



CORRELAÇÕES ENTRE INDICADORES DE CITAÇÃO E ALTMÉTRICOS:

análise de artigos brasileiros com colaboração internacional

Danielle Pompeu Noronha Pontes¹
João de Melo Maricato¹
Marcia Regina da Silva²

Resumo: Investiga a correlação de indicadores de citação e altmétricos em relação à países e idiomas. Analisou-se dados de 22.170 artigos extraídos da base Scopus (2017-2021). Verificou-se que as maiores correlações estão entre os indicadores de citação (Crossref/Scopus). A mídia com maior correlação com os indicadores de citação foi o Mendeley. Em geral, existe correlação semelhante entre os indicadores dos países, sendo uma das exceções os dos Estados Unidos. Em relação a língua, as correlações mais altas são dos indicadores dos artigos em inglês, sendo uma das exceções os artigos em português, que apresentaram correlações maiores entre Crossref/Twitter e Scopus/Mendeley.

Palavras-Chave: Almetria, Correlação, Bibliometria, Citação

1 INTRODUÇÃO

A crescente utilização acadêmica de ferramentas de mídias sociais tem despertado nas comunidades científicas interesse em rastrear o impacto acadêmico por meio de novos indicadores (ARAÚJO; CARAN; SOUZA, 2016). Avaliar o resultado da pesquisa e medir o seu impacto é um componente essencial para evolução da ciência. A contagem tradicional de citações é uma prática reconhecida para a avaliação do impacto acadêmico, no entanto, beneficia mais os países da América do Norte e Europa em detrimento daqueles em desenvolvimento. As bases de dados tradicionais utilizadas para a avaliação científica não representam de forma efetiva países em desenvolvimento. Em contrapartida, as métricas alternativas podem auxiliar na medição mais ampla do impacto dos trabalhos acadêmicos podendo beneficiar os países não centrais (ALPERIN, 2013).

A velocidade de comunicação provocada pela conectividade através da web social vem remodelando o panorama da comunicação científica, permitindo o surgimento de processos mais ágeis de comunicação científica, impulsionando o uso de métricas alternativas oriundas

¹ Universidade de Brasília (UnB)

² Universidade de São Paulo (USP)

de mídias e redes sociais. Nesses ambientes o acesso à informação é democrático e desburocratizado em relação às bases tradicionais. Compreender o potencial das novas métricas é importante para universidades, agências de financiamento e governos dos países em desenvolvimento. Existe a possibilidade de que tais métricas proporcionem um olhar regionalizado da pesquisa, orientando decisões dos autores (sobre onde publicar) e dos sistemas de avaliação, em prol do desenvolvimento regional.

Contemporaneamente, os indicadores altmétricos estão amplamente disponíveis, sendo necessário compreender se eles podem ajudar a avaliar os resultados acadêmicos (THELWALL, 2020). A altmetria tem sido usada para medir o impacto mais amplo das atividades de pesquisa (ALHOORI *et al.*, 2014). No entanto, ainda são poucos os estudos que analisam os potenciais dos indicadores altmétricos em relação aos países e idiomas. Desse modo, os principais resultados das pesquisas publicados, até o momento, trazem dados “internacionais” (grandes bases de dados e idioma inglês), não contemplando características particulares de países, regiões e idiomas.

A pesquisa de Sedighi (2021) é um exemplo de trabalho com viés de um país. O autor avaliou o impacto dos produtos científicos iranianos selecionados nas áreas de ciências sociais e ciência da computação indexadas na Scopus usando indicadores altmétricos (artigos publicados entre 2014-2018). A relação entre variáveis como colaboração entre pesquisadores, periódicos de acesso aberto foram analisadas por meio de testes de correlação.

