

A aplicabilidade da Folksonomia nos estudos alométricos

The applicability of Folksonomy in altmetric studies

Skrol Salustiano¹, Fabio Castro Gouveia⁽²⁾

(1) IBICT-UFRJ, Rio de Janeiro, skrol@ufrj.br.

(2) Fiocruz, Rio de Janeiro, fabio.gouveia@fiocruz.br.

Resumo

A popularização e crescimento das redes sociais abriu novas possibilidades para a pesquisa científica. Nos estudos métricos da informação possibilitou o surgimento de novos campos de estudos como a Folksonomia, utilizada para etiquetagem social e a Altmetria, cuja premissa é medir a disseminação de documentos científicos nas redes sociais. **Objetivo:** Identificar a convergência entre os saberes e, principalmente, se é possível construir um modelo para o desenvolvimento de indicadores alométricos. **Metodologia:** Levantamento bibliográfico e de dados realizada no Twitter, onde foram coletadas publicações que tivessem a hashtag “openaccess”. O conteúdo foi analisado para identificar, principalmente, se existiam links para produções científicas. Dessa forma, tornou-se possível compreender a dinâmica da utilização das hashtags no ambiente acadêmico e como esses resultados podem ser aplicados nos Estudos Métricos da Informação. **Resultado:** Com os tweets capturados, analisados e classificados foi possível extrair informações sobre o tipo de conteúdo compartilhado e as áreas do saber predominantes. **Considerações Finais:** O dataset permitiu observar a possibilidade de construção de um modelo para identificar a convergência entre os saberes e, principalmente, se é possível o desenvolvimento de indicadores alométricos.

Palavras-chave: Folksonomia, Altmetria, Redes Sociais, Indexação, Estudos Métricos

Abstract:

The popularization and growth of social networks has opened new possibilities for scientific research. In information metrics studies it has enabled the emergence of new fields of study such as Folksonomy, used for social labeling, and Altmetrics, whose premise is to measure the dissemination of scientific documents on social networks. **Objective:** To identify the convergence between the two fields of knowledge and, especially, if it is possible to construct a model for the development of altmetric indicators. **Methodology:** Bibliographic and data survey conducted on Twitter, where publications that had the hashtag "openaccess" were collected. The content was analyzed to identify, mainly, if there were links to scientific productions. Thus, it became possible to understand the dynamics of the use of hashtags in the academic environment and how these results can be applied in Information Metrics Studies. **Result:** With the captured, analyzed and classified tweets it was possible to extract information about the type of shared content and the predominant knowledge areas. **Final Considerations:** The dataset allowed us to observe the possibility of building a model to identify the convergence between the knowledge areas and, especially, if it is possible to develop altmetric indicators.

Keywords: Folksonomy, Altmetrics, Social Networks, Indexing, Metric Studies

1. Introdução

Nas últimas duas décadas, com a consolidação dos ambientes digitais surgiram novas oportunidades para o acompanhamento do fazer ciência, por meio de dados que sinalizam uma possível relevância e/ou interesse ao assunto abordado. Como na previsão de Bossy (1995), a internet se tornou um campo para observação da “ciência em ação”.

Neste novo ambiente que se abriu para a pesquisa científica surgiram novas maneiras para organizar, representar e recuperar

informações, baseadas em rótulos (tags), que servem para marcar e agrupar conteúdos similares. Guedes e Dias (2010) ao analisarem a nomenclatura apropriada para a indexação de objetos digitais, observaram que “um mesmo documento é acessado, manipulado, organizado e recuperado por uma infinidade de usuários em diversas partes do mundo e ao mesmo tempo.” (GUEDES; DIAS, 2010, 47), o que favorece a descentralização na organização das informações.

Assim, esse sistema de atribuição de tags ou etiquetagem digital dos conteúdos surge como uma adaptação aos modelos de taxonomia e passou a ser chamado de Folksonomia¹. O diferencial é que dentro dessa nova dinâmica, as palavras-chave que serão indexadas passam a ser construídas e geridas pelos termos livres atribuídos pelos usuários ao compartilharem os conteúdos.

