
**PANDEMIA DE COVID-19 NO BRASIL: UMA ANÁLISE SOBRE NOTÍCIAS E
COMENTÁRIOS DE USUÁRIOS**

COVID-19 PANDEMIC IN BRAZIL: AN ANALYSIS OF NEWS AND USERS' COMMENTS

PANDEMIA DE COVID-19 EN BRASIL: UN ANÁLISIS DE NOTICIAS Y COMENTARIOS DE USUARIOS

LUCAS DARLINDO FREITAS RODRIGUES¹

ANA LÚCIA PRADO²

FÁBIO MANOEL FRANÇA LOBATO³

Submissão: 28/01/2022
Aprovação: 05/03/2022
Publicação: 31/03/2022

¹ É aluno de graduação do curso de Ciência da Computação na Universidade Federal do Oeste do Pará (UFOPA). Atuou como membro do Laboratório de Computação Aplicada (LACA-UFOPA) com bolsa de Iniciação Científica e Tecnológica fomentada pelo CNPq. Possui experiência nas áreas da Análise de Mídias Sociais e Mineração de Textos.

Orcid ID: <https://orcid.org/0000-0001-9280-0547> E-mail: lucas.rodrigues@discente.ufopa.edu.br

¹ É Professora da Faculdade de Comunicação, da Universidade Federal do Pará. Coordenadora local do projeto de pesquisa nacional, financiado pelo CNPq, edital Universal 18/2021: "Índice de Credibilidade Jornalística: formulação de indicadores de fortalecimento do jornalismo para o combate aos ecossistemas de desinformação". Coordena o Laboratório de Multimídia e Dados (Labmuds/Facom).

Orcid ID: <https://orcid.org/0000-0002-8901-2491> E-mail: anaprado@ufpa.br

¹ É professor na Universidade Federal do Oeste do Pará e na Universidade Estadual do Maranhão. É líder do Grupo de Estudo e Pesquisa em Computação Aplicada e atualmente está como bolsista de Produtividade Desenvolvimento Tecnológico e Extensão Inovadora do CNPq. Tem especial interesse em Inteligência Computacional, Análise de Redes Sociais e Mineração de Opiniões.

Orcid ID: <https://orcid.org/0000-0002-6282-0368> E-mail: fabio.lobato@ufopa.edu.br

RESUMO

Em meio à pandemia do novo coronavírus, a proliferação da desinformação e de notícias falsas disputam espaços de fala em ambientes virtuais com o jornalismo, sobretudo, nas mídias sociais. Tal fenômeno inclui discursos de ódio e enunciados negacionistas, que põem em xeque a investigação científica sobre as consequências da contaminação em larga escala provocada pelo SARS-CoV-2. À luz desses fatos, este trabalho examina dois aspectos-chave desse fenômeno: qual o conteúdo produzido pelos usuários sobre notícias relacionadas à pandemia e como os internautas interagem com esse conteúdo. Para tal, cinco portais jornalísticos foram analisados utilizando-se de métodos de processamento de linguagem natural e de ciência de dados, como análise de sentimentos, modelagem de tópicos, análise de legibilidade e temporalidade. Os resultados alcançados permitem-nos apresentar um panorama da recepção das notícias e desinformação nos principais portais de notícias brasileiros.

Palavras-chave: Desinformação. Mídias sociais. Covid-19. Jornalismo on-line. Brasil.

ABSTRACT

Amidst the pandemic of the new coronavirus, the proliferation of misinformation and fake news, dispute spaces of speech in virtual environments with journalism, especially in social media. This phenomenon includes hate speech and denial statements, which call into question a scientific investigation of the contamination consequences on a large scale. Considering these facts, this work examines two critical aspects of this phenomenon: the content produced by users about news related to the pandemic and how internet users interact with this content. To this end, five journalistic portals were analyzed using natural language processing and data science methods, such as sentiment analysis, probability modeling, readability, and temporality analysis. The results obtained allow us to present an overview of the reception of news and disinformation in the leading Brazilian news portals.

Keywords: Misinformation. Social media. Covid-19. On-line journalism. Brazil.

RESUMEN

En medio de la pandemia del nuevo coronavirus, la proliferación de desinformación y noticias falsas disputan espacios de discurso en entornos virtuales con el periodismo, especialmente en las redes sociales. Este fenómeno incluye el discurso de odio y las declaraciones de negación, que cuestionan la investigación científica de las consecuencias de

la contaminación a gran escala. A la luz de estos hechos, este trabajo examina dos aspectos clave de este fenómeno: ¿Cuál es el contenido producido por los usuarios sobre noticias relacionadas con la pandemia? y ¿Cómo los usuarios de internet interactúan con este contenido? Para esto, analizamos cinco portales periodísticos utilizando métodos de procesamiento del lenguaje natural y ciencia de datos, como el análisis de sentimientos, el modelado de temas, el análisis de legibilidad y temporalidad. Los resultados alcanzados nos permiten presentar un panorama de la recepción de noticias y desinformación en los principales portales de noticias brasileños.

Palabras clave: Desinformación. Redes Sociales. Covid-19. Periodismo On-line. Brasil.

INTRODUÇÃO

Desde o final de 2019, o mundo enfrenta a primeira grande pandemia do século XXI, a do novo coronavírus. A ciência corre para oferecer rapidamente respostas que possam ajudar a prevenir, a tratar e a curar a infecção por SARS-CoV-2 e as variantes do vírus. Contudo, além da crise sanitária provocada pelo vírus, um outro fator se configura como forte inimigo a ser enfrentado - cujos efeitos ainda estão a ser dimensionados - o crescimento do fenômeno da desinformação, ou *fake news*, como ficou conhecido pelo senso comum. No Brasil, o cenário pandêmico se agravou ainda mais pelo tensionamento entre os campos da Ciência e da Política, que se esgarçam, sobretudo, no campo midiático.