Mas-Bleda e Thelwall (2016) comparam a pesquisa espanhola e britânica publicada em 2012, utilizando uma série de indicadores bibliométricos e altmétricos, para estudar os pontos fortes das pesquisas espanholas. O trabalho inclui a análise de patentes, menções de apresentações on-line, de currículo de cursos, da Wikipedia, menções no Twitter e contagens de leitores/capturas no Mendeley. Os resultados mostraram que o Mendeley é a fonte altmétrica com maior cobertura, com 80% dos artigos da amostra, seguido pelo Twitter (34%).

O estudo apresentado por Alperin (2015) explorou os níveis de cobertura de indicadores altmétricos entre países, assuntos e idiomas de artigos publicados na América Latina. Ao fazer isso, concentra atenção especial ao maior país da região, o Brasil. Explorando uma região em desenvolvimento de forma ampla e um país emergente em detalhes, a pesquisa abre uma janela a discussão sobre a universalidade e os limites de algumas das reivindicações existentes da literatura altmétrica, fornecendo uma compreensão mais clara sobre o potencial de altmetria em diferentes países.

Diante do exposto, o presente estudo propõe analisar a relação entre os indicadores altmétricos e os de citação de autores vinculados à instituições brasileiras, publicados em colaboração internacional (em inglês, português ou espanhol). Para isso, foram analisados artigos classificados simultaneamente nas áreas de ciências sociais e ciência da computação. A escolha desse corpus, embora não seja objetivo central do trabalho, pode trazer a possibilidade de visualização da produção e relações entre as áreas, intimamente relacionadas com a Ciência da Informação, podendo contribuir com maior entendimento do campo. Os dados foram extraídos da base de dados Scopus, no período de 2017 a 2021. A escolha da base deu-se por conveniência, visto que ela possui sistema de recuperação de informações robusto, ser relativamente fácil extraír dados da base e por ela ser uma das maiores fontes de dados bibliográficas do mundo. Para se analisar as correlações entre os dados, optou-se por adotar a correlação de Spearman. A escolha em se utilizar a correlação de Speraman ao invés de Perarson, é recomendável nos casos em que haja alguns pontos muito afastados dos demais, ou quando exista uma relação monótona crescente ou decrescente entre o conjunto de dados (SOUSA, 2019). De maneira geral, os indicadores bibliométricos e altmétricos possuem tal característica.

A presente pesquisa busca responder algumas questões: Existem diferenças ou semelhanças nas correlações entre citação e indicadores altmétricos em razão dos países que publicaram em colaboração internacional com o Brasil? O idioma dos artigos publicados por pesquisadores brasileiros em colaboração internacional influenciam nos níveis de correlação entre os indicadores altmétricos e de citação? Quais as correlações entre as citações da base de dados Scopus e da Crossref e quais delas apresentam maior correlação com os indicadores altmétricos? Quais indicadores altmétricos (Mendeley ou Twitter) apresentam maiores índices de correlação (possibilidade de explicar citação posterior) no contexto dos artigos de pesquisadores atuantes no Brasil com artigos em colaboração internacional?

2 PROCEDIMENTOS METODOLÓGICOS

Este estudo descritivo utiliza indicadores altmétricos e de citação para analisar correlações entre eles. Foram levantados artigos de pesquisadores vinculados ao Brasil que mais contaram com colaboração internacional indexados na base de dados Scopus no período entre 2017 e 2021. Os dados bibliométricos incluíram artigos que continham, dentre outros metadados, o

DOI, a quantidade de citações computadas na Scopus, o idioma original do artigo e o ano de publicação.

Foram coletados artigos, oriundos de revistas, em estágio final de publicação, com temática de intersecção entre as áreas de Ciência Social (SOCI) e Ciência da Computação (COMP). A busca foi restrita aos autores com afiliação em instituições do Brasil em coautoria com autores de 11 países (Alemanha, Austrália, Canadá, China, Espanha, Estados Unidos, Hong Kong, Índia, Itália, Reino Unido e Portugal) com maiores colaboração.