2. Objetivos

Com base nessa premissa, esse estudo ainda em andamento, apresenta os primeiros resultados, da pesquisa que busca analisar as potencialidades da Folksonomia aplicada aos Estudos Métricos da Informação (EMI), com foco na Altmétrie, para verificar se teoricamente existe a convergência entre os saberes que possa auxiliar na construção de um modelo prático para o desenvolvimento e modelagem de indicadores alométricos. Para atingir os objetivos, a pesquisa teve sua metodologia baseada na revisão narrativa, que identificar e discutir as inter-relações entre Altmétrie e Folksonomia. Para isso foram utilizados como bibliografia estudos recorrentemente citados, incluindo autores nacionais e internacionais.

3. Procedimentos Metodológicos

A pesquisa teve dois momentos distintos: o primeiro foi a análise do conceito da Folksonomia, quando buscou-se identificar as possibilidades de aplicação para a organização do conhecimento. Nesta etapa foi realizada a revisão narrativa, que buscou os avanços realizados em mais de uma década de estudos da Folksonomia.

Na segunda parte, em consonância com o objetivo da pesquisa de identificar o impacto da participação pública na organização do conhecimento, foi definido como termo de levantamento a hashtag "#openaccess".

O Twitter² foi escolhido como ambiente mais adequado para a realização do levantamento. A captura das informações foi

realizada com a ferramenta TAGSExplorer³, com período compreendido entre 20 de julho de 2022 a 2 de setembro de 2022. A janela temporal tinha como objetivo inicial completar o período de três meses. No entanto, ao atingir o volume de 200 mil tweets a ferramenta começou a apresentar erros na captura dos conteúdos. Dessa forma, para não prejudicar a qualidade do dataset a coleta foi interrompida antecipadamente, mas sem prejuízos a qualidade das informações coletadas que foram além do esperado.

Dentro da janela temporal foram extraídos 298.516 tweets ou publicações, que foram tratados e modelados na ferramenta Rstudio.

A primeira triagem ou visualização foi de identificar quais eram as principais hashtags associadas ao conteúdo coletado. Foi identificado muitos conteúdos com apenas uma única hashtag e em sua maioria não tinham links associados. Estes foram excluídos, pois não entregariam o principal resultado busca: qual tipo de informação estava sendo compartilhada nos links.

Em seguida, foram também eliminados da análise usuários com menos de cinco mil seguidores. Essa opção aconteceu pela identificação deste público ter média de apenas quatro retuites. Como o TAGSExplorer não oferece outra métrica de engajamento (curtidas, salvamentos e cliques nos links), esta métrica serviu para apontar a dispersão, apropriação e envolvimento com os conteúdos publicados.

Após a limpeza dos dados, o dataset resultou em um conjunto 19.437 tweets, cujas hashtags das publicações totalizaram 174.933, o que representa uma média de nove hashtag por publicação.

Esse volume de tweets selecionados ainda passou por outros tratamentos realizados no próprio Rstudio, com o objetivo de extrair o tipo de publicação, posteriormente classificadas em: Links para Artigos, Links Diversos, Links Mistos e Sem Links.

Dentro dos processos descritos acima o principal foi o de descompactação das URLs encurtadas. Para o processo foram utilizados os pacotes DypLR e Longurl. O primeiro foi

¹ Termo cunhado por Thomas Vander Wal (2007) que significa a junção das palavras: "folk" pessoas; e "taxonomia" = classificação. Ver também: Furner (2009)

² <https://tags.hawksey.info>

³ <https://twitter.com>

para copiar as URLs para uma nova coluna, no entanto na mesma linha. Em seguida foi utilizado o pacote Longurl que extraiu as urls originais.

Com base nos links foi possível identificar as revistas científicas ou periódicos e novamente com o DyplyR criar a relação entre tipo de periódico e área do saber. Os endereços que não pertenciam a publicações científicas foram classificados como "Diversos", pois a maioria era informativos de eventos (Congressos, seminários, webinars, entre outros). Essa classificação dos links diversos foi realizada de forma manual: foram agrupados os links similares e posteriormente foram analisados.

4. Resultados

Com base nos dados coletados observou-se que a maior parte do conteúdo foi para o compartilhamento de links. Deste volume, conforme pode ser observado no Apêndice A, do Gráfico 1, as publicações com links para documentos científicos representaram 24%.

