

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

ANDRÉ MEDEIROS MENDES

**ANÁLISE DA DESINFORMAÇÃO NO CONTEXTO DA VACINAÇÃO CONTRA O
CORONAVÍRUS NO BRASIL**

CURITIBA

2023

ANDRÉ MEDEIROS MENDES

**ANÁLISE DA DESINFORMAÇÃO NO CONTEXTO DA VACINAÇÃO CONTRA O
CORONAVÍRUS NO BRASIL**

**Analysis of misinformation in the context of the coronavirus vaccination in
Brazil**

Dissertação apresentada como requisito para obtenção do título de Mestrado em Ciência da Computação do Programa de Pós-Graduação em Computação Aplicada da Universidade Tecnológica Federal do Paraná.

Orientador: Prof. Dr. Luiz Celso Gomes Jr.

CURITIBA

2023



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.



ANDRE MEDEIROS MENDES

ANÁLISE DA DESINFORMAÇÃO NO CONTEXTO DA VACINAÇÃO CONTRA O CORONAVÍRUS NO BRASIL

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre Em Computação Aplicada da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Engenharia De Sistemas Computacionais.

Data de aprovação: 18 de Abril de 2023

Dr. Luiz Celso Gomes Junior, Doutorado - Universidade Tecnológica Federal do Paraná

Dr. Renato Fileto, Doutorado - Universidade Federal de Santa Catarina (Ufsc)

Dr. Roberlei Alves Bertucci, Doutorado - Universidade Tecnológica Federal do Paraná

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 20/04/2023.

AGRADECIMENTOS

Certamente estes parágrafos não irão atender a todas as pessoas que fizeram parte dessa importante fase de minha vida. Portanto, desde já peço desculpas àquelas que não estão presentes entre essas palavras, mas elas podem estar certas que fazem parte do meu pensamento e de minha gratidão.

Agradeço ao meu orientador Prof. Dr. Luiz Celso Gomes Junior, pela sabedoria com que me guiou nesta trajetória.

Gostaria de deixar registrado também, o meu reconhecimento à minha família, pois acredito que sem o apoio deles seria muito difícil vencer esse desafio.

Enfim, a todos os que por algum motivo contribuíram para a realização desta pesquisa, meu muito obrigado.

“Não é o conhecimento, mas a utilidade do conhecimento, que constitui a verdadeira ciência.” - Francis Bacon

RESUMO

A desinformação que circula nas redes sociais a respeito da pandemia do coronavírus tem sido um obstáculo no combate à doença, impactando diretamente no avanço da imunização em diversas partes do mundo. Este estudo buscou avaliar possíveis influências entre as taxas de vacinação no Brasil e as interações nas mídias sociais relacionadas a postagens de fontes questionáveis. Utilizando análises de redes complexas e modelos econométricos, medimos o impacto de informações questionáveis sobre vacinas, considerando questões geográficas, políticas e sociais. Utilizamos a lista de páginas classificadas como emissoras de desinformação pela CPMI Fake News, instalada pelo Senado Federal Brasileiro em setembro de 2019, como fontes questionáveis de informação. Um procedimento de análise de grafos foi implementado para expandir a lista original e identificar outras contas que tenham compartilhado os mesmos conteúdos duvidosos. A partir da lista expandida, obtivemos as postagens consideradas questionáveis no Facebook. Os resultados das análises de dados de painel mostram um impacto negativo nas taxas de vacinação, com diferentes variações dependendo da orientação política das fontes disseminadoras consideradas.

Palavras-chave: desinformação; ciência de dados; vacinação; coronavírus.

ABSTRACT

The misinformation circulating on social media regarding the coronavirus pandemic has been an obstacle in the fight against the disease, directly impacting the advancement of immunization in various parts of the world. This study sought to evaluate possible influences between vaccination rates in Brazil and social media interactions related to posts from questionable sources. Using complex networks analysis and econometric models, we measured the impact of questionable information about vaccines, considering geographic, political and social issues. We used the list of pages classified as emitters of disinformation by CPMI Fake News, installed by the Brazilian Federal Senate in September 2019, as questionable sources of information. A chart analysis procedure was implemented to expand the original list and identify other accounts that shared the same dubious content. From the expanded list, we obtained the positions considered questionable on Facebook. The results of the panel data analyzes show a negative impact on vaccination rates, with different variations depending on the political orientation of the disseminating sources considered.

Keywords: misinformation; data science; vaccination; coronavirus.

LISTA DE FIGURAS

Figura 1 – Framework Infodemic WHO	11
Figura 2 – Metodologia do processamento dos dados	18
Figura 3 – Nuvem de Palavras	19
Figura 4 – Correlação entre a Vacinação e as lista CPMI e lista Geral	21
Figura 5 – Análise temporal Vacinas x Facebook	24
Figura 6 – Grafo representando as conexões entre páginas no Facebook	26
Figura 7 – Comunidades da rede	27
Figura 8 – Possibilidades de novas conexões na rede	27
Figura 9 – Reta de regressão mostrando a influência das interações no Facebook por inclinação política do estado (UF)	29
Figura 10 – Reta de regressão mostrando a influência das interações no Facebook por faixa de IDH	29
Figura 11 – Mapas Coeficientes de Regressão e p-values	32

LISTA DE TABELAS

Tabela 1 – Classificação política das páginas de maior centralidade em cada comunidade	28
Tabela 2 – Fixed Effects	31
Tabela 3 – Resultado da adaptação do modelo para capturar influência estadual .	33