Esse contexto adorna de forma importante o debate sobre a pandemia, principalmente pela atuação direta do presidente da República, Jair Bolsonaro, cujo viés negacionista tem confrontado abertamente as orientações da ciência quanto às medidas de prevenção e tratamento da infecção causada pela Covid-19. Desde antes da posse, o então candidato prometia um intenso combate à corrupção, implementação de uma agenda de costumes conservadora e perseguição implacável aos adversários. Eleito, Bolsonaro tem levado adiante como prioridade o cumprimento de promessas feitas aos seus eleitores, mantendo uma postura de palanque e arrebanhando ativamente muitos apoiadores que travam um embate de ideias quase sempre pautadas em desinformação. Com a pandemia de Covid-19, esse

fenômeno se tornou uma ferramenta poderosa para quem usa a desqualificação do conhecimento científico como estratégia política. Assim, cria-se um terreno fértil para o surgimento de tratamentos preventivos com drogas sem comprovação científica, minimização dos efeitos da doença e negligência em adotar medidas de isolamento social, só para citar alguns dos agravantes de ordem ideológica que se juntam à emergência sanitária

Nesse sentido, Guimarães e Carvalho (2020) alertam para como o conjunto de crenças do indivíduo podem influenciar na forma os indivíduos absorvem as informações e as compartilham, sobretudo:

A situação se torna ainda mais delicada quando trata-se de fármacos que permeiam a questão do tratamento e da profilaxia à doença. Algumas drogas tornaram-se alvos de discussões e manchetes na mídia, principalmente a Hidroxicloquina, Cloroquina, Ivermectina e a Azitromicina. A Hidroxicloriquina e a Cloroquina são utilizadas como antimaláricos, como tratamento para doenças reumáticas e para o lúpus, mas começaram a ser usadas - em associação com a Azitromicina - para tratar e prevenir contra a COVID-19. (GUIMARÃES E CARVALHO, 2020, P.2)

Se por um lado a expansão do ambiente comunicacional da internet possibilita a produção, a aquisição e o compartilhamento de informações, e é um elemento valioso na guerra contra a proliferação de notícias falsas, sobretudo o trabalho do jornalismo profissional; por outro, é nesse território das redes telemáticas que as “desordens da informação” crescem e se multiplicam de forma mais visível. Estudos como os de Naeem, Bhatti e Khan (2020), Recuero e Soares (2020) se detêm em analisar como a desinformação ligada à Covid-19 é disseminada no ambiente virtual e colocam em risco a saúde mundial. Boa parte dessas mensagens falsas são publicadas por meio das mídias sociais e aplicativos de mensagens instantâneas. Mas não só, a desinformação também se faz presente na interação dos usuários nas áreas destinadas a comentários das notícias. Nesse contexto, é proposto no presente trabalho analisar como os usuários de portais de notícias interagem às informações

sobre a pandemia de Covid-19, publicadas ao longo de 15 meses desde janeiro/2020 até março/2021. A intenção é compreender como os internautas acessam essas notícias publicadas na *internet* a partir de recortes por horário de publicação ou assuntos/temáticas, por exemplo. Como já mencionado anteriormente, esse esforço de compreensão parte do engajamento entre público e notícias, com foco nos comentários publicados no espaço destinado para tal fim pelos portais investigados.

Para fins conceituais, adota-se neste trabalho o termo “desinformação” e “desordem da informação”, sugeridos pela Unesco (2019), ao invés da expressão *fake news*, de uso comum para designar o fenômeno de notícias falsas.

A metodologia adotada neste artigo baseia-se em Processamento de Linguagem Natural (PLN), direcionada aos internautas e a forma de ação nas mídias sociais, similar ao proposto por Aggarwal, Rabinovich e Stevenson (2020), centrada nos comentários dos usuários. Esses portais foram escolhidos como objeto de investigação, pois destacam-se entre os maiores no quesito veiculação de notícias generalistas, com atualização contínua, direcionados a um público heterogêneo, quatro dos quais compõem grandes conglomerados de mídia: Folha de S. Paulo, do Grupo Folha; G1, do Grupo Globo; CNN Brasil, da Novus Media, ligado a CNN internacional; Portal R7, do Grupo Record. O portal Metrôpoles não está ligado a nenhum grupo midiático do país.

Um outro fator de escolha do referido objeto, advém do fato que esse cenário, com farta produção de desinformação e manifestações de hostilidade, é também um espaço para o que jornalismo atue como contraponto na disputa de narrativas que se tencionam no campo midiático destinado à manifestação do usuário/leitor. Nesse ambiente, percebe-se um baixo nível de moderação ou com baixa eficiência na filtragem de comentários, favorecendo uma maior disseminação de mensagens de aspectos negativos ou indesejáveis, gerando sentimentos similares aos gerados por notícias, porém com maior potencial de extensão em relação às características da discussão entre usuários *online* (LIU et al., 2020).

PANDEMIA E JORNALISMO

A circulação de notícias acerca da Covid-19 aos usuários de internet ainda no início da pandemia era caracterizada por um baixo nível de informações confiáveis sobre a doença, o que ocasionou um efeito psicológico negativo para a população, conforme Liu et al. (2020). Sentimentos de raiva, medo ou tristeza foram visualizados no público em geral. Também houve um aumento na temática de saúde mental, demonstrando um maior aumento dos casos de problemas mentais no período. Apesar disso, as mídias de caráter jornalístico serviram como um importante canal informativo a respeito de novos casos de infecção e formas de prevenção. Nesse sentido, ao analisar a credibilidade e a confiabilidade das mídias sociais em um cenário eleitoral, Gandia et al. (2019) constataram que as redes possuem um nível de credibilidade inferior aos dos noticiários de TV aberta e a cabo.

Os portais de notícias, em geral, se integram a esse esforço de combater a proliferação da desinformação sobre o Covid-19, produzindo notícias, reportagens, artigos e outros gêneros jornalísticos. Alguns dos principais portais de notícia permitem que os usuários postem mensagens diretamente em sua plataforma, incluindo também funcionalidades de interação com os comentários por parte de outros usuários (*e.g. Curtir e Denunciar*) – posicionando-se, assim, como mídias sociais (RODRIGUES et al., 2021). Esta característica midiática abre um leque para se estudar a percepção do conteúdo, bem como analisar as reações dos usuários frente a determinadas opiniões expostas. Apesar de implementarem filtros de conteúdo para evitar a proliferação de discursos de ódio por meio de listas de palavras-proibidas, que impedem a publicação do comentário, métodos mais sofisticados não são observados. Consequentemente, há uma disseminação de desinformação e de discursos de ódio em tais mídias. Como impactos direto do trabalho, têm-se a mensuração e discussão do panorama da recepção das notícias, distorção de conteúdo e desinformação nos principais portais de notícias brasileiros.

O restante do artigo encontra-se organizado como segue. Os materiais e métodos são descritos na Seção dois, perpassando a coleta, a preparação dos dados e os métodos de análise e validação. Os resultados são apresentados e discutidos na Seção três. Por fim, as conclusões, impactos e direcionamento de trabalhos futuros são apresentados na Seção quatro.

