom desempenho do país na produção científica no tema resulta da cooperação entre brasileiros situados no Brasil e membros da diáspora brasileira de CT&I no exterior.

2 MATERIAIS E MÉTODOS

A pesquisa envolveu abordagem multimétodos e se constitui em seis etapas (Figura 1).

Figura 1- Etapas da Metodologia



Fonte: Elaborado pelos autores (2024).

A primeira etapa foi o **estudo bibliométrico** sobre o tema de COVID-19 entre os anos de 2020 a 2022, na base de dados *OpenAlex*. Trata-se de uma base de dados pública que contém informações bibliográficas sobre produções científicas indexadas em diversas outras fontes, incluindo autores, instituições, periódicos, citações, entre outras (Priem; Piwowar; Orr, 2022). Dentre as vantagens, destacam-se: (i) grande cobertura de fontes, cerca de 226 mil, incluindo uma melhor cobertura de trabalhos para além da língua inglesa, de autores do Sul Global e das diversas áreas do conhecimento; (ii) mecanismo de pesquisa eficiente e bem documentado; (iii) dados de uma maior diversidade de tipos de publicações; e (iv) formato livre na licença CC0 (*Creative Commons zero*), o que permite transparência e reuso.

As buscas na *OpenAlex* foram realizadas por meio do *Bigquery*, acessada por sua *API* (*Application Programming Interface*), entre 31 de maio a 10 de junho de 2023. A estratégia de busca foi elaborada a partir dos termos utilizados no *Topical Cluster 1500* da base *SCIVAL/Scopus*, abrangendo COVID-19, SARS-CoV-2 e Coronavírus. Além disso, foi realizada uma revisão de literatura sobre a pandemia para aprimorar a estratégia de busca (Cintra, 2022; Mota, Ferreira, Leal, 2020; Santos Neto, 2022; Silva *et al.*, 2022). A *query* foi composta pelas palavras-chave COVID, COVID-19, SARS-CoV-2 e Coronavírus e derivações. Foram selecionados quatro tipos de publicações: *proceedings-article*, *journal-article*, *book-chapter* e *report-series*.

A segunda etapa foi a **listagem de todos os autores dos artigos em coautoria internacional** com autores filiados a instituições no Brasil, incluindo estes últimos. Esta listagem foi a base para a terceira etapa que envolveu duas atividades simultâneas. Realizou-se a **busca destes autores na Plataforma Lattes** para checar se estavam ou estiveram filiados a instituições no Brasil em algum momento de suas trajetórias, ou se fizeram etapas de sua formação no país. Isso é possível de ser observado porque se trata de uma plataforma cumulativa, pois mesmo os currículos que não são atualizados, permanecem ali. Ao mesmo tempo, foi realizada a busca de todos os artigos dos autores da listagem na *OpenAlex*, independente do período, o que foi denominado de **busca da vida toda**. Nestes registros longitudinais foi aplicado um algoritmo para identificar se em algum momento de suas trajetórias estiveram filiados a instituições no Brasil. Como resultado da terceira etapa, foi identificada uma listagem preliminar de

potenciais membros da diáspora científica brasileira.

Na quarta etapa, foi feita a **triangulação** de todas as informações obtidas sobre todos os autores dos artigos que levaram à criação da listagem preliminar de membros da diáspora científica brasileira (artigos com potencial de ter membros da diáspora). A planilha resultante inclui as seguintes variáveis principais: Identificadores de Objetos Digitais (*Digital Object Identifier* - DOI), Identificadores Únicos de Pesquisador (*Open Researcher and Contributor ID* - ORCID), nome do pesquisador, outros nomes do pesquisador usados em citações, última filiação institucional no exterior, país da instituição, ID do Lattes e nacionalidade.

A quinta etapa foi uma ampliação da etapa anterior, centrada agora não nos potenciais membros da diáspora, mas nos próprios artigos que potencialmente contêm contribuições de membros da diáspora, e consistiu na **checagem manual** das informações reunidas para classificar todos os coautores dos artigos em termos da nacionalidade e cooperação com o Brasil. Como se trata de uma etapa que demanda muito tempo, foi selecionada uma amostra de 10% dos artigos com maiores índices de visualizações e citações segundo a base Dimensions. Foram buscados detalhes como país de realização da formação e filiação empregatícia no Brasil. As principais fontes foram *Orcid*, *Lattes*, *Linkedin*, *sites* das instituições de filiação dos pesquisadores e pesquisa direta na plataforma Google, além das informações contidas em cada artigo.