SUMÁRIO

1	INTRODUÇÃO	10
1.1	Objetivos	11
1.2	Justificativa	11
1.3	Estrutura do trabalho	12
2	REFERENCIAL TEÓRICO	13
2.1	As fake news e a sociedade	13
2.2	Impactos da desinformação na vacinação	13
2.3	Regressão Linear	15
2.3.1	Regressão com Dados em Painel	15
2.3.2	Fixed Effects	16
2.4	Ciência das Redes	16
2.4.1	Centralidade	17
2.4.2	Comunidades	17
3	METODOLOGIA E PROCESSAMENTO DE DADOS	18
3.1	Fontes de Dados	18
3.1.1	Postagens do Facebook	18
3.1.2	Fontes de desinformação	20
3.1.3	Dados sobre a vacinação, casos e óbitos do coronavírus	21
3.1.4	Dados sociais e políticos	22
3.2	Modelo de Regressão Linear - Fixed Effects	22
3.3	Análise Exploratória	23
3.3.1	Análise Temporal	23
3.3.2	Análise de rede e identificação da orientação política	25
3.3.3	Questões políticas e socioeconômicas	26
4	RESULTADOS	30
4.1	Regressão Linear	30
4.2	Análise Geográfica	32
4.3	Limitações e discussões futuras	34
5	CONCLUSÃO	36

REFERÊNCIAS	37
------------------------------	-----------

1 INTRODUÇÃO

Várias tecnologias de produção de vacina tiveram seus testes acelerados em decorrência do agravamento da pandemia de COVID-19 no mundo. As vacinas se tornaram a solução mais promissora e ansiosamente esperada. A melhor alternativa para controlar a pandemia é a imunização da população através da vacinação (LIMA; ALMEIDA; KFOURI, 2021).

Tanto a eficácia das novas vacinas quanto o isolamento social têm sido assuntos recorrentes em plataformas digitais e grandes alvos de desinformação e fake news. Essas postagens com conteúdo duvidoso, quando disseminadas por canais de alta capilaridade, têm o potencial de influenciar um grande número de pessoas, causando um impacto direto na imunização da sociedade. Um exemplo desse impacto é a queda na taxa de imunizados com a vacina MMR ¹, que caiu de 99% na geração pré-Internet da população em São José do Rio Preto, São Paulo, Brasil, para 32,9%, após a publicação de um artigo pelo médico Andrew Wakefield em 1998, o qual sugeria uma ligação entre essa vacina e o autismo (VASCONCELLOS-SILVA; CASTIEL, 2020).

De acordo com uma pesquisa realizada pela VotoLatino em 2021 (LATINO, 2021), entre os entrevistados que ainda não tinham sido vacinados, 51% disseram que provavelmente não serão vacinados contra a COVID-19 em momento algum. O estudo constatou que a desinformação em torno da eficácia das vacinas desempenha um papel fundamental na hesitação entre os latinos. A pesquisa também constatou que o Facebook é a principal fonte de desinformação em relação à vacina, representando 49%, seguido por notícias locais, com 39%.

Esse estudo analisa os impactos nos índices de vacinação causados por postagens realizadas por fontes questionáveis, considerando a inclinação política estadual com base no resultado das eleições presidenciais de 2018, o IDH e a renda per capita de cada estado. Para identificar as fontes questionáveis, utilizamos como base uma lista de páginas que foram classificadas como canal de comportamento desinformativo pela Consultoria Legislativa a pedido da Comissão Parlamentar Mista de Inquérito - Fake News (CPMI Fake News) (BRASIL, 2019), instaurada no Senado Federal do Brasil.

Os resultados obtidos demonstram que o aumento nas interações dos usuários com as postagens oriundas de fontes questionáveis resultaram em um impacto negativo nos índices de vacinação em todo o Brasil. Esses resultados apresentaram uma instabilidade que dificultou a interpretação. Uma possível causa é a ausência de variáveis e de dados mais qualificados que ajudem a estabilizar o modelo, como fontes mais confiáveis de desinformação e outras variáveis sócio-econômicas.

¹ 'Vacina tríplice viral contra Sarampo, Caxumba e Rubéola - <https://pt.wikipedia.org/wiki/VASPR>'

1.1 Objetivos

O objetivo principal desse estudo é analisar o possível impacto da desinformação nos índices de vacinação contra coronavírus no Brasil.

Os objetivos específicos são:

- Verificar a possibilidade de medir o impacto de postagens no Facebook contendo conteúdo duvidoso sobre a COVID-19 sobre os índices de vacinação no Brasil
- Verificar se a inclinação política das páginas consideradas como fontes de desinformação é relevante para o possível impacto nos índices de vacinação.
- Verificar se variáveis socio-econômicas e geográficas interferem no possível impacto da desinformação nos índices de vacinação.

1.2 Justificativa

De acordo com a Declaração transregional sobre a infodemia COVID-19, assinada pelos 132 Estados-Membros da Organização Mundial da Saúde, “Em tempos de crise sanitária do COVID-19, a disseminação da “infodemia” pode ser tão perigosa para a saúde e segurança humana quanto a própria pandemia.”²

Em abril de 2020 a Organização Mundial da Saúde emitiu um relatório técnico sobre o gerenciamento da infodemia, onde ela apresenta um framework com 4 áreas de atuação (figura 1). Uma das áreas é denominada "Quantify Impact", onde é destacada a importância de se entender a infodemia para combatê-la.