MATERIAIS E MÉTODOS

Nesta seção, apresenta-se a metodologia utilizada para o desenvolvimento do presente estudo. O modelo conceitual do processo, que cobre desde a coleta dos dados até a validação dos resultados, baseando-se nos trabalhos de Silva et al. (2017), de Almeida, Cirqueira e Lobato (2017) e Cirqueira et al. (2016), é apresentado na Figura 1. As etapas encontram-se descritas nas subseções a seguir.

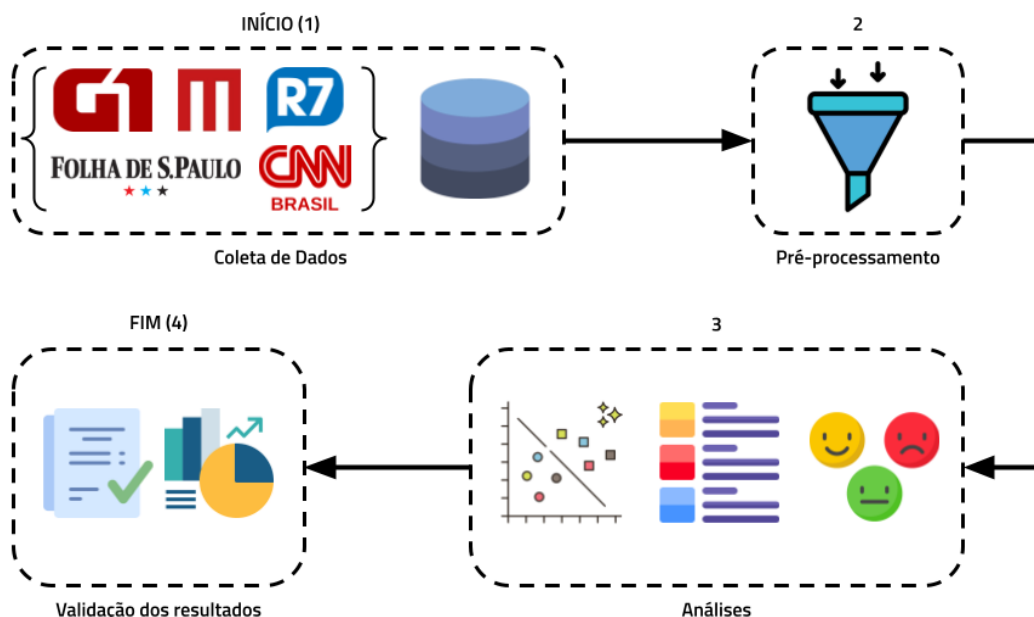


Figura 1 - Estrutura de funcionamento do *framework* experimental.

COLETA DE DADOS

A etapa para a coleta e a construção da base de dados é subdividida em duas principais partes: a seleção de palavras-chave e a coleta efetiva de dados. Para a primeira parte, utilizou-se a metodologia Delphi baseada em Baines et al. (2018). Nela, dois especialistas (um jornalista e um analista de mídias sociais) listaram separadamente as palavras-chave candidatas e depois foram validadas por meio de rodadas de discussões até se chegar em um consenso. Após isso, um terceiro analista listou sinônimos com base em buscas nas redes a fim de se garantir uma cobertura confiável. As palavras-chave consideradas tangentes ao tema alvo e efetivaram um cruzamento entre elas, definindo termos únicos. As validações foram aplicadas por várias vezes perpassando por 54 itens ao total, resultando em uma lista contendo 12 palavras-chave únicas e cinco sinônimos, totalizando 17 termos validados a serem utilizados na coleta. O resultado deste processo encontra-se disposto no Quadro 1.

Termo de Referência	Sinônimos
Coronavírus	N/A
SARS-CoV-2	N/A
COVID-19	COVID; COVID 19
OMS	Organização Mundial da Saúde
<i>Lockdown</i>	<i>Lock-down; Lock down</i>
Pandemia	N/A
Quarentena	N/A
Isolamento Social	N/A

Cloroquina	N/A
Hidroxicloroquina	N/A
Ministério da Saúde	N/A
Distanciamento Social	N/A

Quadro 1 - Palavras-chave definidas após validação.

Para a etapa de coleta, desenvolveu-se um *web crawler* para efetuar a aquisição dos dados desejados com base no Quadro 1. Os locais restringiram-se a cinco portais de notícias selecionados a partir de sua popularidade de acesso no ranqueamento Alexa¹ (PRANTL, D. e PRANTL, M., 2018). Como mencionado acima, os sítios escolhidos foram o Portal G1², Metrôpoles³, Folha de São Paulo⁴, Portal R7⁵ e CNN Brasil⁶. Os dados coletados de forma comum nas plataformas e de acordo com cada fonte estão sumarizados nos Quadros 2 e 3.

Notícias	
Item	Tipo de dado
Título	Textual
Resumo	Textual
Corpo	Textual

¹ <https://www.alexa.com/topsites/countries/BR>

² <https://g1.globo.com/>

³ <https://www.metropoles.com/>

⁴ <https://www.folha.uol.com.br/>

⁵ <https://noticias.r7.com/>

⁶ <https://www.cnnbrasil.com.br/>

Data de Publicação	Data/Hora
<i>Link</i>	<i>URL</i>
Plataforma	Textual
Palavra-chave	Textual

Quadro 2 - Dados coletados das notícias.

Comentários	
Item	Tipo de dado
Autor	Textual
Data de Publicação	Data/Hora
Comentário	Textual
Título da Notícia	Textual
Palavra-chave	Textual
Plataforma	Textual

Quadro 3 - Dados coletados dos comentários

Os dados são armazenados em um banco de dados ao final para preparo na etapa subsequente, o pré-processamento.

PRÉ-PROCESSAMENTO

Esta etapa é fundamental para a remoção de inconsistências e aumento da confiabilidade do conjunto de dados a ser utilizado (RAMÍREZ-GALLEGO et al., 2017). Aplicaram-se métodos adaptados de Cirqueira et al. (2016) e Cirqueira et al. (2018), o qual define duas principais categorias para a execução das tarefas (M), a transformação de um conteúdo textual (TR) e a remoção de conteúdo ou itens (RM). A partir da frase-exemplo “*A vacinação em 2021 irá progredir!*”, a aplicação das técnicas de caixa baixa (TR), acentuação (RM), remoção de caracteres especiais, numéricos e *stopwords* (RM) e também de espaços adicionais (RM) geram ao final como saída a frase “*vacinacao ira progredir*” (sic).