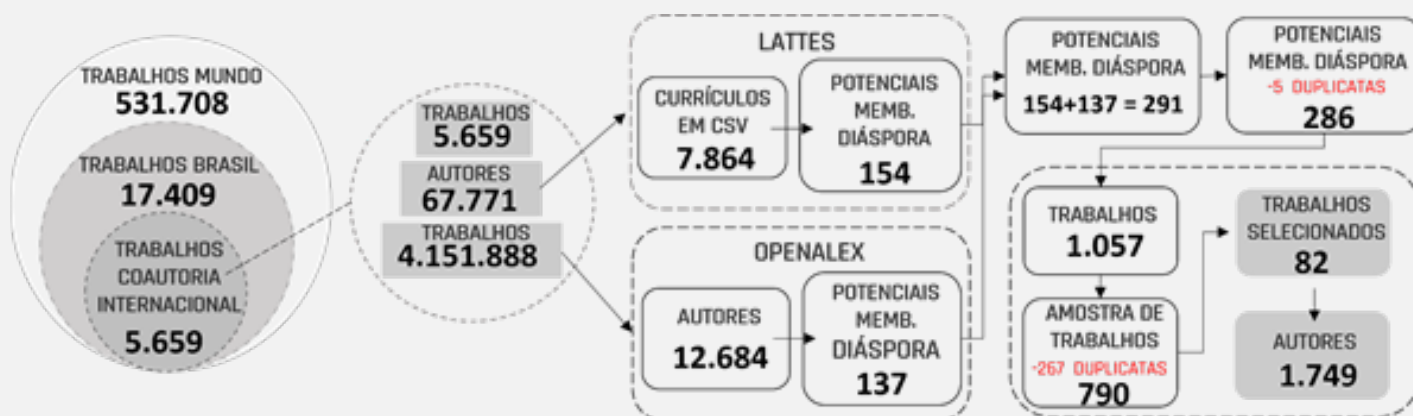
A sexta e última etapa foi a obtenção da **listagem dos membros da diáspora científica brasileira**, bem como a classificação dos demais autores dos artigos nas seguintes categorias: 1) Pesquisador estrangeiro que encontra-se no Brasil; 3) Pesquisador estrangeiro no exterior que colabora com pesquisador no Brasil; 4) Pesquisador brasileiro que encontra-se no Brasil; 5) Pesquisador estrangeiro no exterior que colabora com a diáspora científica brasileira (sem participação de pesquisador no Brasil); 6) Consórcio (Grupo de Investigação) e 7) Não se aplica.

3 RESULTADOS E DISCUSSÃO

O estudo bibliométrico resultou na identificação de 531.708 trabalhos científicos relacionados à COVID-19 no mundo. No cenário brasileiro, foram identificados 17.409 trabalhos, sendo que 5.659 (32,5%) contaram com colaborações internacionais. A partir desses trabalhos com coautoria internacional, extraiu-se o total de 67.771 autores. Foi possível identificar 4.151.888 de trabalhos no OpenAlex da vida toda destes autores (Figura 2).

A coleta de dados na Plataforma Lattes buscou identificar os autores a partir dos títulos dos artigos, sendo que 7.864 autores foram associados a algum currículo. Após a extração de informações sobre formação, atuação profissional, publicações e outros detalhes relevantes, foram selecionados 154 autores como potenciais membros da diáspora, a partir da exclusão dos currículos com endereços profissionais em branco ou no Brasil.

Figura 2 - Resultados da busca dos membros da diáspora.



Fonte: Elaborado pelos autores (2024).

Na *OpenAlex*, a partir dos trabalhos da vida toda (4.151.888), aplicou-se um filtro de seleção, buscando nos trabalhos pelo menos uma filiação à instituição no Brasil, resultando na identificação de 12.684 autores. Utilizando consultas condicionadas, foram identificados 137 autores que saíram do Brasil (ou seja, que tiveram pelo menos uma publicação com filiação à instituição no exterior depois de uma publicação com filiação no Brasil), que foram considerados potenciais membros da diáspora.

