



24° ENANCIB

Encontro Nacional de Pesquisa em Ciência da Informação
Perspectivas Contemporâneas na Ciência da Informação
• Vitória - ES • Ancib • PPGCI/UFES



XXIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXIV ENANCIB

ISSN 2177-3688

GT Especial

**CARACTERIZANDO A ATENÇÃO ONLINE DE PESQUISAS SOBRE COVID-19 NO BRASIL:
GRAU DE EXPOSIÇÃO, GÊNERO E LOCALIZAÇÃO DOS USUÁRIOS NO TWITTER(X)**

***CHARACTERIZING ONLINE RESEARCH ATTENTION ABOUT COVID-19 IN BRAZIL:
DEGREE OF EXPOSURE, GENDER AND LOCATION OF USERS ON TWITTER(X)***

Ronaldo Ferreira de Araújo – Universidade Federal de Alagoas (UFAL)

Karlson Tellicio Bezerra de Lima – Laboratório de Estudos Métricos da Informação na Web
(Lab-iMetrics)

Modalidade: Resumo Expandido

Resumo: A pandemia de COVID-19 trouxe desafios sem precedentes para a sociedade global, destacando a importância de pesquisas científicas para entender e combater o vírus. A disseminação destas pesquisas nas mídias sociais desempenhou um papel significativo na conscientização pública e na orientação de políticas de saúde. Este estudo exploratório visa analisar a atenção online dada a pesquisas sobre Covid-19 compartilhadas no Brasil durante a pandemia (2020-2022), examinando no Twitter(X), o grau de exposição das publicações, o gênero e a localização geográfica dos usuários. As pesquisas foram recuperadas na base Dimensions e os dados de atenção obtidos via Altmetric e API do Twitter(X) com análises e visualizações no Python. Foram identificadas 17.600 pesquisas, com atenção de 62.979 usuários e 234.346 tweets. O grau de exposição é de pouco mais de 183 milhões, tendo mais usuários homens (47,2%), que mulheres (23,9%) e maior concentração de tweets na Região Sudeste (58,2%) e de usuários em São Paulo (23,32%).

Palavras-chave: altmetria; Covid-19; Brasil; atenção online; Twitter(X).

Abstract: The COVID-19 pandemic has brought unprecedented challenges to global society, highlighting the importance of scientific research to understand and combat the virus. The dissemination of this research on social media has played a significant role in raising public awareness and guiding health policy. This exploratory study aims to analyze the online attention given to research on Covid-19 shared in Brazil during the pandemic (2020-2022), examining on Twitter(X), the degree of exposure of publications, the gender and geographic location of users. The searches were retrieved from the Dimensions database and the attention data obtained via Altmetric and Twitter(X) API with analyzes and visualizations in Python. 17,600 searches were identified, with the attention of 62,979 users and 234,346 tweets. The degree of exposure is just over 183 million, with more male users (47.2%) than female users (23.9%) and a greater concentration of tweets in the Southeast Region (58.2%) and users in São Paulo (23.32%).

Keywords: altmetrics; COVID-19; Brazil; online attention; Twitter(X).

1 INTRODUÇÃO

A pandemia de Covid-19 representou um desafio sem precedentes para a saúde pública global, afetando milhões de pessoas e alterando significativamente a dinâmica social e econômica em todo o mundo. No Brasil, um dos países mais impactados (Silva; Lima, 2021) a disseminação de informações sobre o vírus, suas variantes, medidas de prevenção e tratamentos se tornou importante para combater a desinformação e orientar a população.

A visibilidade e o alcance das pesquisas sobre Covid-19 compartilhadas em mídias sociais como o Twitter(X) foram essenciais para garantir que informações precisas e atualizadas chegassem a um público amplo e diversificado (Bashirian *et al.*, 2023). No entanto, entender quem está compartilhando essas informações, como estão sendo disseminadas e qual é o impacto dessas ações nas diferentes regiões do Brasil ainda é uma área que necessita de investigação. Diante disso questionamos: Como as pesquisas sobre Covid-19 foram disseminadas no Twitter(X) no Brasil, e qual é o perfil dos usuários envolvidos nessa disseminação?