ANÁLISES

Após a etapa de pré-processamento, os métodos de análise são aplicados no conjunto final. A visualização geral dos dados compreende a parcela mais simples em comparação às demais, porém capaz de entregar informações que não são facilmente detectadas (ALI et al., 2016). Com isso, é possível identificar a quantidade de notícias, comentários, interações, médias, modos de distribuição do conteúdo adquirido, dentre outros. Dentro da área de PLN, a modelagem de tópicos permite que informações textuais sejam verificadas e organizadas como uma composição tópico-palavras. Para definir os parâmetros do número de tópicos (T) e palavras (P), considerou-se a facilidade para rotulação e assimilação de um assunto para um conjunto de palavras resultantes do tópico. Também buscou-se evitar combinações de T x P que gerassem ambiguidade para compreensão dos resultados. Ao final, foi definido como T (5) x P (5), evitando uma quantidade elevada de T x P, o que poderia causar repetições ou incongruência entre os tópicos.

Outra análise realizada é relacionada a identificação de sentimentos, a qual permite visualizar um dado conteúdo textual de forma sumarizada em três principais polaridades, negativo, neutro e positivo (SAIF et al., 2016). Com isso, utilizou-se para essa tarefa a

ferramenta *Polyglot* (CHEN & SKIENA, 2014), a qual utiliza um dicionário léxico para determinar qual o principal sentimento presente em sua entrada, tendo uma saída numérica que varia entre -1 e 1. Os valores negativos são considerados como sentimento negativo; Para zero atribui-se como neutro; e um para os positivos - sentimento positivo.

VALIDAÇÃO

A validação dos dados é feita de forma conjunta entre os pesquisadores de forma qualitativa. Nela, selecionam-se alguns comentários de forma aleatória e avalia-se o resultado da análise para cada comentário selecionado – aderente à estratégia conhecida por *ground truth*. A finalidade dessa etapa é a detecção de itens inconsistentes, incoerências ou informações que possam estar incorretas, o que pode demandar um novo preparo e execução das fases anteriores (*e.g.*, pré-processamento ou modelagem).

RESULTADOS E DISCUSSÕES

Na etapa de coleta de dados, o período de publicação das notícias e os comentários coletados abrangem um total de um ano e três meses - entre janeiro de 2020 a março de 2021. Os metadados extraídos do conjunto de dados formado são listados para cada notícia e comentários nos Quadros 4 e 5, respectivamente.

Notícias	
Metadado	Valor
Quantidade coletada	14.314
Média de publicação mensal	954,27
Plataformas acessadas	5

Quadro 4 - Metadados obtidos para as notícias na base de dados coletada.

Comentários	
Metadado	Valor
Quantidade coletada	58.434
Média de publicação mensal	3.896
Média de publicação por notícia	4,1
Plataformas acessadas	3

Quadro 5 - Metadados obtidos para os comentários na base de dados coletada.

Com base nos metadados dos Quadros 4 e 5, foram coletadas um total de 14.314 notícias, uma média de 954 notícias publicadas mensalmente. Quanto aos comentários, foram coletadas 58.434 postagens, com uma média de 3.896 ao mês e 4,1 por notícia. Nesse caso, os comentários foram limitados a análises apenas nas plataformas que disponibilizam tal seção, as quais são restritas ao Portal G1, Metrôpoles e Folha de São Paulo.

O acesso a essas notícias e a consequente interação por parte dos internautas são fatores que podem variar de acordo com o assunto principal abordado. Um dos principais itens que indexa a notícia aos sistemas de busca como o *Google* são as suas palavras-chave. A visualização na quantidade das notícias publicadas por palavra-chave é apresentada na Figura 2.

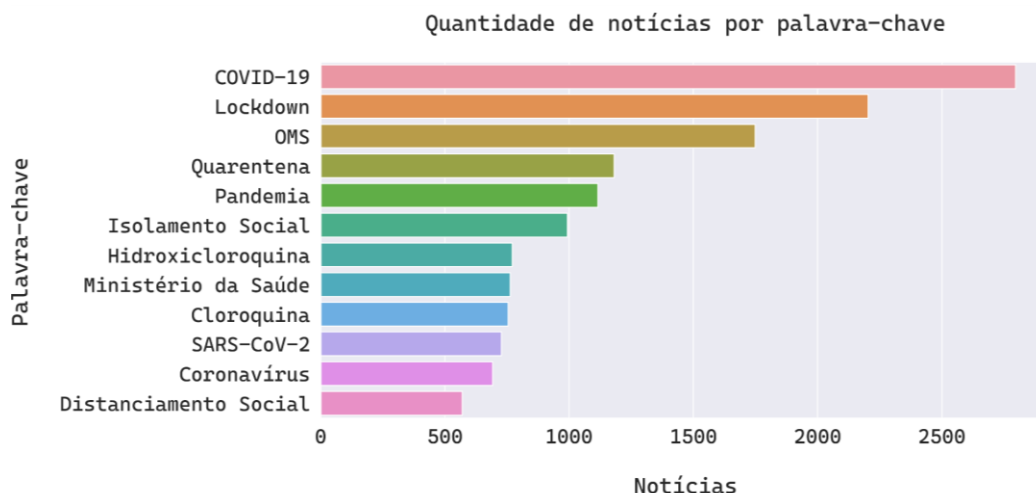


Figura 2 - Quantidade de notícias por palavra-chave.

Além das palavras-chave, o volume de publicação de notícias oscilou nas plataformas entre os meses. Em contrapartida, não se pode afirmar que uma maior quantidade de postagens cause um engajamento superior, pois é dependente de um sistema de comentários, nível de crescimento e compartilhamento de uma dada notícia e da popularidade da plataforma. A Figura 3 exibe o volume de notícias publicadas mensalmente por plataforma no período estudado.