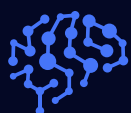
As amostras resultantes da *OpenAlex* (137 autores) e do *Lattes* (154 autores) foram combinadas, excluindo-se 5 duplicatas entre as listas de autores, resultando em 286 autores, aos quais correspondiam 1.057 artigos sobre COVID-19 com cooperação internacional. Removidas as duplicatas, a amostragem final foi de 790 artigos. Destes, 82 foram selecionados para checagem manual. Estes artigos abrangiam 1.749 autores no total, dos quais 297 participaram em mais de um artigo da amostra. Os autores da amostra incluíam pesquisadores no Brasil e no exterior. A análise da amostra levou à identificação de 60 membros da diáspora (Tabela 1).

Na etapa de verificação manual, foram identificadas algumas inconsistências provenientes do *OpenAlex*, o que resultou na recuperação de dados incorretos. Estas inconsistências abrangem a falta de *ORCID*, *ORCID* incorretos, discrepâncias nos nomes dos autores e a omissão de alguns autores listados em artigos específicos, o que demandou uma verificação e busca manual adicional.

Tabela 1 - Classificação dos autores em relação à coautoria dos trabalhos.

Classificação dos autores	Frequência
Pesquisador estrangeiro no exterior que colabora com pesquisador no Brasil	876
Pesquisador brasileiro que encontra-se no Brasil	473
Não se aplica*	179
Pesquisador estrangeiro no exterior que colabora com a diáspora científica brasileira (sem participação de pesquisador no Brasil)	150
Membro da diáspora científica brasileira	60
Outros	6
Pesquisador estrangeiro que encontra-se no Brasil	4
Consórcio**	1
Total	1749

Fonte: Elaborado pelos autores (2024).



Nota: *Artigos que não continham autores filiados às instituições no Brasil e que entraram na análise por erro da base *OpenAlex*. **Consórcio de pesquisa que assina com um nome coletivo.

4 CONCLUSÃO

A metodologia conseguiu identificar membros da diáspora científica brasileira a partir da localização de autores filiados a instituições no Brasil em algum momento, e também da identificação de suas nacionalidades. A triangulação das informações das várias bases de dados permitiu aproximações sucessivas dos artigos que poderiam ter membros da diáspora. Cerca de 50% dos artigos da amostra de checagem manual revelou a presença de membros da diáspora.

A base *OpenAlex* e a *Plataforma Lattes*, que possuem abordagens distintas na coleta e organização dos dados, oferecem vantagens e desvantagens neste trabalho. Enquanto a *API* da *OpenAlex* fornece uma visão mais abrangente, com dados bibliométricos e informações sobre a mobilidade dos autores, a *Plataforma Lattes* oferece detalhes mais específicos sobre a formação, trajetória profissional e produção científica dos pesquisadores. Essa complementaridade entre as duas fontes é crucial para uma compreensão completa do panorama da pesquisa dos membros da diáspora nos temas como colaboração internacional, mobilidade acadêmica e impacto da produção científica.

Apesar da abrangência da *OpenAlex*, é relevante notar as informações inacuradas de alguns campos, que representam um desafio considerável para assegurar a precisão dos dados relacionados à diáspora.

O artigo tratou e apresentou a metodologia desenvolvida para localizar membros da diáspora aplicada ao caso brasileiro. Com algumas adaptações, como a utilização da base *ORCID*, pode ser aplicada para o caso de outras diásporas científicas.

AGRADECIMENTOS

Este trabalho conta com financiamento da Fapesp (Contribuição da Diáspora Científica Brasileira na cooperação internacional do país - Processo: 20/04208-9). Os autores agradecem a equipe do projeto maior que apoiaram a realização da pesquisa.

REFERÊNCIAS

CAÑIBANO, C.; WOOLLEY, R. Towards a socio-economics of the brain drain and distributed human capital. **International Migration**, Geneva, v. 53, n. 1, p. 115-130, 2015. DOI: <https://doi.org/10.1111/imig.12020>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/imig.12020>.