O objetivo do estudo é analisar a atenção online dada a pesquisas sobre Covid-19 compartilhadas no Brasil, caracterizando essa atenção em termos de grau de exposição, gênero e localização geográfica dos usuários.

O Brasil é o sétimo país do mundo que mais disseminou pesquisas sobre Covid-19 no Twitter(X) (Bashirian *et al.*, 2023) e ao compreender melhor o papel desta mídia, identificando os perfis dos usuários mais engajados e as regiões com maior atividade, este trabalho pode contribuir para o desenvolvimento de estratégias de comunicação mais eficazes e direcionadas. Além disso, ao fornecer dados sobre o alcance e a visibilidade das pesquisas científicas, o estudo pode auxiliar na promoção de práticas informadas de saúde pública e no combate à desinformação.

2 MATERIAL E MÉTODO

Trata-se de uma pesquisa exploratória e descritiva desenvolvida no âmbito dos estudos de altmetria, termo guarda-chuva para o campo que analisa métricas geradas a partir de interações com pesquisas em fontes da web social (Barcelos; Maricato, 2023). A pesquisa tem o intuito de analisar a atenção online dada a pesquisas sobre Covid-19 compartilhadas no Brasil durante a pandemia, especialmente no Twitter(X).

As publicações sobre COVID-19 foram coletadas na base Dimensions usando a opção de busca completa e a *string*: "2019-nCoV" OR "COVID-19" OR "SARS-CoV-2" OR ("coronavirus" OR "corona virus"). Os dados foram exportados em arquivo .CSV e a lista de DOIs foi importada para o Altmetric.com para coletar os dados altmétricos das publicações. No Altmetric.com aplicamos o filtro para obter apenas menções no Twitter (X) do Brasil para as postagens entre 2020 e 2022, e exportamos um arquivo .CSV com dados de cada usuário <user_ID> e cada tweet <tweet_ID>. Além disso, recuperamos as informações sobre cada usuário (nome, localização, informações biográficas, seguidores) do Twitter (X) via *Application Programming Interface* (API). Com os dados dos usuários que compartilharam as pesquisas partiu-se para a medição da audiência potencial dos tweets e a caracterização em termos de gênero e geolocalização.

As comunidades de atenção em torno de publicações científicas no Twitter(X) são identificadas com base no engajamento com o artigo e no seu potencial alcance, estando este último relacionado ao público potencial do tweet e aferido pela soma dos seguidores dos usuários que indica seu grau de exposição (Haustein; Bowman; Costas, 2015; Wouters; Zahedi; Costas, 2018; Aperin; Gomez; Haustein, 2019).

Estudos anteriores desenvolvidos para a caracterização demográfica de usuários do Twitter e aferição de gênero encontraram muitas inconsistências e falta de padrão dos dados (Oliveira *et al.*, 2015) o que demanda um esforço de processamento de dados. Para tanto, as análises desses aspectos foram realizadas em Python (bibliotecas *Panda*, *Numpy*, *Regex* e *jsonlines* para o processamento de dados e *Matplotlib*, *GeoPandas* e *Seaborn* para visualizações) e linguagem R.

A base de dados dos usuários foi processada com Python com a finalidade de uniformizar informações. Para cada usuário, seu conjunto de caracteres (*string*) do campo de localização do usuário foi convertido para caracteres do Código Padrão Americano para o Intercâmbio de Informação (*American Standard Code for Information Interchange* - ASCII) equivalentes, removendo qualquer caractere que não seja alfabético ou espaço. Espaços em branco duplicados são removidos e a *string* é convertida para minúsculas, retornando a localização limpa e normalizada. Para atribuir a localização de um usuário de forma uniforme, é feita uma checagem em seu campo de localização: verifica-se, para todos os nomes de cidades listados no Censo Demográfico 2022 do Instituto Brasileiro de Geografia e Estatística (IBGE, 2022) se todas as palavras que o constituem estão contidas no campo de localização.