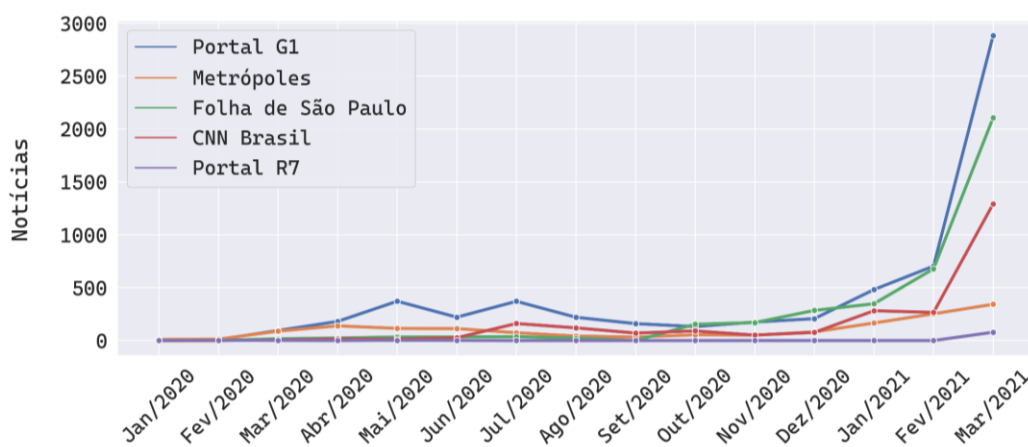


Figura 3 - Volume mensal de notícias por plataforma.

Os meses com as maiores taxas de postagens em 2020 são julho e dezembro, com 1.294 notícias publicadas, sendo esses os mais representativos para as cinco plataformas sob análise. Quando visualizado o ano de 2021, o destaque vai para o mês de março com 6.705 notícias. Esses dados representam os cenários da pandemia no país, tal como o crescimento contínuo da quantidade de casos e do número de mortos pela Covid-19. Em sequência, houve a chegada das vacinas e problemas na aquisição das mesmas, bem como escândalos envolvendo a interlocução com diversos fabricantes, que teve como consequência a instauração da “CPI da Covid”.

Quando direcionado ao volume por hora, chama-se atenção para o período da tarde, horário com maior movimentação de notícias relacionadas à política e boletins de fim do dia, lançados às 23 horas pela Folha de São Paulo. Essa representação é sumarizada na Figura 4.

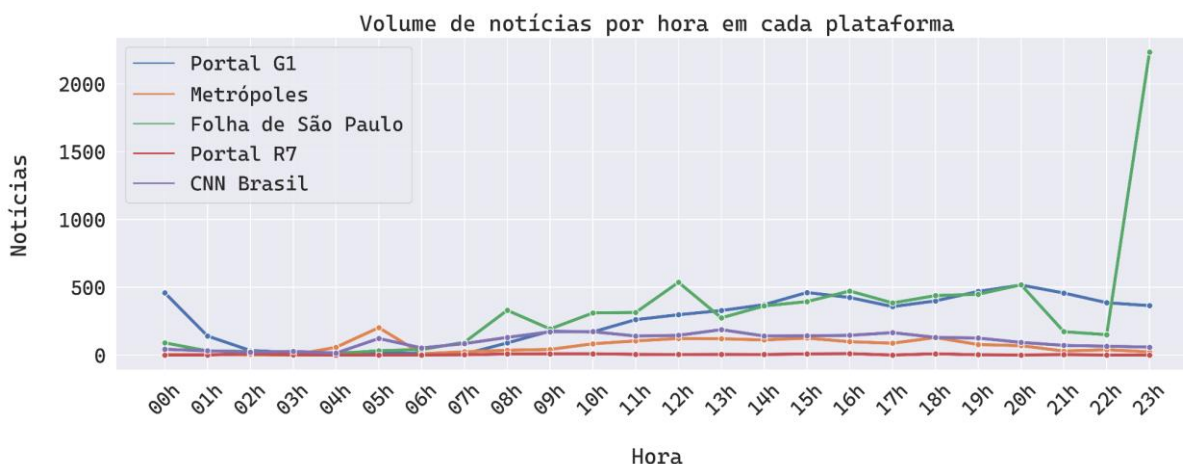


Figura 4 - Volume de notícias por hora em cada plataforma.

Percebe-se que a frequência de publicação de notícias é mais significativa entre 12h e 23h, os horários das 12h (1.104 notícias) e 16h com (1.149 notícias). Em sincronia, o volume de comentários também tende a seguir um padrão similar a depender do horário,

principalmente pela maior presença de pessoas conectadas. Essa representação é apresentada na Figura 5.



Figura 5 - Volume de comentários por hora em cada plataforma.

O G1 assume o destaque quanto à quantidade de interações dentre as três plataformas devido a ter uma maior abrangência de conteúdo, maior fluxo de acesso e um sistema simplificado para a publicação de comentários que utiliza um sistema proprietário, o que não ocorre por exemplo, com o Metrôpoles que faz o uso do *Disqus*. O engajamento de uma notícia pode ser visualizado por meio da somatória entre a quantidade de comentários e as avaliações positivas e negativas totais atribuídas. Tais dados permitem visualizar quais os assuntos com maior repercussão, além de exibir qual o interesse do usuário em interagir e quanto uma publicação repercutiu. O engajamento das cinco principais notícias em relação às suas interações, com suas palavras-chave e contabilizando os comentários e interações positivas e negativas é evidenciado na Figura 6 e no Quadro 6.

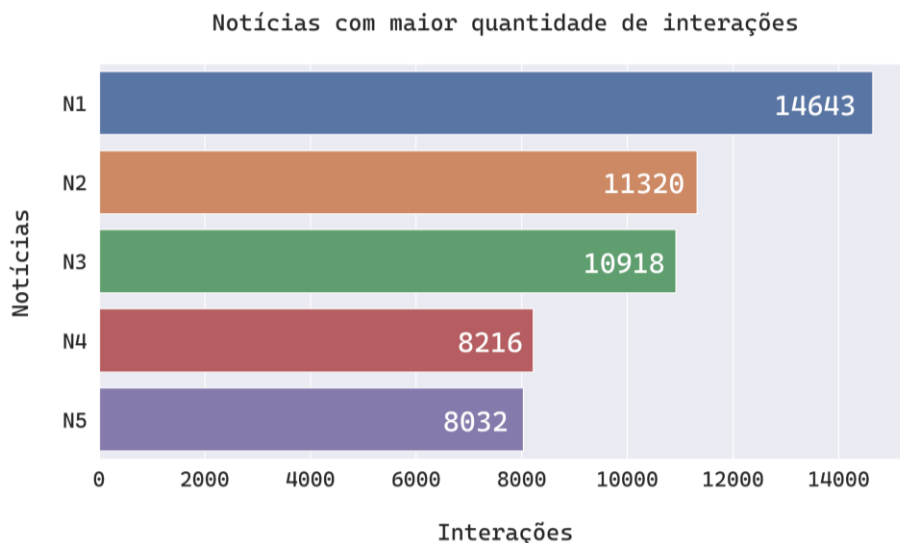


Figura 6 - Cinco notícias com maior engajamento em interações.