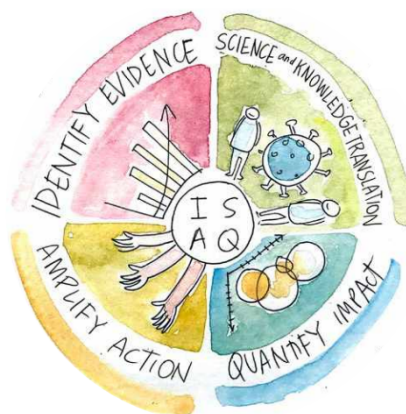


Figura 1 – Framework Infodemic WHO

O Ministério da Saúde informou em seu livro Saúde Brasil 2020-2021 (TRANSMISSÍVEIS, 2022) que no final de 2020 o Brasil apresentava 7.716.405 casos de COVID e 195.725

² https://onu.delegfrance.org/IMG/pdf/cross-regional_statement_on_infodemic_final_with_all_endorsements.pdf

óbitos confirmados da doença, e todos os municípios brasileiros apresentavam, pelo menos um caso confirmado da doença, e 90,4% apresentavam óbitos confirmados. O Brasil apresentava-se com o maior número absoluto de casos e óbitos em relação aos demais países da América Latina e o terceiro maior entre todos os países do mundo, ficando atrás dos Estados Unidos da América e da Índia.

A possibilidade de obtenção dos dados do Facebook através do Crowdtangle e a disponibilidade dos dados sobre a vacinação no Brasil proporcionam insumos para realizarmos as análises para alcançar os objetivos desse trabalho. Além dos dados, temos a oportunidade temporal, onde a proximidade dos fatos ocorridos faz com que o momento seja oportuno para a realização do estudo, possibilitando analisar os impactos nos dados reais de vacinação.

1.3 Estrutura do trabalho

Esse trabalho contextualiza na seção 2 os trabalhos relacionados e fundamentos importantes para a compreensão da pesquisa. A seção 3 descreve a metodologia e processamento de dados. A seção 3.3 apresenta a análise exploratória realizada nesse estudo. Na seção 4 são detalhados os resultados obtidos. Na seção 4.3 são apresentadas as limitações e possíveis trabalhos futuros e na seção 5 é apresentada a conclusão do estudo.

2 REFERENCIAL TEÓRICO

2.1 As fake news e a sociedade

As fake news são um tema altamente relevante no cenário político e social brasileiro. Entre as categorias de fake news mais frequentes se destaca a de natureza política, seguida pelas categorias de epidemiologia e de tratamento e prevenção da COVID-19 (BARCELOS *et al.*, 2021). As fake news contribuem para o aumento da descrença na ciência e nas instituições de saúde pública. Além disso, podem evidenciar a disseminação deliberada de notícias falsas sobre a COVID-19, como a promoção de tratamentos ineficazes e a ideia de isolamento vertical como argumento contra o distanciamento social (COUTO; BARBIERI; MATOS, 2021).

No cenário brasileiro, Gomes-Jr. e Frizzon (JR.; FRIZZON, 2019) analisaram fake news políticas usando técnicas de processamento de linguagem natural e ciência de redes. Usando como base métricas de centralidade de redes sobre pessoas e entidades citadas nas fake news, os autores caracterizaram a interdependência entre interesse da população e publicação de fake news. Os resultados sugerem que produtores de fake news se aproveitam do interesse da população sobre determinados temas.

2.2 Impactos da desinformação na vacinação

A exposição à desinformação diminui a receptividade em relação à vacinação e gera desconfiança em boa parte da população. Loomba et al. (LOOMBA S., 2021) mostram em um estudo realizado em 2021 que o índice de pessoas que definitivamente tomariam uma vacina contra o coronavírus tão logo estivesse disponível caiu de 54,1% para 47,9% no Reino Unido e de 42,5% para 36,1% nos EUA após a exposição dos entrevistados a fake news. Além disso, o número de pessoas que estavam inseguras mas inclinadas a não tomar a vacina aumentou 2,7% no Reino Unido e houve um aumento de 3,3% no grupo de pessoas que definitivamente não tomariam a vacina.

A aceitação da vacina também está associada à capacidade de um indivíduo identificar fake news, segundo mostram os resultados de um estudo publicado em março de 2021 no *Journal of Public Health*, onde indivíduos que obtiveram um baixo índice de detecção de fake news eram mais propensos a serem hesitantes ou contrários à vacinação (LOUZZANI-TOUHAMI K, 2021).

Em março de 2019 o Facebook implantou uma política de desinformação sobre vacinas. Pesquisadores norte americanos utilizaram modelos de regressão de séries temporais para avaliar o impacto dessa política no endosso (curtidas) dos usuários em postagens de conteúdo duvidoso. Para identificar essas postagens foram classificadas 172 páginas como anti e pró-vacina do Facebook e coletadas as postagens em um período de seis meses antes e seis

meses depois da implantação da política. O estudo concluiu que a política impactou moderadamente o número de endossos de conteúdo anti-vacina na plataforma e não causou impacto em conteúdos pró-vacina (GU *et al.*, 2022).

Mensurar o impacto da desinformação sobre a aceitação da vacina é uma questão científica aberta, complexa e desafiadora. Ao tentar correlacionar a vacinação com o engajamento das pessoas nas redes sociais Facebook e Twitter, Valensise et al. (VALENSISE *et al.*, 2021) analisaram 146 milhões de conteúdos produzidos por 20 milhões de contas entre janeiro de 2020 e abril de 2021, considerando os tweets e retuites que continham conteúdo duvidoso sobre a COVID-19 em diversos países (não incluindo o Brasil).

