

FERRAMENTAS DE ANÁLISE DE DADOS ALTMÉTRICOS: UM RELATO DE EXPERIÊNCIA

Vildeane Borba

Doutora em Comunicação e Informação
Universidade Federal de Pernambuco (UFPE)
E-mail: vildeane.borba@ufpe.br

Sônia Caregnato

Doutorada em Information Studies
Universidade Federal do Rio Grande do Sul (UFRGS)
E-mail: sonia.caregnato@ufrgs.br

INTRODUÇÃO

O desenvolvimento das tecnologias digitais, e em especial a internet, tem modificado a forma como se cria, se gerencia e se interage com a informação, especialmente quando a web influencia diferentes práticas comunicacionais, tais como quando se conecta, modifica, compartilha, discute, visualiza e/ou “curte” (like) em plataformas de mídias sociais, também chamadas de sites de rede social *on-line*, como Facebook e Twitter.

Dentre as diversas informações discutidas e disseminadas em plataformas de mídias sociais, a informação científica chama a atenção quando publicações acadêmicas, sejam elas artigos de revisão, originais, capítulos de livros, anais de eventos, teses e/ou dissertações, são divulgadas neste novo locus. Meschede e Siebenlist (2018) afirmam que o campo científico está operando no mundo conectado, na medida em que cientistas estão usando, apresentando e colaborando com outros, por meio de portais e ferramentas de mídias sociais.

A utilização de mídias sociais, cujas atividades se concentram em difusão e visibilidade de conteúdos científicos, fez surgir uma nova subárea de estudo no âmbito dos estudos métricos da informação, chamada altmetria. Ela se utiliza das fontes de dados emergentes, isto é, das plataformas de mídias sociais, para contribuir na avaliação da pesquisa, complementando os indicadores tradicionais.

Nesse sentido, por ser uma área em desenvolvimento, a coleta e a análise de dados relativos à produção científica advindos de plataformas de mídias sociais têm sido um desafio. Um dos aspectos que necessita ser estudado em maior profundidade é o papel

dos agregadores de dados, que são ferramentas que reúnem dados altmétricos e deles derivam indicadores.

Assim, este artigo tem o objetivo de relatar a experiência de utilização dos agregadores *Altmetric.com* e do *Webometric Analyst*, em um estudo sobre a presença dos artigos de revistas científicas brasileiras da Ciência da Informação nas plataformas de mídias sociais. Ressalta-se que o objeto deste artigo são os agregadores e não os dados altmétricos das revistas.

ALTMETRIA E AGREGADORES DE DADOS ALTMÉTRICOS

A origem do termo altmetria relaciona-se à utilização da hashtag #altmetrics em um tweet por Jason Priem, em 2010. Posteriormente, uma publicação dele com colaboradores, conhecida como *Manifesto da Altmetria*, estabelece o potencial da altmetria para melhorar a comunicação científica e a avaliação da pesquisa, diferenciando-a dos indicadores tradicionais e definindo-a como a descoberta de vestígios invisíveis de impacto acadêmico, através da observação de atividades em ferramentas e sistemas *on-line*. (ERDT et al., 2016; HAUSTEIN; BOWMAN; COSTAS, 2016; PRIEM et al., 2010; SANKAR; KAVITHA, 2016)

A altmetria visa complementar os indicadores tradicionais a partir de dados abertos de fontes advindas de plataformas de mídias sociais. Tais mídias foram desenvolvidas a partir da web 2.0, que está inserida nos preceitos na ciência aberta. Segundo Bartling e Friesike (2014, p. 10), a ciência aberta “refere-se a uma cultura científica que se caracteriza por sua abertura, ao compartilhamento de resultados para um público amplo”.