ID	Título	Palavras-chave	Interações			Total
			Comentários	Positivas	Negativas	
N1	"Mourão diz que posição do governo 'é uma só': isolamento e distanciamento social"	Distanciamento Social	486	9.018	5.139	14.643
N2	"'Não houve convergência técnica entre nós', diz médica que rejeitou convite de Bolsonaro para assumir Ministério da Saúde"	Cloroquina; Ministério da Saúde	2.150	9.170	0	11.320
N3	"Pazuello mente ao afirmar que ministério nunca recomendou 'tratamento precoce' para Covid"	Cloroquina; Hidroxicloroquina	1.284	9.634	0	10.918
N4	"Venda de remédios sem eficácia comprovada contra a Covid dispara"	Cloroquina; Hidroxicloroquina	2.068	6.148	0	8.216
N5	"Ex-integrantes da PGR apontam crime de Bolsonaro na pandemia e pedem a Aras para denunciá-lo"	Cloroquina; Hidroxicloroquina	1.202	6.830	0	8.032

Quadro 6 - Cinco notícias com maior engajamento em interações.

O maior índice de engajamento foi para a notícia N1, com um total de 14.643 interações, divididos em 486 comentários com 9.018 e 5.139 avaliações positivas e negativas, respectivamente. Pode-se inferir que o maior número de reações positivas se deve ao fato de que a notícia reporta a posição de um indivíduo que ocupa o posto máximo na República, quando da substituição do titular da presidência.

Algumas das notícias estão associadas às falas de figuras políticas de relevância, como o presidente do país, repercutidas a partir de comentários em entrevistas ou redes sociais. Quando se trata de transpor a esfera da política pública para a prescrição de medicamentos, instauram-se situações como, por exemplo, a compra disparada de remédios sem comprovação científica para o tratamento precoce contra a Covid-19, é o que está retratado na Notícia N4. No caso em tela, o uso indiscriminado da Cloroquina ou Hidroxicloroquina além de não gerar efeitos propalados, houve escassez da droga para tratar quem realmente a necessita. A pesquisa científica séria provou que a Cloroquina não serve para prevenir e/ou tratar a Covid-19.

Uma das características presentes em mídias sociais é quanto ao tamanho dos comentários publicados, comumente mais curtos e informais. Postagens curtas com críticas, mensagens xenofóbicas ou com características de discurso de ódio são algumas das variações presentes por todo o conjunto de dados. Respectivamente, citam-se os exemplos: “O Brasileiro não tem consciência, não adianta.”; “estão de parabens seus campineiros burrosssss” (sic); e “E quando o vírus contamina um idiota, ele se torna idiota ao quadrado?”. Os comentários citados servem para demonstrar que uma publicação curta pode ser um meio simples, porém, rápido e efetivo de disseminação para discursos de ódio em geral.

Em visa de mostrar quais os tópicos, assuntos e palavras mais abordados na interação entre os usuários, além de possibilitar sua comparação com as palavras-chave da Figura 2 em relação a temática principal, listam-se os tópicos e termos mais utilizados nos comentários por meio da modelagem de tópicos no Quadro 7.

Tópico	Assunto	Palavras
1	Saúde no Governo Bolsonaro	governo vacina sao bolsonaro presidente
2	Produção de vacinas	brasil producao país vez eua
3	Tratamento com Cloroquina	covid pessoas tratamento cloroquina precoce
4	Vacina / Coronavac	bula vacina vc coronavac toma
5	Política	ser povo deve político presidente

Quadro 7 - Tópicos e palavras mais frequentes para os comentários.

Dentre os temas mais abordados, encontram-se ligações direcionadas aos recentes escândalos na saúde brasileira, assim como o vínculo do tópico aos tratamentos não comprovados e à vacinação. Palavras assimiladas como “bolsonaro”, “producao”, “cloroquina” e “coronavac”, relacionadas aos Tópicos 1, 2, 3 e 4 fazem essa conexão entre quatro dos cinco tópicos principais abordados. Além disso, o Tópico 5 também se vincula ao conteúdo geral por tratar-se da temática política e dessa forma, tangencia os demais assuntos discutidos. Listam-se no Quadro 8 alguns comentários conectados aos tópicos do Quadro 7.

N	Comentário
1	“idiota esse já era, só de seguir o Bozo já prova isso não precisa da covid-19. Esse já era idiota com a Gado-17.”
2	“Se trata com cloroquina e hidroxicloroquina, aproveita e chama o Covard 17 pata te visitar, sem máscara, se abracem, tirem selfies, e conversem sobre a gripezinha inventada pela China, depois chamem os africanos para conduzirem os caixões, teu e do porco, digo Bozo.”
3	“A chinezada tá ferrada. Terão que vender muito ching ling pra recuperar o prejuí . E esses números aí divulgados pelo partido comunista são duvidosos , é claro que são.”

Quadro 8 - Comentários extraídos diretamente do conjunto de dados.

Em referência ao Quadro 8, identificam-se traços visíveis de ironia e de discursos de ódio, com ligação análoga ao apresentado nos tópicos do Quadro 7. A forma na qual as diversas temáticas são abordadas nos exemplos mostram a insatisfação e negatividade de uma parte dos internautas sobre a forma na qual a pandemia é gerenciada. Além disso, percebe-se ataques de cunho político, xenofóbico e outras manifestações pejorativas.

Além dos principais tópicos abordados, a forma a qual é feita a comunicação entre os usuários pode apresentar indícios de ofensas, seja pelo teor de uma notícia quanto ao seu tema principal, ou em referência a outro comentário publicado por outra pessoa e que desencadeia uma série de reações. Para isso, a análise de sentimentos foi feita visando identificar qual o teor dessas postagens e é apresentada na Figura 7. Butler (2021) alerta para o fato de que a linguagem atua com o poder de ferir as pessoas justamente porque os seres humanos são seres linguísticos (BUTLER, 2021, p. 12). Não é objetivo deste trabalho discutir a questão do discurso de ódio, mas apenas referir que o espaço de comentários de notícias configura-se simbolicamente como uma ferramenta para ferir o outro, um dado expresso pela análise de sentimentos e os trechos selecionados no Quadro 8.