Para medir o impacto dessas mensagens nas intenções de vacinação, os pesquisadores utilizaram a pesquisa diária de tendências e impacto COVID-19 do Facebook em uma amostra de 1,6 milhão de usuários. Os resultados mostraram uma forte compatibilidade com um modelo nulo, o que sustentou a hipótese de que padrões alterados de consumo de informações não são um preditor confiável de mudança comportamental coletiva. Os resultados da pesquisa também demonstraram que a atenção mais ampla nas mídias sociais parece se resolver na polarização, com os propensos a vacinas e os hesitantes a vacinas mantendo suas posições. Vale ressaltar que os resultados obtidos nesse estudo se basearam na intenção de vacinação do público consultado, em um período onde, segundo a organização Our World in Data ¹, menos de 8% da população mundial tinha recebido pelo menos 1 dose da vacina. Nosso estudo analisou o impacto da desinformação nos índices reais de vacinação em um período onde os benefícios da vacinação já podiam ser notados pela população com a diminuição dos índices de internamento e mortes no país, o que pode explicar as diferentes conclusões apresentadas nos dois trabalhos.

Em um outro estudo, Pierri et al. (PIERRI *et al.*, 2021) mediram o impacto da desinformação nos índices de vacinação nos Estados Unidos e concluíram que a desinformação online está ligada à hesitação e recusa precoce da vacinação COVID-19. Utilizando como principal variável independente um conjunto de 55 milhões de mensagens do twitter sobre vacinação, colhidas entre 04/01/2021 e 25/03/2021, e uma lista de fontes classificadas como não confiáveis, os pesquisadores mediram a correlação entre a desinformação, os índices de vacinação e a hesitação vacinal. A pesquisa identificou geograficamente os usuários e mediu o impacto conforme a inclinação política do estado, avaliando o impacto entre os Republicanos e Democratas. Foram utilizados modelos de regressão para explicar a taxa de vacinação e a hesitação vacinal, e a causalidade de Granger para identificar a direção da influência entre as variáveis. O estudo demonstrou uma associação negativa entre as taxas de adesão à vacinação e desinformação online (Pearson $r = -0,49$, $p < 0,001$), comprovando que um aumento na quantidade média de desinformação online está significativamente associado a uma diminuição nas taxas diárias de vacinação por milhão. O estudo encontrou uma interação significativa entre partidarismo político e desinformação. Os condados republicanos sofrem uma influência maior, mas a medida que os níveis de desinformação aumentam, os condados democratas e republicanos convergem para

¹ https://ourworldindata.org/covid-vaccinations?country=OWID_WRL

o mesmo nível. A análise da direcionalidade demonstrou que a desinformação ajuda a prever a hesitação da vacina, fracamente no nível estadual ($p = 0,0519$) e fortemente no nível municipal ($p < 0,001$). Os resultados desse estudo são similares aos que obtivemos na análise do impacto da desinformação no Brasil, com correlação negativa em todos os estados, diferenciados pela inclinação política e fatores sociais.

2.3 Regressão Linear

O termo “Análise de Regressão” define um conjunto vasto de técnicas estatísticas usadas para modelar relações entre variáveis e prever o valor de uma ou mais variáveis dependentes a partir de um conjunto de variáveis independentes (CHEIN, 2019). O objetivo de testar qualquer uma dessas variáveis é verificar se existe alguma interação estatística entre elas, podendo assim serem utilizadas para ajudar na explicação nas mudanças na variável independente. Nesse estudo aplicaremos regressão linear múltipla, onde são utilizadas mais de uma variável independente para explicar as alterações da variável dependente. A regressão linear múltipla pode ser representada pela seguinte fórmula matemática:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \epsilon \quad (1)$$

Na fórmula Y é a variável dependente, α é o coeficiente linear, β_i representa o coeficiente de regressão das variáveis independentes, X_i são as variáveis independentes e ϵ é o erro aleatório associado à variável Y.

2.3.1 Regressão com Dados em Painel

Nesse estudo utilizamos regressão linear em dados de painel, que consiste em informações de várias unidades amostrais acompanhadas ao longo do tempo. Dessa forma, as observações são consideradas em duas dimensões, sendo uma delas a unidade amostral e a outra temporal (HANCK MARTIN ARNOLD; SCHMELZER, 2021). No nosso caso, os quantitativos de postagens, interações e compartilhamentos realizados no Facebook irão variar durante o tempo porém não irão variar entre os estados do país em cada recorte temporal, pois não existe uma divisão geográfica nas postagens.

O modelo geral para Dados em Painel pode ser representado através da seguinte fórmula matemática:

$$Y_{it} = \beta_{0it} + \beta_{1it} x_{1it} + \dots + \beta_{nit} x_{kit} + \epsilon \quad (2)$$

Na fórmula, i significa os diferentes indivíduos observados pelo tempo t . β_0 representa o parâmetro de intercepto, ou seja, onde a reta irá cruzar o eixo y , e β_k o coeficiente angular correspondente à k -ésima variável explicativa do modelo.

2.3.2 Fixed Effects

Segundo Duarte et al., "O modelo de efeitos fixos pretende controlar os efeitos das variáveis omitidas que variam entre indivíduos e permanecem constantes ao longo do tempo. Para isto, supõe que o intercepto varia de um indivíduo para o outro, mas é constante ao longo do tempo; ao passo que os parâmetros resposta são constantes para todos os indivíduos e em todos os períodos de tempo". No nosso caso, esses efeitos podem ser vistos como as variáveis de controle IDH e renda per capita, que não variaram durante o período de amostra utilizada na pesquisa.