Este resultado demonstra que existe um processo descentralizado de indexação proporcionado pela Folksonomia que cumpre "a principal finalidade de um sistema de recuperação da informação [que] é fornecer aos usuários a informação na forma e momento exigidos por eles" (CARNEIRO, 1985, p. 223).

Com essa premissa é possível observar que as redes sociais cumprem esse pré-requisito, pois permitem que informações sejam compartilhadas e recuperadas, seja no formato de um *tweet*, um post no Facebook ou uma imagem no Instagram. Essa diversidade de formatos segue em consonância com o pensamento de Leiva e Fujita ao afirmarem que "qualquer objeto pode ser indexado, ou seja, reduzido a representações conceituais que facilitem seu armazenamento e recuperação em bases de dados" (GIL LEIVA; FUJITA, 2012, p. 65).

Além disso, também demonstra a possibilidade da utilização das hashtags para a identificação do tipo de produto científico compartilhado, e a área do saber predominante, para cada tipo de área. Para os Estudos Métricos da Informação, esses

indicadores podem auxiliar na construção e desenvolvimento de novos sistemas capazes de auxiliar instituições de pesquisas, universidades e pesquisadores a incorporar camadas mais refinadas aos dados altmétricos.

Ao analisar o instrumental de ferramentas utilizadas é possível observar que o TAGSExplorer, para a captura de tweets, é uma ferramenta que entregou volume de conteúdos superior ao esperado. No entanto, após ultrapassar o limite de 200 mil tweets começou a enviar mensagens de erro, com a informação que a coleta seria refeita. Esse tipo de erro pode ser prejudicial para grandes volumes de dados. No entanto, com essa observação, para trabalhos posteriores serão realizadas quebras temporais, para evitar que aconteça a sobrecarga da ferramenta.

Enquanto, a limpeza, modelagem e análise dos conteúdos com a utilização do software R, demonstrou grandes possibilidades para esse tipo de estudo. As bibliotecas e/ou pacotes auxiliaram na modelagem dos dados, principalmente, no caso desta pesquisa, na leitura de links encurtados.

4.1 Discussão teórica

A popularização dos meios digitais trouxe para o grande público a possibilidade de abandonar o papel de consumidor passivo de informações para ser o seu produtor. Essa nova realidade, diferente da percepção de Johnson (2003) que acreditou que esse novo ambiente acabaria por ser tornar um "caos informacional", na prática acabou por criar uma "indexação democrática", conforme postulado por Brown, Hilderley, Griffin e Rollason (1996), ao desenvolver um projeto para a indexação e recuperação de imagens, por pessoas comuns.

Essa democratização também foi identificada por Lévy (2010) ao definir a contribuição do público como "inteligência coletiva", pois existe uma contribuição baseada em percepções e inteligência individual. A percepção é complementada por Gil Leiva (1997) ao destacar que a explosão informacional proporcionou que na prática a indexação também fosse realizada por computadores, em um processo parcialmente

ou totalmente automatizado. Enquanto Santos (2019) afirma que a Folksonomia pode ser utilizada como um conceito que “pode contribuir para os estudos métricos da informação científica” (SANTOS, 2019, p. 2).

Ao mesmo tempo, com esses dados preliminares, é possível identificar como as áreas do saber estão envolvidas tanto com o movimento do Open Access, como na utilização dos ambientes digitais para o compartilhamento e troca de informações sobre conteúdos científicos.

No Apêndice A, Gráfico 2, é possível identificar pelas hashtags a predominância de duas áreas: Ciências Exatas e da Terra e Ciências Biológicas, no compartilhamento de conteúdos.

4.2 A Folksonomia como suporte a indexação tradicional

Como afirmou Bossy (1995, não paginado, tradução nossa) “a diversidade de novos padrões de comunicação na rede eletrônica às vezes rompe as fronteiras entre circulação formal e informal, entre atividades que ocorrem dentro e fora dos laboratórios”. A premissa permite trazer para o debate a Folksonomia e a Almetria como campos ainda em fase de amadurecimento, mas que carregam em suas bases a utilização de técnicas e modelos tradicionais, adaptados para o novo ambiente. Essa utilização acaba por não gerar rupturas ou grandes impactos na forma como a academia e os pesquisadores observam nos resultados apresentados.