Considerou-se apenas artigos nos idiomas Inglês, Espanhol e Português. Limitou-se a artigos que continham o DOI, resultando em um total de 21.877 registros (esse número inclui dupla contagem, pois um mesmo artigo pode ter sido objeto de colaboração entre dois ou mais países, sendo o número de DOIs únicos 15.702 artigos). Posteriormente os DOIs dos artigos foram pesquisados através de APIs que possibilitaram a coleta de indicadores altmétricos de menções (Twitter), Citações (Crossref) e Leitores/capturas (Mendeley). Os dados foram analisados nos programas Microsoft Excel.

Os indicadores destas fontes foram correlacionados entre eles e com os indicadores de citação da Scopus (extraídos da própria base). Um coeficiente de correlação mede o grau pelo qual duas variáveis estão associadas entre si. Quanto maior for o valor absoluto do coeficiente, mais forte é a relação entre as variáveis. Também foram realizadas algumas análises com o número absoluto das citações, para fins de comparações entre os anos de publicação. Neste caso, foram excluídas todas as duplicidades tomando como base o DOI do documento, resultando em 15.702 artigos.

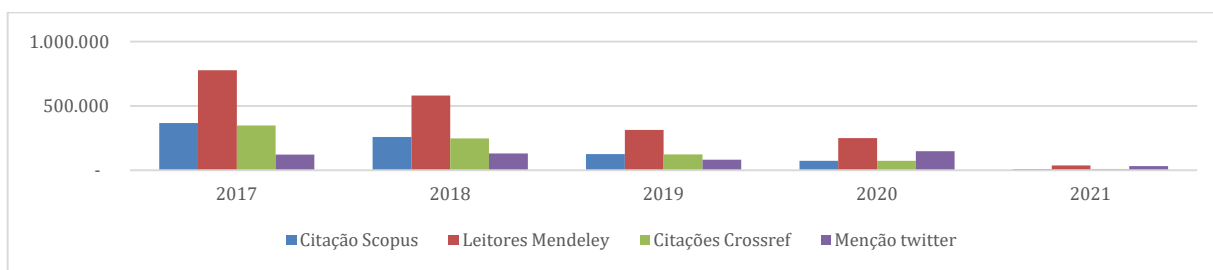
3 RESULTADOS E DISCUSSÕES

As maiores ocorrências de indicadores são apresentados pelo Mendeley (1.959.507), seguido de citações da Scopus (833.301) e, praticamente empatado, aparecem as citações da Crossref (803.269). O Twitter, no conjunto de dados, tem pouca expressividade com 515.538 menções. No entanto, no ano de 2020 observa-se que ele tem proeminência quando comparado aos indicadores de citação (Crossref e Scopus). Isso sugere que o Twitter proporciona a medição e o impacto de maneira mais imediata, quando comparados aos demais indicadores. Do total de menções, leituras e citações, o Mendeley conta com 47%, o Scopus com 20%, o Crossref com 19% e o Twitter com 12% (em relação a totalidade dos artigos únicos, 15.702).

Desse total de artigos, 40% (6.315) foram publicados no ano de 2017 seguido por artigos de 2018 que concentram 31% das publicações totalizando 4.825 artigos. Os artigos publicados em 2019 representam 18% com 2.853, 2020 representa 9% com 1.509 e os publicados em 2021 representam 1% com 200 artigos. Ou seja, o número de artigos publicados na base Scopus com intersecção entre as áreas de ciências sociais e ciência da computação apresenta decréscimo de aproximadamente 10% de publicações ao ano.

No Gráfico 1 fica claro que as menções vão diminuindo, em média 10%, em relação ao ano anterior. Ou seja, quanto mais antigo o artigo (dentro do período observado), mais citações, menções e leituras/capturas ele teve. É possível ainda notar que há bastante proximidade entre o número de citações da Crossref e Scopus em todos os anos analisados (Gráfico 1).