O modelo de efeitos fixos pode ser representado da seguinte maneira:

$$Y_{it} = \alpha + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + \epsilon_{it} \quad (3)$$

Na fórmula, α_i representa um intercepto para cada indivíduo analisado. Segundo Duarte et al., "as diferenças de comportamento entre os indivíduos deverão ser captadas pelo intercepto".

Uma das aplicabilidades do modelo de efeitos fixos é quando as observações são obtidas para toda a população amostral e o que se deseja fazer são inferências para cada indivíduo dos quais dispõe-se de dados.

Nesse estudo consideramos os dados do Facebook para toda a população e queremos entender o possível impacto causado nos índices de vacinação em cada estado.

2.4 Ciência das Redes

"Ciência das redes é um campo acadêmico interdisciplinar que estuda redes complexas tais como redes de telecomunicações, redes de computadores, redes biológicas, redes cognitivas e semânticas, e redes sociais"².

Uma rede complexa é um tipo de grafo que apresenta uma estrutura topológica não-trivial de conexão, composta por um conjunto extremamente elevado de vértices (nós) que são conectados por meio de arestas (conexões, ligações ou links) (BARABASI, 2016).

Matematicamente, uma rede, ou grafo, $R = (V, A)$ é formada por um conjunto de N vértices, $V = v_1, v_2, \dots, v_N$ e um conjunto de M arestas, $A = a_1, a_2, \dots, a_M$. O vértice representa uma característica local da rede e as arestas estabelecem algum tipo de relação entre vértices.

² https://pt.wikipedia.org/wiki/Ci%C3%Aancia_das_redes#cite_note-5

As redes possuem algumas propriedades que possibilitam a sua análise e são úteis para resolver determinados tipos de problemas, como o grau de um vértice (quantidade de conexões com outros vértices), a sua centralidade, o peso da aresta e o conceito de comunidades.

2.4.1 Centralidade

A centralidade pode ser entendida como uma medida da influência ou importância de um determinado vértice em uma rede. Os principais tipos de centralidade são:

- Proximidade: distância média a que cada vértice está de todos os outros vértices na rede
- Intermediação: número dos caminhos mais curtos numa rede que passam por esse vértice.
- Grau: quantidade de ligações que um vértice específico possui na rede.

2.4.2 Comunidades

Comunidades, também chamadas de clusters ou agrupamentos, são grupos de vértices que tem grande probabilidade de compartilhar propriedades comuns ou tem papéis semelhantes no grafo (MACHADO, 2011).

A qualidade das comunidades obtidas pelos algoritmos disponíveis na literatura pode ser comparada através da modularidade, uma métrica que indica o quanto os elementos de uma comunidade estão conectados. Para identificar as comunidades são necessários algoritmos heurísticos, como o método Louvain (BLONDEL *et al.*, 2008). Este algoritmo possui complexidade linear no número de arestas, e é capaz de identificar comunidades com alta modularidade através de um algoritmo iterativo de duas fases. Observamos que o critério guloso do algoritmo é muito simples: ao construir comunidades, o algoritmo analisa os vértices em sequência (de 0 a $n-1$) (AIRES; NAKAMURA, 2017).

3 METODOLOGIA E PROCESSAMENTO DE DADOS

A metodologia utilizada para o processamento dos dados pode ser dividida em três etapas: captura dos dados, filtro das informações e classificação, conforme Figura 2