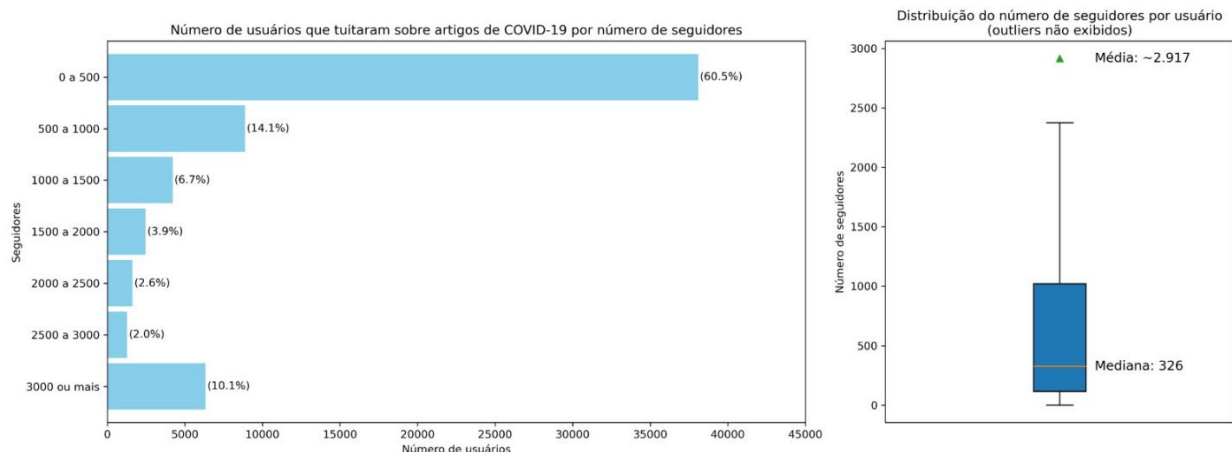
Para possibilitar análises de gênero de usuários, o conjunto de caracteres (*string*) que compõem o campo 'nome' foi processado, convertendo-os em caracteres ASCII equivalentes, em minúsculo, removendo-se espaços em branco duplicados. Foi utilizada a biblioteca de R chamada "genderBR" [<https://cran.r-project.org/web/packages/genderBR/index.html>] para prever o gênero do usuário a partir da primeira palavra que constitui o campo de nome, removendo títulos como "Profa." ou "Dra.". Após isso, cerca de 3% dos usuários tiveram seu gênero inferido com outra biblioteca, desta vez em Python, chamada Gender Guesser [<https://gender-guesser.com/>].

3 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Os dados coletados indicam um total de 17.600 pesquisas foram compartilhadas por 62.979 usuários que juntos postaram 234.346 tweets. Destacamos nos resultados o impacto potencial das postagens com base no número de seguidores e as diferenças na participação entre gêneros e regiões. Para tanto, apresentamos uma descrição detalhada dos principais achados, divididos em três categorias principais: grau de exposição, aferição de gênero dos usuários e localização geográfica dos compartilhamentos.

O grau de exposição encontrado, pela soma dos seguidores é de pouco mais de 183 milhões, tendo média de 2.917 e mediana de 326. A distribuição do nível de exposição está representada no Gráfico 1 considerando a variação entre as faixas de seguidores por usuário (A) e indicação de média e mediana (B).

Gráfico 1 – Distribuição do nível de exposição por faixas de seguidores dos usuários



