

MODELAGEM DE INDICADORES ALTMÉTRICOS: uma análise da influência do tempo de publicação

Sérgio Salustiano da Silva

Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, RJ, Brasil
sallustiano@gmail.com

Fábio Castro Gouveia

Fundação Oswaldo Cruz, Rio de Janeiro, RJ, Brasil
fgouveia@gmail.com

1 INTRODUÇÃO

Os estudos relacionados à Ciência da Informação demonstram potencial para várias análises, inclusive com novos cenários a partir de abordagens cibernéticas. Para contribuir com esses estudos este trabalho apresenta a definição de três indicadores (Exposição, Interação e Apropriação) que visam auxiliar no acompanhamento do impacto da produção científica e também identificar como é o consumo destes conteúdos nas redes sociais, pois periódicos com maior visibilidade tendem a gerar maior impacto, enquanto que periódicos e conteúdos com menor influência entre os seus pares tendem a ser menos divulgados. Assim, diante desse cenário são apresentados os conceitos que balizaram a pesquisa e os códigos para manipular os dados extraídos da plataforma *Altmetric.com*.

2 METODOLOGIA E TRATAMENTO DOS DADOS

A pesquisa foi alicerçada na cibermetria por considerar para a análise dos dados as inter-relações entre os conteúdos e os processos comunicativos, além de também se ater nas particularidades de consumo, ferramentas e serviços utilizados pelos usuários. Ao mesmo tempo, o subcampo inclui “a aplicação das tradicionais técnicas informétricas a qualquer tipo de informação disponível na internet” (SIQUEIRA, 2016, p.36).



O *corpus* da análise foi constituído do de 3.000 documentos (artigos, livros, registros de ensaios clínicos, conjunto de dados, capítulos de livros e notícias) recuperados pela API da ferramenta *Altmetric.com*. A coleta dos dados foi realizada entre os dias 1 a 30 de junho de 2017 quando foram baixados diariamente os 100 artigos com maior índice de atenção no dia.

A tabulação dos dados foi realizada com o *Google Refine* utilizado para converter arquivos provenientes da ferramenta da *Altmetric.com* do formato *.JSON* para *.CSV* (os dados em estado bruto estão disponíveis no *Figshare* (<https://figshare.com/s/9Icf3e6Of8250aecfa5f>), em seguida a modelagem dos dados foi realizada na ferramenta *Rstudio*, versão 1.1.414.

A partir desse cenário, observamos que o modelo adequado para taxonomia própria de exploração científica das redes sociais deve estar aliado com as métricas que podem ser utilizadas por cada grupo de ferramentas. Também foi levado em consideração a utilização de modelos de mensuração consolidados no mercado como a *Barcelona Declaration of Measurement Principles* (AMEC, 2010), cujo documento compilou um conjunto de sete diretrizes para medir os ambientes digitais; e o *The Social Media Measurement Conclave* (2013), uma coalizão de empresas *Business to Business* (B2B) e *Business to Consumer* (B2C), agências de relações públicas e de mídia social e associações industriais, que publicaram o *Complete social media measurement standards*. Com base nesses documentos foi possível extrair três indicadores que servirão para agregar os serviços e categorizar as redes sociais segundo as suas apropriações gerais, conforme apresentadas na Figura 1:



FIGURA 1 - DIVISÃO DE CATEGORIAS UTILIZADAS PARA AGRUPAR AS REDES SOCIAIS



Fonte: Elaborado pelos autores, 2018.

- **Exposição** – São considerados indicadores de exposição as publicações que sejam apenas para a divulgação científica. Esse conteúdo identificado como publicações de primeiro nível, geralmente tem o objetivo de apresentar, ou divulgar resultados de uma pesquisa, ou direcionar o público para um conteúdo específico.
- **Interação** – Definido como uma interação de segundo nível e envolve uma ação em resposta ao conteúdo disponibilizado. Esse tipo de envolvimento pode ser identificado por meio de comentários, incorporação de conteúdo etc.
- **Apropriação** – Definido como a Apropriação de terceiro nível cuja ação pode ser pela apropriação da pesquisa original para o desenvolvimento ou aplicação em uma citação ao trabalho origi-



nal, em artigos, palestras, desenvolvimento de produtos comerciais, matérias em revistas especializadas, entre outros.