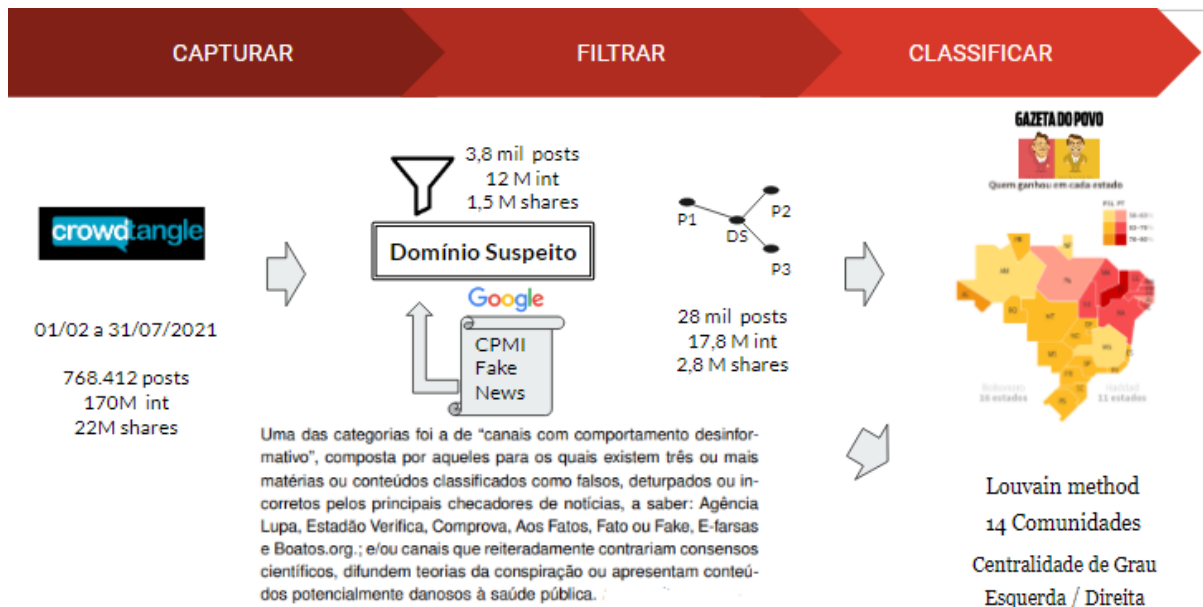


Figura 2 – Metodologia do processamento dos dados

3.1 Fontes de Dados

3.1.1 Postagens do Facebook

Os posts sobre vacinação foram extraídos do Facebook, através da plataforma Crowdtangle¹. Essa plataforma é uma iniciativa do Facebook para possibilitar análises e pesquisas científicas, tanto com os dados do Facebook como de outras plataformas de mídias sociais. O conjunto de dados extraído do Crowdtangle possui 41 colunas, das quais foram utilizadas neste trabalho a página de origem, a data da postagem, a classificação da página dentro do Facebook, o link compartilhado, o número de interações e o número de compartilhamentos de cada postagem, agrupados por semana do ano. Para a filtragem de postagens contendo possível desinformação sobre a vacina do coronavírus foi elaborada a fórmula de consulta abaixo, onde os termos relacionados à COVID-19 foram combinados com termos relacionados à vacinação para tentarmos capturar somente postagens que tratassem prioritariamente da vacina contra o coronavírus:

(pfizer, coronavac, janssen, covishield, astrazeneca, covid, sputnik, covaxin, coronavírus,

¹ <https://help.crowdtangle.com/en/articles/4201940-about-us>

com menos de 25 mil curtidas ou seguidores não estão sendo monitoradas pelo Crowdtangle e, portanto, não estão presentes nos dados utilizados nesse trabalho.

Como a exportação dos dados do Crowdtangle possui uma limitação de 300.000 registros por relatório, foram emitidos três relatórios com a separação de meses (janeiro - março, abril - maio, junho - julho 2021) e realizada a consolidação dos arquivos.

A consulta resultou em 768.412 postagens, com um total de 170.287.015 interações e um total de 22.988.314 compartilhamentos. Existem 69.828 registros de postagens com o total de interações igual a zero e 240.804 registros de postagens sem compartilhamentos, o que demonstra que o compartilhamento de uma postagem é mais difícil de acontecer do que uma interação, o que pode demonstrar um engajamento maior do usuário em relação à postagem quando o mesmo compartilha seu conteúdo.

Para confirmarmos se a consulta havia capturado postagens com o conteúdo desejado, realizamos a leitura de uma amostra aleatória contendo 100 postagens. Apesar de existir conteúdos tanto contra como a favor da vacina, o assunto principal das postagens era a vacinação da COVID em toda a amostra, demonstrando que o resultado da consulta estava dentro do esperado.

3.1.2 Fontes de desinformação

Para classificar as possíveis fontes de desinformação, cruzamos o resultado obtido no Crowdtangle com a lista de páginas geradas pela Consultoria Legislativa, a pedido da CPML Fake News, classificadas como disseminadoras de desinformação³. Segundo o relatório emitido pela Consultoria Legislativa, a metodologia para classificação da página como “canais com comportamento desinformativo” foi a seguinte:

Uma das categorias foi a de “canais com comportamento desinformativo”, composta por aqueles para os quais existem três ou mais matérias ou conteúdos classificados como falsos, deturpados ou incorretos pelos principais checadores de notícias, a saber: Agência Lupa, Estadão Verifica, Comprova, Aos Fatos, Fato ou Fake, E-farsas e Boatos.org.; e/ou canais que reiteradamente contrariam consensos científicos, difundem teorias da conspiração ou apresentam conteúdos potencialmente danosos à saúde pública. A consulta nas agências de notícias foi feita em suas páginas na internet, no repositório de matérias checadas ao longo de sua atuação.

Consideramos como possível desinformação as postagens onde a página de origem ou o domínio do link compartilhado estavam na lista da CPML. Desta forma, foram marcadas 3.765 postagens de 424 páginas distintas, com um total de 12.000.646 interações e um total

³ DOC 74 e anexos em <https://legis.senado.leg.br/comissoes/docsRecCPI?codcol=2292>, acesso em 12/11/2022

de 1.570.086 compartilhamentos. Após essa classificação, foi elaborado um grafo para identificar outras páginas que compartilharam os conteúdos classificados como questionáveis. Os detalhes da elaboração desse grafo estão descritos na Seção 3.3.2. Com isso, o número de postagens subiu para 28.229, com 17.887.803 interações e 2.830.723 de compartilhamentos. É importante enfatizar que estas páginas não são necessariamente fontes exclusivas de desinformação. Diversas páginas são referentes a fontes tradicionais de informação mas que em algum momento compartilharam ou publicaram conteúdo que foi considerado pela CPMI da Fake News como desinformação e por esse fato estamos considerando como fontes duvidosas nesse trabalho.

Para verificarmos se as postagens realizadas por esse grupo específico de páginas causava um efeito diferente nos índices de vacinação em relação ao grupo geral, analisamos a correlação de Pearson do quantitativo semanal de postagens com os índices de vacinação semanal, tanto para o grupo de páginas da CPMI como também para o restante das páginas presentes no relatório do Crowdtangle mas que não estavam presentes na lista da CPMI. Para as páginas contidas na lista da CPMI, a correlação com os índices de vacinação foi de -0,44 ($p\text{-value} < 0,01$) enquanto que para o restante das páginas a correlação foi de -0,27 ($p\text{-value} < 0,01$), conforme gráfico na Figura 4. Esta verificação sustenta a escolha do grupo de páginas informadas na CPMI como fonte de informações questionáveis, dado que a maior correlação das interações com as postagens desse grupo pode indicar uma influência mais relevante na diminuição da vacinação.