A coleta e análise de dados altmétricos são realizados através de agregadores de dados, que são definidos como “ferramentas e plataformas que agregam e oferecem eventos online como métricas derivadas de provedores de dados altmétricos”. (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2016, p. 70) Os agregadores de dados mais populares de plataformas de mídias sociais e que coletam dados de diversas fontes são: Impactstory, Plum Analytics, PLOS ALM, Altmetric.com e Webometric Analyst, estes dois últimos serão mais aprofundados, pois foram o foco deste artigo.

As estruturas de coleta de dados dos agregadores altmétricos se apresentam de diferentes formas. Os agregadores primários são aqueles que rastreiam os dados altmétricos diretamente dos provedores de dados, isto é, das plataformas de mídias sociais como o Facebook, Twitter e Mendeley entre outros, tendo como exemplo o Altmetric.com e o PLOS ALM. Os agregadores secundários são aqueles que coletam os dados altmétricos de agregadores primários, como o Impactstory, Plum Analytics e Webometric Analyst e os agregadores terciários são os que recolhem tanto de agregadores primários quanto de secundários, como é o caso do Snowball Metrics. (ERDT et al., 2016)

O Altmetric.com foi desenvolvido em 2011, por Euan Adie, e lançado em fevereiro de 2012, sendo, a partir deste mesmo ano, financiado pela Digital Science. Pesquisadores afirmam que o Altmetric.com é um dos frameworks emergentes mais populares que rastreia, analisa e coleta atividades *on-line* contínuas em torno de resultados de pesquisa. (ADIE; ROE, 2013; ERDT et al., 2016; MUKHERJEE; SUBOTIC; CHAUBEY, 2018)

Esse agregador tem interfaces específicas para atender a diferentes públicos, entre eles editores, instituições, pesquisadores, financiadores, e para pesquisa em desenvolvimento, contribuindo no monitoramento, navegação, pesquisa e avaliação em torno da pesquisa publicada em fontes não convencionais. (DIGITAL SCIENCE, 2016) As ferramentas disponíveis nele abrangem o Altmetric explorer for publishers Institutions and Funders; Altmetric for books; Altmetric Badges; Altmetric API e Free tools.

Já o Webometric Analyst é um software gratuito, projetado para realizar análises automáticas variadas na web para fins de pesquisa em Ciências Sociais. Ele pode criar diagramas de rede de coleções de sites, estimar o impacto *on-line* de coleções e recuperar informações em larga escala sobre blogs e vídeos do YouTube. (THELWALL, 2009)

O programa reúne dados da web de diferentes fontes e os processa de várias maneiras, tendo como principais análises: um diagrama de rede que pode ser produzido para ilustrar a conectividade em um conjunto de sites e a partir de uma tabela de resumo. (THELWALL, 2009) Para as métricas alternativas, o Webometric processa dados do Mendeley, Altmetric.com, Google books, Worldcat, Youtube, Twitter entre outros, através de interface distinta para cada tipo de fonte de coleta a ser realizada.

METODOLOGIA

Na análise da utilização dos agregadores de dados Altmetric.com e Webometric Analyst, o *corpus* utilizado para testar os dois agregadores constituiu-se de 13 periódicos brasileiros na área da Ciência da Informação, englobando 2.724 artigos científicos publicados no período de 2011 a 2018 que disponibilizavam seus DOIs.

O Webometric Analyst foi baixado do site do grupo de pesquisa Statistical Cybermetrics, sendo executado no sistema operacional Windows. A entrada de dados no software foi realizada através de arquivo de notas (.txt), contendo as informações de título, autor e ano em que pelo menos um destes elementos não seja nulo, <Blank -tab- title - tab- -authors -tab-year>. Para este software, não é obrigatória a existência do DOI do artigo.

Para a estrutura de entrada do formato do arquivo no Webometric Analyst, isto é, título, autor e ano, deve-se excluir os caracteres: ?, [], (), ^, ou espaços duplos, no intuito de evitar erros. Também é necessária a retirada de todos os acentos, pois a codificação do software foi configurada para o inglês.