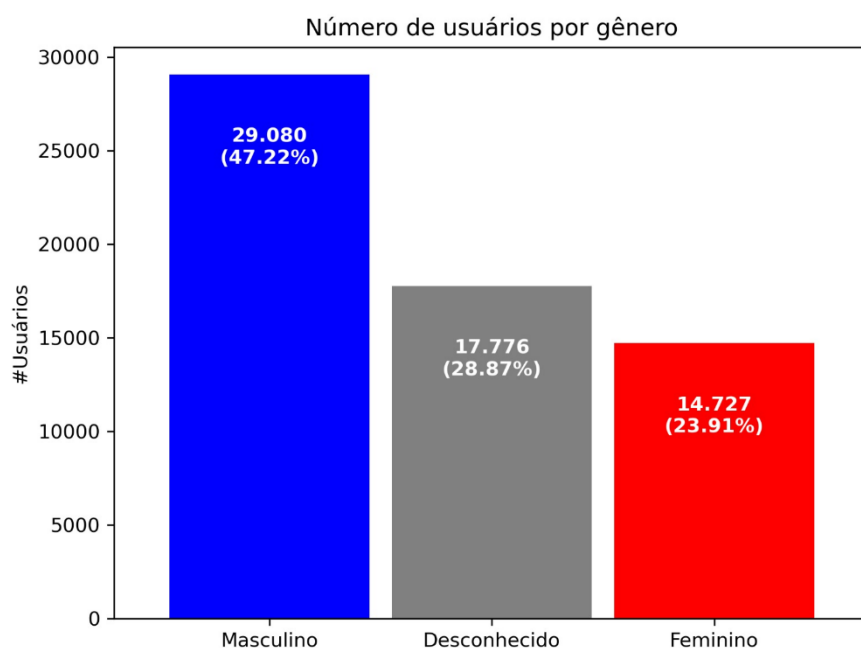
Fonte: dados da pesquisa (2024).

A exposição, conforme colocada por Haustein, Bowman e Costas (2015) e Alperin, Gomez e Haustein, 2019) indica, sob um aspecto estrutural do Twitter, o público potencial que um tweet pode atingir, dado o número de seguidores do usuário. Embora a maioria (74,6%) dos usuários possuam até 1.000 seguidores, vale destacar que na faixa de 3.000 seguidores ou mais, encontram-se 17 usuários com mais de 1 milhão de seguidores.

Adicionalmente, a análise de gênero dos usuários que compartilham informações sobre Covid-19 é importante para entender as dinâmicas de engajamento e participação, e identificar a predominância de gêneros na disseminação de informações pode revelar diferenças de comportamento e preferências na busca e compartilhamento de conhecimento científico.

O Gráfico 2 apresenta o perfil dos usuários por gênero. Dos 62.979 usuários identificados, 1.396 foram desconsiderados na representação por não terem mensagens (tweets) vinculadas, o que corresponde 2%. Assim o gráfico apresenta a distribuição dos 61.583 por gênero.

Gráfico 2 – Perfil dos usuários que postaram pesquisas sobre Covid-19 no Brasil por gênero



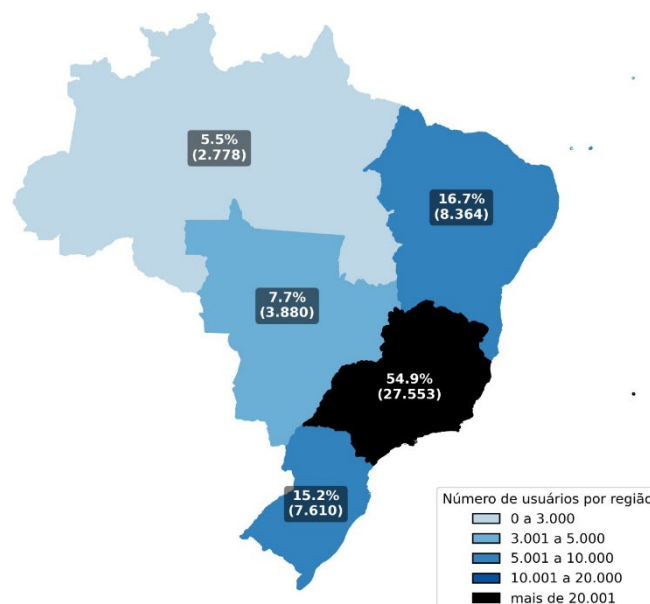
Fonte: dados da pesquisa (2024).

Os dados confirmam que 47,2% dos usuários são homens e 23,9% são mulheres. Estudos que analisaram geovisualizações de dados do Twitter incluindo dados demográficos e de gênero de usuários que compartilharam conteúdos sobre Covid-19 também identificaram mais homens do que mulheres em todas as faixas etárias analisadas, com proximidade de diferença similar ao que encontramos, está a faixa de 40 anos ou mais, com 22,4% de homens e 9,43 de mulheres (Lin *et al.*, 2023). O alto valor de perfis com gênero não identificado (28,8%) pode ser atribuído a contas não pessoais (empresas, coletivos, mídia, organizações, etc).