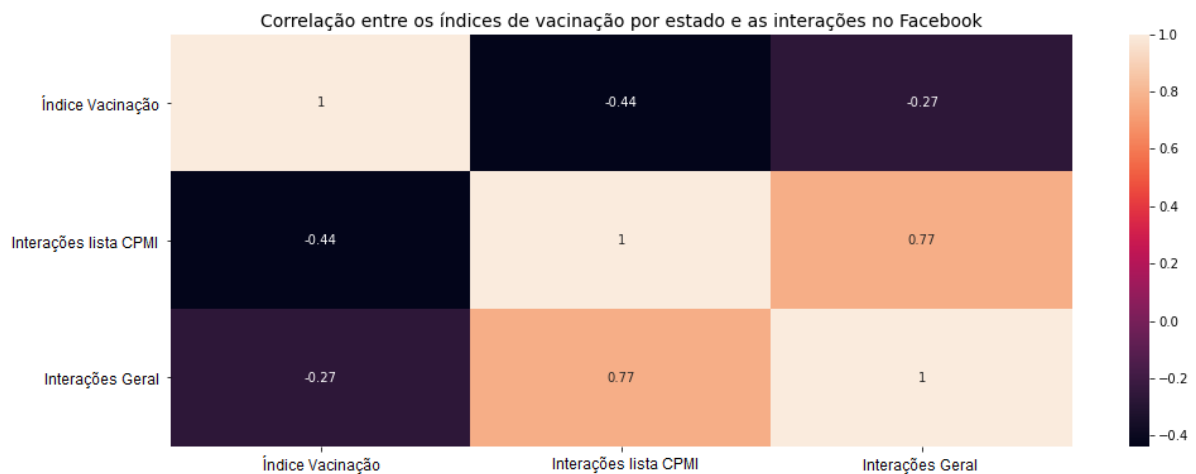


Figura 4 – Correlação entre a Vacinação e as lista CPMI e lista Geral

3.1.3 Dados sobre a vacinação, casos e óbitos do coronavírus

Os dados sobre a vacinação no Brasil foram extraídos do DATASUS (SAUDE, 2021), consolidados e armazenados em um banco de dados Postgres. A base de dados contém informações sobre 88,5 milhões de registros de vacinação, nos 26 estados brasileiros mais o Distrito

Federal e estão divididos em gênero, idade e categoria do paciente (grupo de risco, profissional de saúde e outros grupos). O período utilizado para análise foi de fevereiro a julho de 2021. Os dados utilizados nesse estudo foram a quantidade de doses aplicadas em cada município e estado e a data de aplicação. Para as análises, agrupamos os quantitativos semanais por estado. Esse agrupamento diminui as possíveis variações ocasionadas pela distribuição diferenciada das vacinas em todo o Brasil, principalmente nas primeiras semanas de vacinação.

Os dados da vacinação foram normalizados com o cálculo do índice de vacinação para cada 100 mil habitantes, agrupados por semana e estado da federação. O índice de vacinação (Iv) foi calculado através da fórmula abaixo, onde V representa a quantidade de vacinas aplicadas por semana em determinado estado (UF) e P a população do estado (UF):

$$Iv = (V/P) * 100.000 \quad (4)$$

Os dados sobre os casos e óbitos estão disponíveis para download através do portal do Ministério da Saúde⁴, que mantém as informações em painéis, com opção de filtro por região ou estado.

3.1.4 Dados sociais e políticos

Foram utilizadas três variáveis de controle para os estados analisados. Para análise da influência política, foi utilizado o resultado da eleição presidencial de 2018, considerando os estados onde o índice de votos válidos do então candidato Bolsonaro foi maior do que 50% como sendo de “inclinação à direita” e os estados onde o índice ficou igual ou abaixo de 50% como “inclinação à esquerda” (POVO, 2018). Também foram incluídos no modelo o IDH estadual, disponível no site AtlasBR (PNUD IPEA, 2019), e a renda per capita para os estados, capturados do site do IBGE (IBGE, 2021).

3.2 Modelo de Regressão Linear - Fixed Effects

Para a análise de impacto da desinformação nos estados foi desenvolvido um modelo de regressão linear em dados de painel com Fixed Effects, onde o índice de vacinação foi considerado como variável dependente, a quantidade de postagens, interações e compartilhamentos no Facebook como variáveis explicativas e IDH, renda per capita e inclinação política como variáveis de controle. Para capturar dependências temporais nas variáveis explicativas e na resposta, lags das respectivas variáveis foram introduzidos no modelo.

A quantidade de postagens foi utilizada para capturar o engajamento das páginas em relação à vacinação. O total de interações foi inserido para tentar capturar o quanto as postagens

⁴ https://infoms.saude.gov.br/extensions/covid-19_html/covid-19/.html, acesso em 20/12/2022

afetavam os seus leitores. A quantidade de compartilhamentos foi inserida para tentar capturar o engajamento dos leitores em replicar o conteúdo consumido.

Durante a análise dos dados notamos a presença de outliers nos dados de interações e compartilhamentos no Facebook. A existência desses outliers se deve principalmente ao fato de pouquíssimas páginas possuírem um grande número de consumidores dos conteúdos por elas produzidos, enquanto a grande maioria das páginas possuem poucos seguidores. Essa característica de concentração em poucas entidades é vista em diversas plataformas de mídias sociais, e no Facebook não é diferente. Nos dados utilizados nesse trabalho temos uma porcentagem de postagens de 0,02% com mais de 100 mil interações, enquanto 97,18% das postagens possuem menos de mil interações.