A API aberta do Altmetric.com possibilita ao usuário fazer a pesquisa de dados altmétricos a partir da pesquisa pelo DOI do artigo.¹ O DOI, em vermelho, pode ser pesquisado em qualquer navegador web, tendo como resultado os metadados disponibilizados pelo sistema.

No entanto, para realização de estudos altmétricos, que necessitam buscar um quantitativo de DOIs relativamente alto, a coleta individual no Altmetric.com é inviável. Então, foi necessário o desenvolvimento de um aplicativo web que permitisse a pesquisa automática de uma coleção maior de DOIs e que seus resultados fossem apresentados de uma forma melhor estruturada para posterior tabulação, comparação e análise. A aplicação Web foi desenvolvida em angular e NodeJS com o banco de dados Cloud Firestore e foi ligado diretamente na API gratuita do Altmetric.com, possibilitando a pesquisa pela entrada de um arquivo de planilha com os números de DOIs organizados por linha.

Figura 1 - Tela de resultados da aplicação web de dados altmétricos a partir da API aberta do Altmetric.com: teste com DOIs do periódico *Em Questão*

#	Nº DOI	Título	Periódico da Publicação
1	10.19132/1808-5245242.12-38	Aspectos éticos da coautoria em publicações científicas	Em Questão
2	10.19132/1808-5245241.267-295	Em nome da moral e dos bons costumes: censura a livros com temática de g...	Em Questão
3	10.19132/1808-5245240.258-277	Cenários da dinâmica de hiperescrevimento e colapso das revistas científicas b...	Em Questão
4	10.19132/1808-5245241.296-302	A altmetria na prática e o papel dos bibliotecários no seu uso e aplicação	Em Questão
5	10.19132/1808-5245233.226-240	Análise bibliométrica da produção científica brasileira e do nordeste em fiote...	Em Questão
6	10.19132/1808-5245233.179-206	Uma proposta de taxonomia para a categorização das falhas no ambiente da...	Em Questão
7	10.19132/1808-5245233.276-290	Para cada universo colecionista, suas fontes: dinâmicas informacionais nas tr...	Em Questão
8	10.19132/1808-5245233.167-180	O catálogo da biblioteca e o linked data	Em Questão
9	10.19132/1808-5245231.129-153	Representação semiótico-discursiva de cibercordões	Em Questão
10	10.19132/1808-5245230.213-234	Inspiência da visualização de indicadores bibliométricos e altmétricos nos Re...	Em Questão

Fonte: dados de pesquisa (2019).

A aplicação web possui uma arquitetura simples de informação, tendo como ferramenta principal para consulta a aba selecionar arquivo, que deve ser inserida a partir de um arquivo de planilha com cada DOI por linha, na primeira coluna da primeira planilha e a coluna não deve conter cabeçalho. A aplicação web registra os números absolutos e percentuais de artigos que obtiveram registros altmétricos e o quantitativo dos que não foram localizados, mostrando os dados descritivos (número de DOI, título e nome do periódico), finalizando através da exportação da captura em arquivo de planilha, conforme Figura 1.

¹ Ver: <https://api.altmetric.com/v1/doi/10.1590/2318-08892018000300005>.

As escolhas de utilização destas duas ferramentas deveram-se ao Mendeley ter sido a mídia com o maior número de registros altmétricos na maioria dos periódicos pesquisados oriundos do Altmetric.com, porém este agregador apresenta dados quantitativos de leitores, sem demonstração de aspectos complementares (demográficos e geográficos). Nesse sentido, o Webometric Analyst possibilita conhecer, além da quantidade de artigos com marcação no Mendeley, a quantidade de marcações por usuário, o perfil ocupacional dos leitores, as áreas do conhecimento, as nacionalidades dos leitores e os artigos científicos que obtiveram o maior número de leitores.