A aferição da localização geográfica dos usuários que compartilham pesquisas sobre Covid-19 (Gráfico 3), bem como a distribuição do volume das postagens (Gráfico 4), é essencial para entender como a disseminação de informações científicas sobre o tema varia entre diferentes regiões do Brasil e seus estados (Gráfico 5).

Dos 62.979 usuários identificados, não se identificou a localização de 13.856, o que corresponde a 22%. O Gráfico 3 apresenta a distribuição dos 49.123 usuários com indicação da localização geográfica.

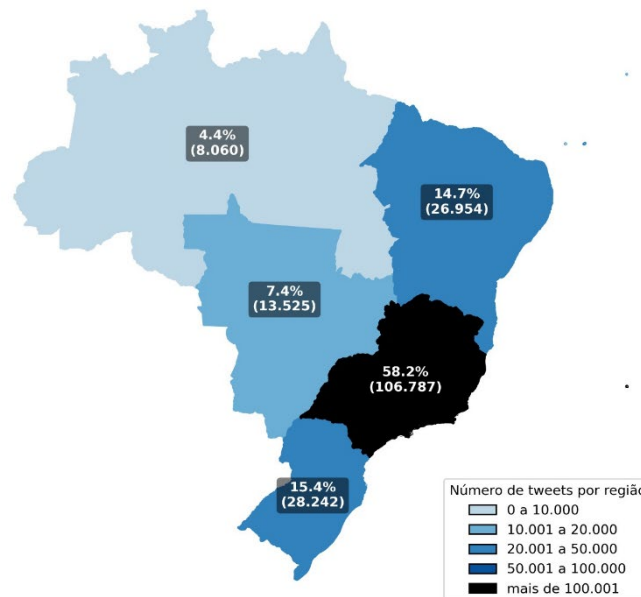
Gráfico 3 – Distribuição dos usuários que postaram pesquisas sobre Covid-19 no Brasil por Região



Fonte: dados da pesquisa (2024).

Dos 234.346 tweets identificados, 183.568 foram feitos por usuários cuja localização não foi identificada, o que corresponde a 21%. Assim o Gráfico 4 apresenta a distribuição de 50.778 tweets.

Gráfico 4 – Distribuição dos tweets sobre pesquisas sobre Covid-19 no Brasil por Região