CARNEIRO, A. M. *et al.* As Universidades Estaduais Paulistas contra a Covid-19. **ComCiência: Revista Eletrônica de Jornalismo Científico**, Campinas, 14 de jul. 2020. [Site]. Disponível em: <https://www.comciencia.br/as-universidades-estaduais-paulistas-contr-a-covid-19/>. Acesso em: 15 maio 2021.



CIUMASU, Ioan M. Turning brain drain into brain networking. **Science and Public Policy**, London, v. 37, n. 2, p. 135-146, Mar. 2010. DOI: <https://doi.org/10.3152/030234210X489572>. Disponível em: <https://academic.oup.com/spp/article-abstract/37/2/135/1619256>.

CINTRA, P. R. **Colaboração científica brasileira em emergências de saúde pública de importância internacional**: um estudo a partir dos casos do Zika vírus e da Covid-19. 2022. Tese (Doutorado em Política Científica e Tecnológica) – Instituto de Geociências, Universidade Estadual de Campinas, Campinas, 2022. Disponível em: <https://repositorio.unicamp.br/acervo/detalhe/1239330>.

LAUDEL, Grit. Studying the brain drain: can bibliometric methods help? **Scientometrics**, Dordrecht, v. 57, n. 2, p. 215-237, 2003. DOI: <https://doi.org/10.1023/A:1024137718393>. Disponível em: <https://link.springer.com/article/10.1023/A:1024137718393>.

MAHER, Brendan; VAN NOORDEN, Richard. How the COVID pandemic is changing global science collaborations. **Nature**, London, v. 594, n. 7863, p. 316-319, 2021. DOI: <https://doi.org/10.1038/d41586-021-01570-2>. Disponível em: <https://www.nature.com/articles/d41586-021-01570-2>.

MEYER, Jean-baptiste; BROWN, Mercy. **Scientific diasporas**: a new approach to the brain drain. Management of Social Transformations MOST, Discussion Paper Series - No. 41. Prepared for the World Conference on Science UNESCO - ICSU, Budapest, Hunw-y, 26 June-I July 1999. Paris: UNESCO, 1999. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000116330>.

MOED, Henk F.; HALEVI, Gali. A bibliometric approach to tracking international scientific migration. **Scientometrics**, Dordrecht, v. 101, n. 3, p. 1987-2001, 2014. DOI: <https://doi.org/10.1007/s11192-014-1307-6>. Disponível em: <https://link.springer.com/article/10.1007/s11192-014-1307-6>.

MOTA, D. M.; FERREIRA, P. J. G.; LEAL, L. F. Produção científica sobre a COVID-19 no Brasil: uma revisão de escopo. **Vigilância Sanitária em Debate: Sociedade, Ciência & Tecnologia**, Rio de Janeiro, v. 8, n. 3, p. 114-124, ago. 2020. DOI: <http://dx.doi.org/10.22239/2317-269x.01599>. Disponível em: <https://visaemdebate.incqs.fiocruz.br/index.php/visaemdebate/article/view/1599>.

PRIEM, Jason; PIWOWAR, Heather; ORR, Richard. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *In*: INTERNATIONAL CONFERENCE ON SCIENCE AND TECHNOLOGY INDICATORS, 26th, 2022, Granada. **Proceedings [...]**. Granada: University of Granada, 2022. Disponível em: <https://arxiv.org/abs/2205.01833>. Acesso em: 16 ago. 2023.

SANTOS NETO, João Arlindo dos. Produção científica sobre a covid-19 na ciência da informação no Brasil: uma pesquisa na Brapci. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde (RECIIS)**, Rio de Janeiro, v. 16, n. 1, p. 11-17, jan./mar. 2022. DOI: <https://doi.org/10.29397/reciis.v16i1.2389>. Disponível em: <https://www.recis.icict.fiocruz.br/index.php/reciis/article/view/2389>.

SILVA, Vania Ferreira da *et al.* Produção científica brasileira sobre a covid-19 na Scopus (2019-2021): uma análise bibliométrica na área das ciências da saúde. **Reciis - Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, Rio de Janeiro, v. 16, n. 2, p. 388-403, abr./jun. 2022. DOI: <https://doi.org/10.29397/reciis.v16i2.2373>. Disponível em: <https://www.recis.icict.fiocruz.br/index.php/reciis/article/view/2373>.