RESULTADOS

A visualização dos registros encontrados com dados altmétricos foi possível a partir do conjunto dos 73 metadados disponibilizados pelo Altmetric.com em arquivo de planilha. O Altmetric.com coleta dados do Mendeley, Twitter, Facebook, Blogs, YouTube/Vimeo, Google+, CiteULike, Reddit e Conotea, sendo estas três últimas não difundidas no Brasil e o Google + uma plataforma que foi finalizada.

O Webometric Analyst, por sua vez, permite a visualização de detalhes obtidos do Mendeley, como número de usuários, local, disciplina, ao passo que sua estrutura de busca permite uma abrangência maior de registros. Enfatiza-se que o Altmetric.com disponibiliza apenas o número de leitores por artigo no Mendeley, tendo o Webometric Analyst apresentado mais detalhes para além de número de leitores.

A partir da utilização das duas ferramentas acima relatadas em artigos de periódicos brasileiros em Ciência da Informação, pode-se inferir que distintas metodologias e ferramentas em altmetria podem gerar diferentes resultados, com discrepâncias e semelhanças pontuais. Alguns trabalhos brasileiros mostraram estes diferentes resultados em periódicos brasileiros em CI devido a utilização de diferentes ferramentas, como é o caso de Nascimento e Oddone (2014; 2015), Manhique e Rodrigues (2019) e Araújo (2015).

Neste estudo foi verificado que apenas 209 artigos científicos dentre os 2.724 artigos com DOI pesquisados apresentaram dados altmétricos a partir do Altmetric.com via API nas plataformas Mendeley, Twitter, Facebook, Blogs, YouTube/Vimeo, CiteULike e Google+, enquanto que no Webometric Analyst, utilizando o mesmo *corpus*, foram encontrados registros altmétricos para 2.022 artigos na plataforma Mendeley.

Nesse sentido, percebeu-se que os números de DOIs de alguns artigos de periódicos brasileiros em CI estavam incorretos, isto é, o link destes DOIs remetia a uma página de erro na CrossRef, presumindo-se que a baixa quantidade de resultados positivos de dados altmétricos no Altmetric.com via API, deve-se a isso. Tal resultado foi verificado principalmente nos periódicos com Qualis Capes B1, que obtiveram registros no Webometric Analyst oriundos do Mendeley, e não alcançaram registros no Altmetric.com, que também abrange esta plataforma de mídia social em seus resultados.

Embora o Webometric Analyst tenha apresentado uma maior cobertura de dados altmétricos para os artigos de periódicos brasileiros em CI, possivelmente pela forma de busca de dados utilizada – neste caso autor, título e ano –, foram percebidas algumas inconsistências no Mendeley, como um artigo com diferentes versões ou com datas de publicação distintas, mostrando divergentes contagens de leitores para um mesmo artigo. Os erros em DOIs e algumas inconsistências percebidas podem ter causado limitações e/ou discrepâncias nos resultados encontrados.

Realizado este levantamento e análise, percebe-se que o *Webometric Analyst* apresenta uma boa cobertura de dados altmétricos para publicações brasileiras em Ciência da Informação para a plataforma de mídia social Mendeley, especialmente pela forma de busca e coleta, porém, esta análise é restrita levando em consideração que só foram obtidos registros do Mendeley. Constatou-se também que a ausência de identificadores digitais na totalidade de artigos de periódicos brasileiros em CI dificulta a coleta e padronização de dados altmétricos, levando em consideração que a ferramenta atualmente mais difundida para indicar dados altmétricos, o Altmetric.com, só pode ser realizada através de identificadores digitais.

CONCLUSÃO

O ambiente da comunicação científica tem sido reconfigurado especialmente pelo uso, difusão e compartilhamento de produtos de pesquisa em plataformas de mídias sociais e os estudos altmétricos têm se desenvolvido com o propósito de coletar os rastros invisíveis não mensurados pelos tradicionais sistemas de avaliação da pesquisa.