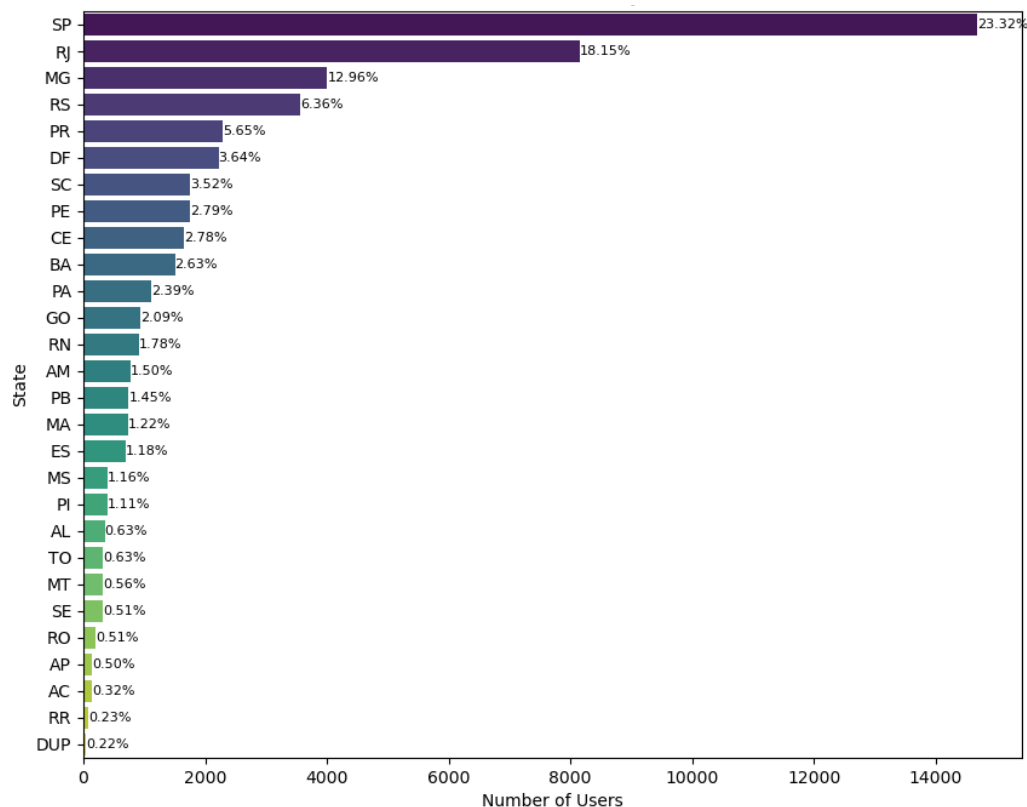


Fonte: dados da pesquisa (2024).

A análise dos dados indica uma relação direta entre o número de usuários e o volume de tweets publicados. A Região Sudeste se destaca tanto em número de usuários (54,9%) quanto em volume de postagens feitas (58,2%), com uma pequena alteração nesta relação quando se analisa a Região Nordeste que é a segunda em número de usuários (16,7%), mas a terceira em volume de tweets (14,7%), ficando atrás da Região Sul, que embora tenha a terceira maior concentração de usuários (15,2%), ocupa o segundo lugar em volume de tweets mencionando pesquisas sobre Covid-19 (15,4%).

As três regiões com maior número de usuários e tweets são as regiões que mais registraram casos confirmados e óbitos por Covid-19 durante a pandemia (2020-2022), sendo o Sudeste com aproximadamente 223,5 milhões de casos e 555 mil óbitos; o Nordeste com aproximadamente 13,8 milhões de casos e 355 óbitos; e o Sul com aproximadamente 9,5 milhões de casos e 235 óbitos (Brasil, 2023).

Gráfico 5 – Participação de usuários por estado



Fonte: dados da pesquisa (2024).

São Paulo ganha destaque com 23,2% dos usuários, seguido do Rio de Janeiro com 18,1% e Minas Gerais com 12,9%. Além de serem os três maiores estados brasileiros em termos de população, segundo o Censo de 2022 (IBGE, 2022), os três também aparecem com os maiores números de óbitos por Covid-19 durante a pandemia (Brasil, 2023).

4 CONSIDERAÇÕES FINAIS

A análise da atenção online às pesquisas sobre Covid-19 compartilhadas no Twitter(X) durante a pandemia no Brasil revelou *insights* significativos sobre o alcance, o perfil dos usuários envolvidos e a distribuição geográfica das interações.

A análise do grau de exposição indicou que, embora a maioria dos usuários (74,6%) tem até 1.000 seguidores, um pequeno grupo de usuários com mais de 1 milhão de seguidores desempenhou um papel significativo na ampliação do alcance das informações, o que destaca a importância dos influenciadores e das contas de grande audiência na disseminação de pesquisas científicas.

A aferição de gênero dos usuários revelou que 47,2% dos perfis são masculinos e 23,9% femininos, com uma alta proporção de perfis não identificados (28,8%), possivelmente devido à presença de contas institucionais e não pessoais. A análise geográfica dos compartilhamentos mostrou que a Região Sudeste se destacou tanto no número de usuários (54,9%) quanto no volume de tweets (58,2%), seguida pelas regiões Sul e Nordeste. A correlação entre o número de casos confirmados e óbitos por Covid-19 e a atividade nas mídias sociais sugere que as áreas mais afetadas pela pandemia tiveram maior engajamento na disseminação de informações científicas.

Estados como São Paulo, Rio de Janeiro e Minas Gerais, que possuem as maiores populações e os maiores números de óbitos por Covid-19, também foram os mais ativos em termos de compartilhamento de pesquisas. Este dado reforça a necessidade de estratégias de comunicação direcionadas, que considerem a distribuição demográfica e a gravidade do impacto da pandemia em diferentes regiões.