Extrair informações quantificáveis das redes sociais, por meio das *tags* vinculadas aos conteúdos demonstra ser a base para os estudos sobre a Folksonomia. No entanto, observa-se que o modelo adequado para taxonomia própria de exploração científica das redes sociais deve estar aliado às suas dinâmicas, ao invés, de apenas tentar buscar uma aproximação ou uma adaptação dos modelos tradicionais para esse novo ambiente.

Essa nova visão sobre a Folksonomia de não se debruçar somente sobre as tags, mas se apropriar de outros elementos, como os links, favorece os estudos do campo e

pode auxiliar no desenvolvimento dos EMI em ambientes sociais.

5. Considerações Finais

Nesta pesquisa demonstramos possibilidades, pontos positivos e negativos na utilização da Folksonomia e, também, apresentamos questões quanto à nova subárea dos Estudos Métricos da Informação: a Almetria. Em comum ambos são utilizados para identificar e classificar publicações geradas em ambientes digitais, principalmente, as redes sociais. No entanto, a convergência entre ambos, apesar de ser possível, ainda é negligenciada.

Assim, na pesquisa conseguimos observar a possibilidade de modelar um novo sistema de aferição alométrica enriquecido pela Folksonomia, que permitem a delimitação de termos capturados para a montagem de índices, rankings de produtividades, dentre outros levantamentos sobre a repercussão da produtividade científica nas redes sociais. Ao mesmo tempo, esse tipo de modelagem requer o aprimoramento dos sistemas de recuperação de recursos informacionais, como forma de refinar a recuperação dos descritores sociais.

Porém, salientamos que a pesquisa se desenvolve somente com uma perspectiva preliminar e sua viabilidade necessita de um modelo prático e funcional, que é a próxima etapa do estudo. No entanto, essas primeiras observações servem como base para o desenvolvimento de modelos funcionais, além de se inserir e estar alinhado aos debates sobre Gestão da Informação nos ambientes digitais, na transparência dos indicadores métricos e em novas metodologias para os Estudos Métricos da Informação.

6. Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e Conselho Nacional de Desenvolvimento Científico e Tecnológico, Processo 430982/2018-6 e 315521/2020-1.