As ferramentas utilizadas para coleta e análise de dados altmétricos devem ser entendidas como facilitadores de dados em plataformas de mídias sociais, porém deve-se estar atento às abrangências e limitações destas ferramentas, assim como a coleção escolhida para teste e captura, entendendo que a natureza das plataformas são divergentes e que não necessariamente foram desenvolvidas para a difusão de conteúdos científicos.

REFERÊNCIAS

ADIE, E.; ROE, W. Altmetric: enriching scholarly content with article – level discussion and metrics. *Learned Publishing*, Bristol, v. 26, n. 1, p. 11-17, 2013.

ARAÚJO, R. F. Mídias sociais e comunicação científica: análise altmétrica em artigos de periódicos da ciência da informação. *Em Questão*, Porto Alegre, v. 21, n. 1, 2015.

BARTLING, S.; FRIESIKE, S. Towards Another Scientific Revolution. In: BARTLING, S.; FRIESIKE, S. (ed.). *Opening Science: the evolving guide on how the Internet is Changing Research, Collaboration and Scholarly Publishing*. Heidelberg: Springer Open, 2014. p. 3-15.

DIGITAL SCIENCE. Altmetric: Who's talking about your research?. *Altmetric*, [s. l.], 2016. Disponível em: <https://www.altmetric.com>. Acesso em: 10 set. 2016.

- ERDT, M.; NAGARAJAN, A.; SIN, J. S. C. *et al.* Altmetrics: an analysis of the state-of-the-art in measuring research impact on social media. *Scientometrics*, New York, v. 109, n. 2, p. 1117-1166, 2016.
- HAUSTEIN, S.; BOWMAN, T. D.; COSTAS, R. Interpreting altmetrics: Viewing acts on social media through the lens of citation and social theories. In: SUGIMOTO, C. R. (ed.). *Theories of Informetrics and scholarly communication*. Frankfurt: De Gruyter, 2016. p. 1-24.
- MANHIQUE, I. L. E.; RODRIGUES, F. A. Indicadores altmétricos em periódicos brasileiros da Ciência da Informação no Brasil. *Revista Ibero-Americana de Ciência da Informação*, Brasília, DF, v. 12, n. 2, p. 515-532, 2019.
- MESCHEDE, C.; SIEBENLIST, T. Cross-metric compatibility and inconsistencies of altmetrics. *Scientometrics*, New York, v. 115, n. 1, p. 283-297, 2018.
- MUKHERJEE, B.; SUBOTIC, S.; CHAUBEY, A. K. And now for something completely different: the congruence of the Altmetric Attention Score's structure between different article groups. *Scientometrics*, New York, v. 114, n. 1, p. 253-275, 2018.
- NASCIMENTO, A. G.; ODDONE, N. E. Uso de altmetrics para avaliação de periódicos científicos brasileiros em Ciência da Informação. *Ciência da Informação em revista*, Maceió, v. 2, n. 1, p. 3-12, 2015.
- NASCIMENTO, A. G.; ODDONE, N. E. Uso de indicadores altmetrics na avaliação de periódicos científicos brasileiros em Ciência da Informação. In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 4., 2014, Recife. *Anais [...]*. Recife: UFPE, 2014. p. 1-2.
- NATIONAL INFORMATION STANDARDS ORGANIZATION. *NISO RP-25: outputs of the NISO Alternative Assessment Metrics Project*. Baltimore: NISO, 2016.
- PRIEM, J.; TARABORELLI, D.; GROTH, P. *et al.* Altmetrics: A manifesto. *Altmetrics*, [s. l.], 2010. Disponível em: <http://altmetrics.org/manifesto>. Acesso em: 30 jul. 2017.
- SANKAR, P.; KAVITHA, E. S. Altmetrics: new research tool in the growth social media. *E-Journal of Library and Information Science*, Madison, v. 3, n. 1, 2016.
- THELWALL, M. *Introduction to Webometrics: quantitative Web Research for the Social Sciences*. San Rafael: Morgan & Claypool, 2009.