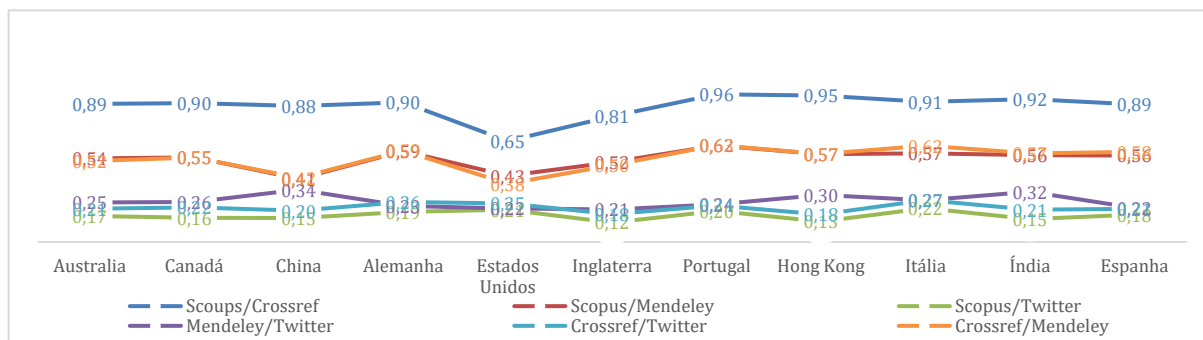
Gráfico 1 - Menções, citações e leituras por ano de artigos publicados por pesquisadores de instituições brasileiras em colaboração internacional



Fonte: Elaborado pelos autores.

Analisando o Gráfico 2, observa-se que a mídia social com a maior correlação com os indicadores de citações (Crossref e Scopus) é o Mendeley. O Twitter tem baixa correlação com os demais canais.

Gráfico 2 - Correlação Spearman entre indicadores de citação e altmetrics entre os países que colaboraram com pesquisadores de instituições brasileiras



Fonte: Elaborado pelos autores.

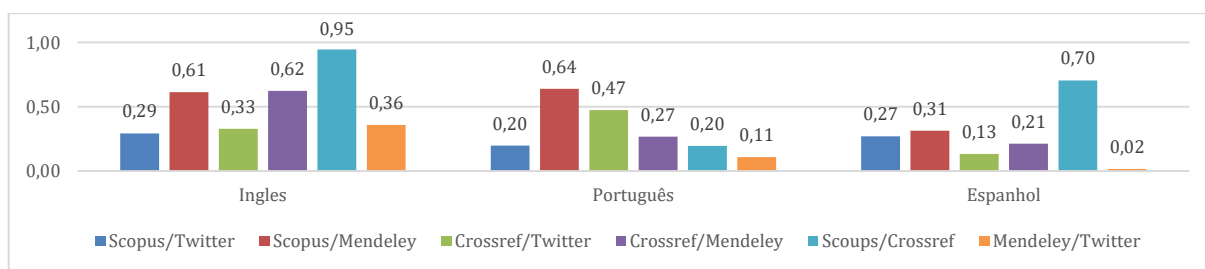
Constata-se que há um padrão entre as correlações dos indicadores calculados para cada país onde, para todos eles, a mais alta correlação ocorre entre Scopus/Crossref (ambas com dados

de citação). Em segundo e terceiro lugar estão as correlações entre Crossref/Mendeley e Scopus/Mendeley. Em quarta posição encontra-se a correlação entre Mendeley/Twitter (0,34). As correlações Crossref/Twitter e Scopus/Twitter encontram-se bem próximas na quinta e sexta posições respectivamente. Destes dados, o que mais chama a atenção é a alta correlação entre as bases Scopus e Crossref, indicando que as fontes são em grande medida compatíveis e que, no caso concreto, ambas tem potencial avaliativo semelhante.