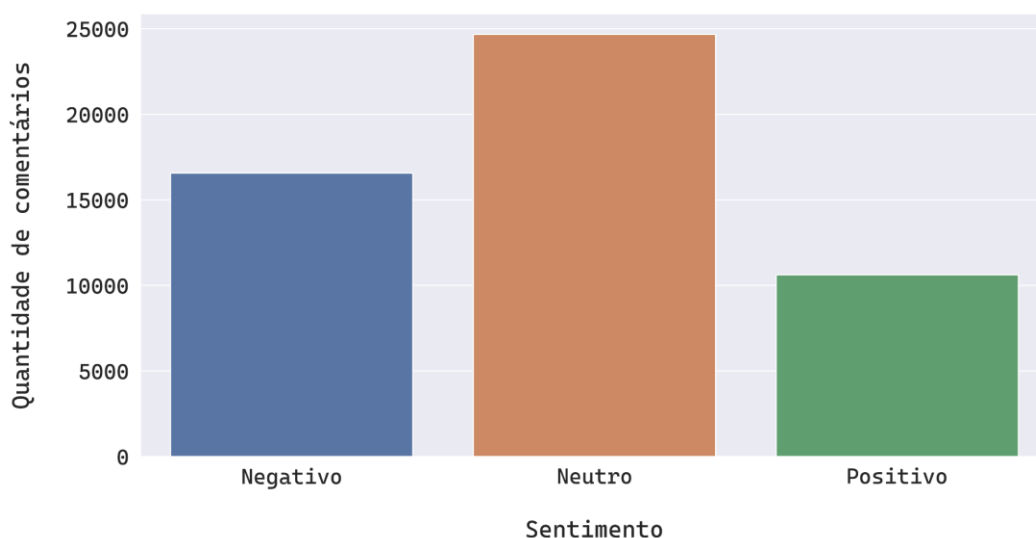


Figura 7 - Análise de sentimentos para os comentários.

Dentre os comentários, a análise de sentimentos resultou em 24.668 neutros (47,57%), 16.565 negativos (31,94%) e 10.628 positivos (20,49%). A diferença entre as polaridades negativa e neutra é de 15,63% (8.103 comentários). A diferença percentual entre os comentários positivos e negativos é de 11,45%, mostrando que há a polarização entre os dois extremos dos sentimentos nos comentários publicados nas plataformas. Esse cenário é potencializado pelas notícias acerca da pandemia, pelos usuários enfrentando um novo estilo de vida repentino ocasionado pela doença, afetando-os de alguma forma, e também por notícias e informações falsas disseminadas nas mídias sociais - que intensificam esses efeitos de forma compartilhada para os que as visualizam e aceitam como fato imediato - além de provocar reações mais exaltadas por parte do público. Os comentários N1 e N2 do Quadro 8 representam parte dos textos publicados dentro dessas características e que envolvem outras, como o extremismo político e a xenofobia. Além disso, a presença de discursos inflamados publicados nas notícias gera um efeito em cadeia, onde um dado comentário ocasiona em uma série de respostas por parte de outros internautas.

CONSIDERAÇÕES FINAIS

Quando se analisa o volume de dados, é possível identificar uma série de elementos que permitem extrapolar para além da quantificação. Há, sem dúvida, um número superior de notícias publicadas durante todo o dia, assim como o de comentários, que seguem a lógica das rotinas produtivas nas redações jornalísticas, quase sempre com volume de apuração, edição e publicação de notícias em grande parte durante manhã, tarde, e começo da noite, o que fica claro quando se olha a taxa de postagem, mantida em uma curva de crescimento a partir do início da manhã, às 7h, e decaindo ao final do dia, entre 20h e 23h.

Outro dado interessante, presente nas mídias sociais, é a baixa quantidade de palavras utilizadas nos textos, além do tom informal com os quais são redigidos. Esse formato é de

fácil produção e bastante efetivo para propagar ofensas, facilmente disseminadas em função da velocidade de propagação da *internet*. Identifica-se objetivamente o uso político dessas manifestações curtas e informais. Assim, comentários com traços xenofóbicos, termos pejorativos e irônicos são direcionados não só a figuras públicas citadas pelas notícias, mas também a outros usuários que interagem com determinada notícia. Isso fica mais visível quando se olha a análise de sentimentos, com grande presença do negativo e baixa existência do positivo na maioria dos comentários, o que reforça a ideia de que um ambiente das interações é muito hostil, quiçá uma "esfera pública" distorcida, permeada não discussões proíficas, mas por linguagem injuriosa, expiação de culpas alheias e canal de vazão dos ressentimentos pessoais e coletivos.

Identificou-se também que os tópicos mais abordados estão relacionados ao Covid-19 na China e no Brasil, ao Presidente da República do Brasil e a escândalos do governo ocorridos durante o período de análise, ampliando os recortes para além da pandemia propriamente dita. Por exemplo, foram as discussões geradas acerca do *impeachment* de Jair Bolsonaro. Os portais de notícias tiveram um papel importante na disseminação de informação sobre esse tema. O Portal G1, por exemplo, propõe a *hashtag* “#FATO ou #FAKE” que serve para confirmar se uma dada notícia, informação, postagem ou conteúdo compartilhados pelas mídias sociais partem do princípio da verdade dos fatos. Outra iniciativa de empresas jornalísticas diz respeito à atualização diária de como a Covid-19 evoluiu pelo país e, com o detalhamento por unidade da federação, ações e formas de prevenção contra a doença. Percebe-se que esse tipo de informação de serviço foi veiculada de modo constante por essas plataformas, destacando o seu papel informativo e de combate ao vírus.

A pandemia no Brasil trouxe diversos desafios e gerou um ambiente de ansiedade e medo no país. Percebe-se um grande volume de notícias sobre o tema e, com as medidas de isolamento, com parte considerável do público permanecendo mais tempo em casa, a

quantidade de interações também aumentou. Esse cenário em combinação com o aspecto de pseudo-anonimato proporciona uma maior proliferação de discursos de ódio, além de uma maior hostilidade nas discussões on-line. Além disso, é perceptível a politização da crise sanitária. Somada às trocas de ministros da Saúde em 2020 e a postura ambivalente do governo federal quanto às medidas efetivas para conter a pandemia, ainda está por se mensurar as reais consequências sociais e econômicas do impacto da pandemia no país. A "CPI da Covid-19", do Senado Federal, representou um esforço de apurar e responsabilizar os agentes públicos que não agiram adequadamente diante da emergência sanitária. O jornalismo teve papel importante no acompanhamento do andamento da Comissão Parlamentar de Inquérito, num movimento que envolveu ativamente os debates paralelos nas mídias sociais.