O objetivo de dividir em três categorias amplas foi por entendermos a necessidade de um apuramento constante, mas, também, foi levado em consideração o objetivo do ambiente, a apropriação do público e o tipo de dado que poderia ser extraído. Com essa premissa, os dados capturados passaram por uma primeira análise para verificar a estrutura dos dados, possíveis duplicatas e a ausência de alguma informação que inviabilizasse a análise. Nesse ponto foram excluídos 167 registros que não tinham as datas de entrada no sistema ou de publicação e 1.154 que eram duplicados.

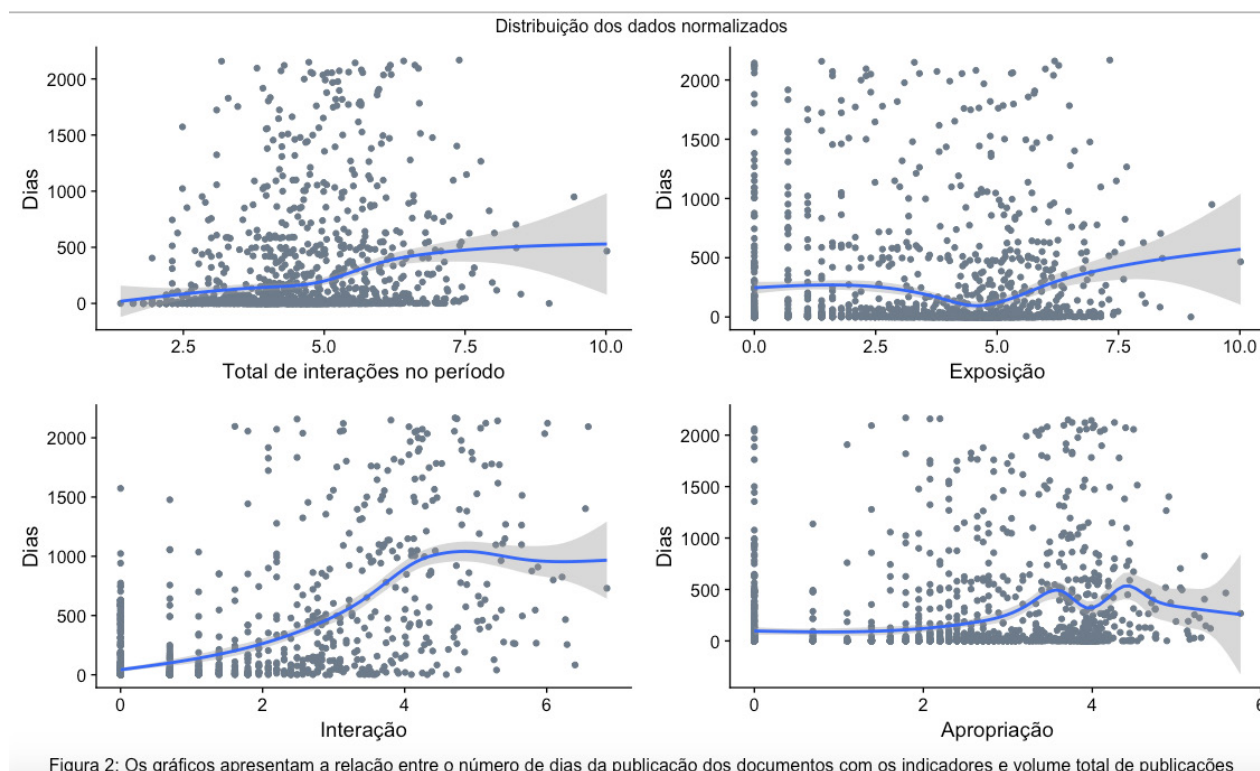
Após essa etapa foram somados os valores de cada rede, para extrair o valor que representasse os Índices (Exposição, Interação e Apropriação), que compreenderam as seguintes redes:

- **Exposição:** *Twitter*, páginas do *Facebook*, *Google+*, *Weibo* e Vídeos;
- **Interação:** *Mendeley*, *Connotea*, *CiteUlike*, *Reddit*;
- **Apropriação:** Notícias, Blogs, Páginas governamentais, *Wikipédia*, *LinkedIn*, *Syllabi*, *RH*, *Q&A*, *Peer_Review*.