Apesar da existência de um padrão semelhante de correlações entre os indicadores dos artigos publicados por pesquisadores atuantes no Brasil com os diversos países em que houve colaboração, alguns casos as correlações são destoantes da média geral. Um exemplo disso são as publicações realizadas em parceria com pesquisadores dos Estados Unidos, como as correlações entre Scopus/Crossref (0,65), Scopus/Mendeley (0,43) e Crossref/Mendeley (0,38) abaixo da média geral. Essa diferença pode estar relacionada, por exemplo, à deficiência de contagem de citações por uma das duas bases. E, no caso dos indicadores altmétricos, é possível que, diferentemente da maioria dos países, as menções não podem explicar citações posteriores. Portanto, artigos publicados em colaboração com Portugal e Itália, por exemplo (cujas correlações são de aproximadamente 0,62) tendem a ter maior possibilidade de que os leitores/capturas do Mendeley expliquem as citações do que aqueles publicados em parceria com os Estados Unidos.

O Gráfico 3 apresenta a distribuição dos resultados da correlação entre os indicadores de citações e os indicadores altmétricos levantados em consideração o idioma de origem do artigo. Os indicadores dos artigos escritos em inglês possuem as correlações mais fortes em comparação com o espanhol e o português na maioria das vezes. Dentre os dados, um ponto que chama a atenção é o fato de que as correlações entre Scopus/Crossref dos artigos em português são as mais baixas entre os três idiomas (0,20), indicando que há debilidades e que uma das duas bases pode ter maior potencial de cobrir artigos em português. Outro fator que reforça essa impressão é o fato de que a correlação entre Crossref/Twitter ser, também, maior nos artigos em português do que nos demais.

Gráfico 3 - Correlação Spearman entre indicadores de citação e altmétricos (idioma de publicação) de artigos em colaboração internacional com pesquisadores de instituições brasileiras



Fonte: Elaborado pelos autores.

De maneira geral, considerando os três idiomas, a mídia social que apresenta a mais alta correlação com os indicadores tradicionais ocorreu entre Scopus/Crossref (0,95). Porém, a maioria dos artigos foi publicada em língua inglesa (97%), seguidos dos publicados em espanhol (aproximadamente 0,35%) e em português (0,15%) e 2% dos artigos não apresentavam a indicação do idioma. Observar-se que entre as citações/menções/leituras os artigos em inglês possuem 97% do total de 4.111.851, um número expressivamente maior que os demais idiomas. Sendo de longe seguido pelo espanhol.

4 CONCLUSÕES

Em geral, nota-se um padrão nas correlações entre citações e indicadores altmétricos em razão dos países que publicaram em colaboração internacional com o Brasil. Destaca-se a alta correlação entre as bases Scopus e Crossref, o que indica que as fontes são em grande medida compatíveis e que, no caso concreto, ambas tem potencial avaliativo semelhante. Um ponto a observar é que o indicador de correlação Scopus/Crossref dos Estados Unidos apresenta uma queda acentuada em relação aos demais países. Entre as correlações estudadas em relação a língua, o idioma Inglês apresenta os índices de correlações mais alto, exceto nas correlações Crossref/Twitter (0,47) e Scopus/Mendeley (0,64) que tem como correlação mais alta o idioma Português. Isso aponta que artigos na língua portuguesa, publicados por pesquisadores brasileiros em colaboração internacional, possuem uma melhor proporção de menção/citação/leitura quando comparados canais alternativos x canais tradicionais em detrimento aos demais idiomas, reforçando a ideia que uma das duas bases tradicionais pode ter maior potencial de cobrir artigos em português. Observou-se que na média entre todos os países estudados o maior índice de correlação é Scopus x Crossref. Analisando a média dos índices de correlações entre as citações da base de dados Scopus e da Crossref com os

indicadores altmétricos, as correlações Crossref x Mendeley e Scopus x Mendeley possuem valores muito próximos (0,57 e 0,56 respectivamente). Em relação aos dados altmétricos, percebe-se que o Mendeley apresenta as maiores correlações, quando comparado aos demais indicadores altmétricos. A significativa vantagem do Mendeley em relação a número absoluto de menções pode ter como motivo o fato de que contagem de citações tradicionais leva um tempo maior para ser acumulada e reflete apenas os impactos científicos dentro da academia (MAS-BLEDA; THELWALL, 2016), enquanto o Mendeley registra de forma mais ágil.