A interação do usuário com as notícias pode significar um fator importante para sacudir certezas e dogmas, principalmente, quando se pensa a função de serviço público que o jornalismo tem prestado no cenário de uma pandemia dessas proporções. São levantadas algumas possibilidades para trabalhos futuros, como uma melhor exploração na forma na qual os usuários interagem a depender da temática da notícia ou por ter envolvimento de uma figura pública ou reconhecida, e, por fim, verificar também a efetividade dos sistemas de denúncia e filtragem de comentários pejorativos nas plataformas a partir de um cruzamento com detectores de discurso de ódio mais sofisticados, bem como o papel das agências de fact-checking. Contudo, do ponto de vista dos fenômenos de desordem da informação, é possível afirmar que *“a consequência de tudo isso é que a desinformação, com esse combustível digital, em contextos de polarização, arrisca ofuscar o papel do jornalismo”* (Unesco, 2019, p.19).

AGRADECIMENTOS

Este trabalho foi parcialmente financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - DT-308334/2020 e 154551/2021-0; pela Fundação Amazônia de Amparo a Estudos e Pesquisas (FAPESPA) - PRONEM-FAPESPA/CNPq nº 045/2021.

REFERÊNCIAS

AGGARWAL, Jai; RABINOVICH, Ella; STEVENSON, Suzanne. Exploration of gender differences in COVID-19 discourse on reddit. **arXiv preprint arXiv:2008.05713**, 2020.

ALI, Syed Mohd et al. Big data visualization: Tools and challenges. In: **2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)**. IEEE, 2016. p. 656-660.

BAINES, Rebecca L.; REGAN DE BERE, Sam. Optimizing patient and public involvement (PPI): Identifying its “essential” and “desirable” principles using a systematic review and modified Delphi methodology. **Health Expectations**, v. 21, n. 1, p. 327-335, 2018.

BUTLER, Judith. **Discurso de Ódio**. Uma política do performativo. São Paulo: Editora Unesp. 2021.

CARVALHO, Wellington; GUIMARÃES, Ádria Silva. Desinformação, Negacionismo e Automedicação: a relação da população com as drogas “milagrosas” em meio à pandemia da COVID-19. **InterAmerican Journal of Medicine and Health**, v. 3, 2020.

CHEN, Yanqing; SKIENA, Steven. Building sentiment lexicons for all major languages. In: **Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**. 2014. p. 383-389.

CIRQUEIRA, Douglas et al. A literature review in preprocessing for sentiment analysis for Brazilian Portuguese social media. In: **2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)**. IEEE, 2018. p. 746-749.

CIRQUEIRA, Douglas et al. Performance evaluation of sentiment analysis methods for Brazilian Portuguese. In: **International Conference on Business Information Systems**. Springer, Cham, 2016. p. 245-251.

DE ALMEIDA, Gustavo RT; CIRQUEIRA, Douglas Rocha; LOBATO, Fábio MF. Improving Social CRM through electronic word-of-mouth: a case study of ReclameAqui. In: **Anais Estendidos do XXIII Simpósio Brasileiro de Sistemas Multimídia e Web**. SBC, 2017. p. 107-110. ISSN 2596-1683.

GANDIA, Rodrigo Marçal et al. Credibilidade e Confiabilidade das Redes Sociais nas Eleições. **Revista Estudos Em Comunicação**, v. 28, p. 201–218, 2015.

LI, Shaohua et al. Generative topic embedding: a continuous representation of documents. In: **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**. 2016. p. 666-675.

LIU, Qian et al. Health communication through news media during the early stage of the COVID-19 outbreak in China: digital topic modeling approach. **Journal of medical Internet research**, v. 22, n. 4, p. e19118, 2020.

NAEEM, Salman Bin; BHATTI, Rubina; KHAN, Aqsa. An exploration of how fake news is taking over social media and putting public health at risk. **Health Information & Libraries Journal**, v. 38, n. 2, p. 143-149, 2021.

PRANTL, David; PRANTL, Martin. Website traffic measurement and rankings: competitive intelligence tools examination. **International Journal of Web Information Systems**, 2018.

RAMÍREZ-GALLEGO, Sergio et al. A survey on data preprocessing for data stream mining: Current status and future directions. **Neurocomputing**, v. 239, p. 39-57, 2017.

RECUERO, Raquel; SOARES, Felipe. O Discurso Desinformativo sobre a Cura do COVID-19 no Twitter: Estudo de caso. **E-Compós**. 2021.

RODRIGUES, Lucas; JUNIOR, Ademir; LOBATO, Fabio. Notícias relacionadas a pessoas com deficiência: uma análise do conteúdo gerado pelos usuários em postagens de mídias sociais. In: **Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional**. SBC, 2019. p. 811-822.

RODRIGUES, Lucas; JUNIOR, Antonio Jacob; LOBATO, Fábio. Análise de discursos em notícias sobre homofobia, racismo e sexismo em comentários de portais brasileiros de notícias. **Anais do Computer on the Beach**, v. 12, p. 467-474, 2021.

SAIF, Hassan et al. Contextual semantics for sentiment analysis of Twitter. **Information Processing & Management**, v. 52, n. 1, p. 5-19, 2016.

SILVA, Wendel et al. A methodology for community detection in Twitter. In: **Proceedings of the International Conference on Web Intelligence**. 2017. p. 1006-1009.

UNESCO, 2019. **Jornalismo, Fake News & Desinformação**. Manual para Educação e Treinamento em Jornalismo. p. 07-31. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000368647?fbclid=IwAR1ltj8iF00MPv69hOx4WViYAHzMUlp8VoYIT0Mepi_TYL_utbV5xIgnnEk. Acesso em: 02 de agosto de 2021.

COMO CITAR ESTE ARTIGO

RODRIGUES, Lucas Darlindo Freitas; PRADO, Ana Lúcia; LOBATO; Fábio Manoel França. Pandemia de Covid-19 no Brasil: uma análise sobre notícias e comentários de usuários. **Revista Culturas Midiáticas**, João Pessoa, v. 16, pp. 41-66, 2022. DOI: <https://doi.org/10.22478/ufpb.2763-9398.2022v16n.61265>.