Em conclusão, este estudo destaca a importância de compreender a dinâmica de disseminação de informações científicas nas mídias sociais para promover uma comunicação mais eficaz e equitativa. A análise do grau de exposição, perfil de gênero e localização geográfica dos usuários fornece uma base sólida para o desenvolvimento de estratégias de comunicação que possam alcançar diferentes públicos de maneira mais eficiente, contribuindo para uma melhor compreensão e combate à pandemia de Covid-19 no Brasil.

Estudos futuros podem expandir essa análise, investigando o impacto das postagens na mudança de comportamento dos usuários e na adoção de medidas preventivas contra a Covid-19. Além disso, o trabalho buscará análises mais aprofundadas que qualifique ainda mais as redes de comunidades atenção identificadas, compreendendo melhor as semelhanças e diferenças entre os grupos, por meio de inclusão de recorte de raça por heteroidentificação e categorização de usuários; e os conteúdos (compreensão de estratégias de comunicação e análise do uso de hashtags, links e emojis).

A inclusão de outras plataformas de mídias sociais e a análise de temporalidade dos compartilhamentos podem oferecer uma visão mais abrangente das dinâmicas de disseminação de informações científicas.

REFERÊNCIAS

ALPERIN, Juan P.; GOMEZ, Charles J.; HAUSTEIN, Stefanie. Identifying diffusion patterns of research articles on Twitter: a case study of online engagement with open access articles. **Public understanding of science**, [S. l.], v. 28, n. 1, p. 2-18, Jan. 2019. Disponível em: <https://doi.org/10.1177/0963662518761733>. Acesso em: 4 out. 2023.

BARCELOS, Janinne; MARICATO, João de Melo. Altmertia: uma análise de seus termos, expressões, conceitos e definições. **Em Questão**, Porto Alegre, v. 29, p. e-129518, 2023. Disponível em: <https://doi.org/10.1590/1808-5245.29.129518>. Acesso em: 28 jan. 2024.

BASHIRIAN, Saeid *et al.* Estimating the online presence of COVID-19 papers in social media: An altmetric study. **J. Med. Libr. Inf. Sci.**, Tehran, v. 4, p. 1-11, 2023. Disponível em: <https://doi.org/10.22037/jmlis.v4i.39781>. Acesso em: 28 jan. 2024.

BRASIL. Ministério de Saúde. **COVID-19 –Painel Coronavírus**. [Brasília]: MS; [2023]. Disponível em: <https://covid.saude.gov.br/>. Acesso em: 12 abr. 2023.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Demográfico 2022**. Rio de Janeiro: Fundação Instituto Brasileiro de Geografia e Estatística. 2010. Disponível em: <https://censo2010.ibge.gov.br/resultados.html>. Acesso em: 10 abr. 2022.

LIN, Binbin *et al.* Sensing the pulse of the pandemic: geovisualizing the demographic disparities of public sentiment toward COVID-19 through social media. **ArXiv**, 2023. Disponível em: <https://arxiv.org/abs/2304.06120>. Acesso em: 28 jun. 2024.

OLIVEIRA, Derick M. de *et al.* Uma estratégia não supervisionada para previsão de eventos usando redes sociais. In: 30th Brazilian Symposium on Databases, 30., 2015, Petrópolis.. **Proceedings [...]**. Petrópolis: SBBd, 2015. p. 137-148.

SILVA, Hudson Pacífico da; LIMA, Luciana Dias de. Política, economia e saúde: lições da COVID-19. **Cadernos de Saúde Pública**. Rio de Janeiro, v. 37, n. 9, 2021. Disponível em: <https://doi.org/10.1590/0102-311X00200221>. Acesso em: 14 jun., 2024.

WOUTERS, Paul; ZAHEDI, Zohreh; COSTAS, Rodrigo. Social media metrics for new research evaluation. In: GLÄNZEL, W.; MOED, H.; SCHMOCH, U.; THELWALL, M. (ed.). **Handbook of quantitative science and technology research**. Berlin: Springer. 2018.