REFERÊNCIAS

- ALHOORI, H. *et al.* Altmetrics for country-level research assessment. *In: THE EMERGENCE OF DIGITAL LIBRARIES – RESEARCH AND PRACTICES. ICADL 2014. LECTURE NOTES IN COMPUTER SCIENCE.* 2014, Suíça. **Anais [...].** Suíça: Springer, 2014. p. 59–64. DOI: https://doi.org/10.1007/978-3-319-12823-8_7. Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-12823-8_7. Acesso em: 5 jan.2022.
- ALPERIN, J. P. Ask not what altmetrics can do for you, but what altmetrics can do for developing countries. **Bulletin of the American Society for Information Science and Technology**, [S. l.], v. 39, n. 4, p. 18–21, 2013. DOI: <https://doi.org/10.1002/bult.2013.1720390407>. Disponível em: <https://www.semanticscholar.org/paper/Ask-not-what-altmetrics-can-do-for-you%2C-but-what-do-Alperin/efe7de8182066599c8792c01d6e19e963a464856>. Acesso em: 27 jan. 2022.
- ALPERIN, J. P. Geographic variation in social media metrics: An analysis of Latin American journal articles. **Aslib Journal of Information Management**, Bingley, v. 67, n. 3, p. 289–304, 2015. DOI: <https://doi.org/10.1108/AJIM-12-2014-0176>. Disponível em: <https://www.semanticscholar.org/paper/Geographic-variation-in-social-media-metrics%3A-an-of-Alperin/a945459db5ce2d9162622ed1060f89c1e04db8c7>. Acesso em: 10 jan. 2022.
- ARAÚJO, R. F.; CARAN, G. M.; SOUZA, I. V. P. Orientação temática e coeficiente de correlação para análise comparativa entre altmetrics e citações. **Em Questão**, Porto Alegre, v. 22, n. 3, p. 184, 2016. DOI: <https://doi.org/10.19132/1808-5245223.184-200>. Disponível em: <https://brapci.inf.br/index.php/res/v/11888>. Acesso em: 5 jan. 2022.
- MAS-BLEDA, A.; THELWALL, M. Can alternative indicators overcome language biases in citation counts? A comparison of Spanish and UK research. **Scientometrics**, Budapest, v. 109, n. 3, p. 2007–2030, 2016. DOI: <https://doi.org/10.1007/s11192-016-2118-8>. Disponível em: <https://link.springer.com/article/10.1007/s11192-016-2118-8>. Acesso em: 10 jan. 2022.
- SEDIGHI, M. Assessing research impact based on altmetrics case study: Iranian scientific products in social sciences and computer science indexed in scopus. **Iranian Journal of Information Processing and Management**, Tehran, v. 36, n. 3, p. 679–708, 2021. DOI: <https://doi.org/10.52547/jipm.36.3.679>. Disponível em: https://jipm.irandoc.ac.ir/browse.php?a_id=4390&slc_lang=en&sid=1&printcase=1&hbnr=1&hmb=1. Acesso em: 5 jan. 2022.
- SOUSA, Á. Coeficiente de Correlação de Pearson e Coeficiente de correlação de Spearman . O que medem e em que situações devem ser utilizados ? **Correio dos Açores**, [S. l.], v. 1, n. 1, p. 19, 2019. Disponível em: <http://hdl.handle.net/10400.3/5365>. Acesso em: 5 jan. 2022.

THELWALL, M. Mendeley reader counts for us computer science conference papers and journal articles. **Quantitative Science Studies**, Cambridge, v. 1, n. 1, p. 347–359, 2020. Disponível em: http://dx.doi.org/10.1162/qss_a_00010. Acesso em: 10 jan. 2022.