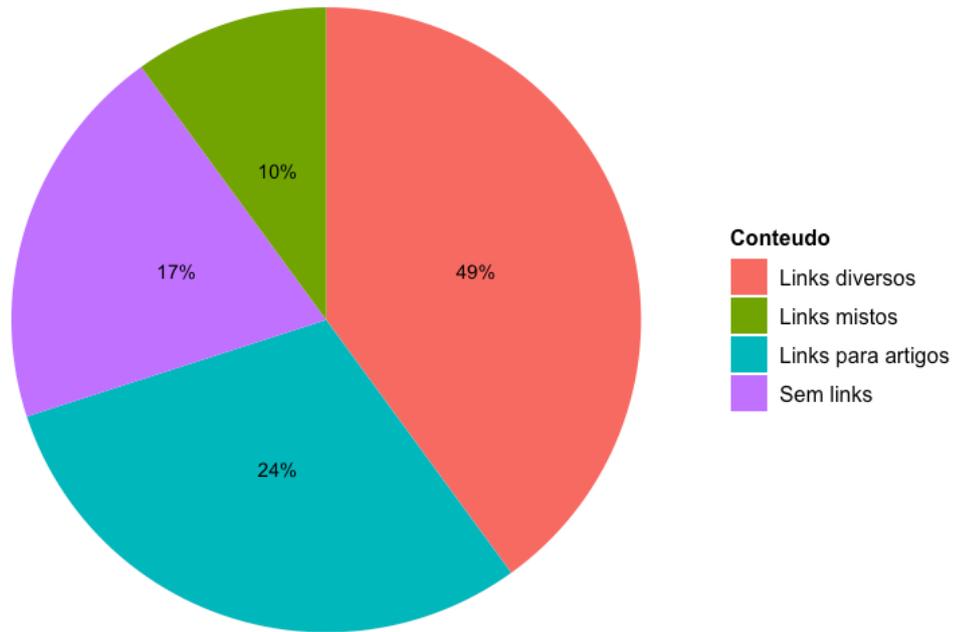
7. Referências

- BOSSY, M. J. The Last of the Litter: Netometrics. 1995. Disponível em: <http://gabriel.gallezot.free.fr/Solaris/d02/2bos sy.html>. Acesso em: 25 ago. 2022.
- BROWN, Pauline; HIDDENLEY, Rob; GRIFFIN, Hugh; ROLLASON, Sarah. The democratic indexing of images. **New Review of Hypermedia and Multimedia**, v. 2, n. 1, p. 107-120, 1996.
- CARNEIRO, MARÍLIA VIDIGAL. Diretrizes para uma política de indexação. **Revista da Escola de Biblioteconomia da UFMG**, v. 14, n. 2, 1985.
- DORA, S. F. San Francisco declaration on research assessment. 2012. Disponível em: <https://sfdora.org/read/read-the-declaration-portugues-brasileiro/>. Acesso em: 11 ago. 2021.
- FURNER, Jonathan. Folksonomies. In: **Encyclopedia of Library and Information Sciences, Third Edition**. CRC Press, 2009. p. 1858-1866
- GIL LEIVA, Isidoro; FUJITA, Mariângela Spotti Lopes (Ed.). **Política de indexação**. Editora Oficina Universitária, 2012.
- GIL LEIVA, Isidoro. **La automatización de la indización, propuesta teórico-metodológica: aplicación al área de biblioteconomía y documentación**. Murcia, España: Universidad de Murcia, 1997.
- GUEDES, R. de M.; DIAS, Eduardo José Wense. Indexação social: abordagem conceitual. **Revista ACB: Biblioteconomia em Santa Catarina, Florianópolis**, v. 15, n. 1, p. 39-53, 2010.
- HAUSTEIN, S.; BOWMAN, T. D.; COSTAS, R. Interpreting “altmetrics”: viewing acts on social media through the lens of citation and social theories. In: **Theories of Informetrics: A Festschrift in Honor of Blaise Cronin**. Berlin: Cassidy R. Sugimoto (Ed.), 2015. p. 372–405.
- HAWKSEY, M. **TAGS**. , 2014. Disponível em: [<https://tags.hawksey.info/>](https://tags.hawksey.info/)
- HOFFMANN, C. P.; LUTZ, C.; MECKEL, M. A relational altmetric? Network centrality on ResearchGate as an indicator of scientific impact. **Journal of the Association for Information Science and Technology**, v. 67, n. 4, p. 765–775, 2016.
- HOTH, Andreas. Information retrieval in folksonomies: Search and ranking. In: **European semantic web conference**. Springer, Berlin, Heidelberg, 2006. p. 411-426.
- JOHNSON, Steven. **Emergência: a dinâmica de rede em formigas, cérebros, cidades e softwares**. Editora Schwarcz-Companhia das Letras, 2003.
- MAI, Jens-Erik. Folksonomies and the new order: authority in the digital disorder. **KO KNOWLEDGE ORGANIZATION**, v. 38, n. 2, p. 114-122, 2011.
- LEVY, Pierre. **Cibercultura**. Editora 34, 2010.
- PANG, B.; LEE, L.; VAITHYANATHAN, S. Thumbs up? Sentiment Classification using Machine Learning Techniques. In: **PROCEEDINGS OF THE ACL-02 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING - EMNLP '02 2002**, Morristown, NJ, USA. **Anais [...]**. Morristown, NJ, USA: Association for Computational Linguistics, 2002. Disponível em: <http://portal.acm.org/citation.cfm?doid=1118693.1118704>. Acesso em: 18 ago. 2022.
- RSTUDIO TEAM. **RStudio: Integrated Development Environment for R**. : Prairie Trillium.Boston, MARStudio, PBC, , [s.d.]. Disponível em: [<https://www.rstudio.com>](https://www.rstudio.com)
- SANTOS, Raimunda Fernanda dos. A Folksonomia e o seu impacto na comunicação científica. **Revista Conhecimento em Ação**, v. 4, n. 2, p. 1-3, 2019.

ZOLLER, Daniel. Posted, visited, exported:
Altmetrics in the social tagging system
BibSonomy. **Journal of Informetrics**, v. 10,
n. 3, p. 732-749, 2016.

8. Apêndice A – Gráficos

Gráfico 1 - Perfil do conteúdo gerado pelo público do Twitter



Dados da Pesquisa, 2022