Enquanto que para o tempo médio de vida dos documentos extraídos foi subtraído a data de publicação pela data da última atualização, antes do documento ser baixado.

3 MODELAGEM DOS DADOS DAS REDES SOCIAIS

Com base nas categorias Exposição, Interação e Apropriação buscou-se identificar como é o consumo de conteúdos científicos nas redes sociais e foram levantadas duas questões: qual o impacto do tempo de publicação para a disseminação do conteúdo e, principalmente, se existe transferência na utilização dos documentos de um indicador para outro.



Fonte: Elaborado pelos autores, 2018.

Com base nos questionamentos foi identificado uma concentração que a variável tempo tem pouco impacto nos indicadores, uma vez que ao observamos os gráficos Exposição Interação e Apropriação (Figura 2) constata-se dinâmicas distintas dentro do mesmo espaço de tempo. Dessa forma, ao mesmo tempo que esse resultado poderia sugerir uma lacuna ou perda durante a transferência de informações, não podemos deixar de reforçar que os Índices Exposição, Interação e Apropriação, em muitos casos, também, podem servir como primeiro impacto para o pesquisador. Além disso, é necessário cuidado ao comparar números obtidos em diferentes redes.

Embora o propósito desta pesquisa não seja o aprofundamento em como as diversas áreas estão se apropriando dos ambientes digitais na divulgação acadêmica, um ponto que deve ser levado em consideração é como cada campo utiliza e com qual finalidade. Como exemplo, podemos citar Page (1999) ao analisar a forma e modelo de publicação na Ciência da Computação ser fortemente baseada na experimentação e a publicação em congressos e conferências é preferida à publicação em revistas. Um



outro dado que ajudar a respaldar esse argumento é o perfil da área que se concentra em redes como *Stack Exchange* e *Reddit*, ao invés do *Twitter*.

4 CONSIDERAÇÕES FINAIS

A pesquisa partiu do pressuposto que existe a possibilidade de refinar e/ou apresentar de forma mais racional os dados oriundos das redes sociais para a produção científica. Dessa forma, foram sugeridos três indicadores que podem representar os caminhos utilizados pelo público para consumir pesquisas científicas. Em paralelo, foi testado o impacto do tempo de publicação na disseminação desses conteúdos, com o objetivo de buscar uma possível relação entre idade dos documentos e volume de envolvimento. Em ambos os casos, foi possível observar que: a) é possível a utilização de novos indicadores para avaliar a dispersão científica, pois apresentam maior clareza na utilização dos documentos pelo público; b) ao avaliar o impacto do tempo de publicação observou-se que a dinâmica própria revelada nos indicadores serve para demonstrar a necessidade e utilidade de pesquisas que priorizem a dinâmica do público dentro destes ambientes.

REFERÊNCIAS

- AMEC. **Barcelona declaration of measurement principles**: validated metrics; social media measurement. 2010. Disponível em: <https://amecorg.com/wp-content/uploads/2011/08/Barcelona_Principles_for_PR_Measurement.pdf>. Acesso em: 22 ago. 2017
- PAGE, A. Best Practices Memo Evaluating Computer Scientists and Engineers For Promotion and Tenure. Computing, 1999.
- PIWOWAR, H. Introduction altmetrics: what, why and where? **Bulletin of the Association for Information Science and Technology**, v. 39, n. 4, p. 8-9, 2013.
- SIQUEIRA, J. C. **Glossário de Neologismos da Ciência da Informação**. Clube de Autores, 2016.
- THE SOCIAL MEDIA MEASUREMENT CONCLAVE. The Conclave Complete Social Media Measurement Standards June 2013. Disponível em: <<http://>

6° EBBC

Rio de Janeiro
17 a 20 de julho



painepublishing.com/wp-content/uploads/2013/10/Complete-standards-document.pdf>. Acesso em: 05 jan. 2018.