Para que os outliers não interferissem nos resultados da regressão linear, substituímos no modelo os campos onde os outliers foram identificados por novas variáveis contendo o cálculo do \log_2 dos valores. Com isso foi possível trabalhar com os valores destoantes sem eliminá-los da base de dados.

A fórmula matemática que representa o modelo é a seguinte:

$$I_{s,t} = \alpha + \sum_{\substack{i \in [0..2] \\ p \in (l,r)}} \beta_i^p \log_2 Int_{t-i}^p + \sum_{\substack{i \in [0..2] \\ p \in (l,r)}} \gamma_i^p Post_{t-i}^p + \sum_{\substack{i \in [0..2] \\ p \in (l,r)}} \theta_i^p Shares_{t-i}^p + \sum_{\substack{i \in [0..2] \\ p \in (l,r)}} \zeta_i^p Obitos_{t-i}^p \\ + \delta Idh_s + \lambda Renda_s + \mu Pol_s + \epsilon$$

(5)

Na fórmula I representa o índice de vacinação, s o estado (UF) e t a semana do ano. A variável i representa o lag temporal da variável, p a inclinação política das páginas no Facebook, onde l são as páginas de esquerda (left) e r as páginas de direita (right). As variáveis $\alpha, \beta, \gamma, \theta, \zeta, \delta, \lambda, \mu$ são os coeficientes lineares, Int a quantidade de interações no Facebook, $Post$ a quantidade de postagens e $Shares$ a quantidade de compartilhamentos no Facebook. A variável Idh representa o IDH médio de cada estado (UF), $Renda$ é a renda per capita do estado (UF), Pol é a porcentagem de votos válidos para o então candidato Bolsonaro em cada estado.

3.3 Análise Exploratória

3.3.1 Análise Temporal

A análise temporal das postagens, interações, compartilhamentos e índices de vacinação mostrou que, próximo da semana 12 de 2021, no final de março, há um aumento considerável nos índices de vacinação, com a chegada e distribuição de vacinas em todo o Brasil. No mesmo período é possível perceber no gráfico (Figura 5) o início de uma tendência de cresci-

mento do tema no Facebook, após um período de menor movimentação em relação ao início do ano. Após as primeiras semanas de aplicação da vacina, as postagens, interações e compartilhamentos oscilaram no decorrer do período, mostrando uma tendência de crescimento a partir da semana 23, início de junho, até o final do período analisado.



Figura 5 – Análise temporal Vacinas x Facebook

Gráfico de linhas comparando o quantitativo de postagens, interações e compartilhamentos no Facebook com o índice de vacinação ao longo do tempo

3.3.2 Análise de rede e identificação da orientação política

Para identificar o relacionamento entre as páginas disseminadoras de desinformação foi gerado um grafo não direcionado, com base no relatório de mensagens do Facebook obtido através do Crowdtangle. Os nós do grafo são as páginas presentes no relatório e uma aresta foi criada sempre que duas páginas compartilhavam conteúdo de um mesmo domínio, seja de um site externo ou de uma outra página do próprio Facebook. Os nós sem conexão com páginas presentes na lista de fontes duvidosas foram desconsiderados.

Para as postagens que não possuíam link compartilhado, ou estavam compartilhando um link do próprio Facebook, o nome da página foi considerado como domínio para a montagem do grafo. Com isso, quando uma página compartilhava a publicação de outra página do Facebook, mesmo sem um link de conteúdo, uma aresta era gerada no grafo entre as duas páginas.

O grafo resultante ficou com 51.384 vértices e 72.301 arestas. O grau médio é de 1,4 e o diâmetro da rede é de 19. O grafo completo pode ser visualizado na Figura 6.

Na área central do grafo pode ser vista as páginas com maior centralidade e suas conexões. Ao redor dessa área estão as páginas com menos expressão dentro dos dados coletados, com um número baixo ou único de conexões.

Filtrando o grafo com a lista de páginas questionáveis e analisando o grafo gerado somente com essas páginas, foi possível perceber uma clara divisão de grupos. Para identificar esses grupos, foi aplicado o método Louvain de detecção de comunidades (BLONDEL *et al.*, 2008), que consiste em um algoritmo heurístico que procura a otimização da modularidade dentro de cada comunidade.

Foram identificadas 14 comunidades na rede, sendo 8 com um número considerável de nós e 6 com um número inexpressivo de nós. As comunidades foram coloridas para melhor visualização, conforme Figura 7.

As páginas com maior centralidade de grau contidas em cada uma das comunidades foram analisadas a fim de identificar possíveis similaridades que justificassem a divisão no grafo. A análise demonstrou que existem dois grandes grupos de comunidades, divididos pela inclinação política, onde 8 comunidades possuem inclinação à direita e 6 comunidades inclinação à esquerda. Essa classificação foi feita com a análise do conteúdo das conexões dos nós de maior centralidade de cada comunidade. A Tabela 1 detalha e contextualiza as comunidades, seus nós centrais e a inclinação considerada.

O grupo de direita possui 165 páginas e realizou 6.966 postagens, tendo 11.734.590 interações e uma média de 1.684,55 interações por postagem. O grupo de esquerda possui 259 páginas, que produziram 21.263 postagens, com 6.153.213 interações, o que representa uma média de 289,38 interações por postagem.

Utilizamos o método do Coeficiente de Jaccard para analisar a possibilidade de novas conexões na rede. O método avaliou a possibilidade de conexão entre cada um dos vértices

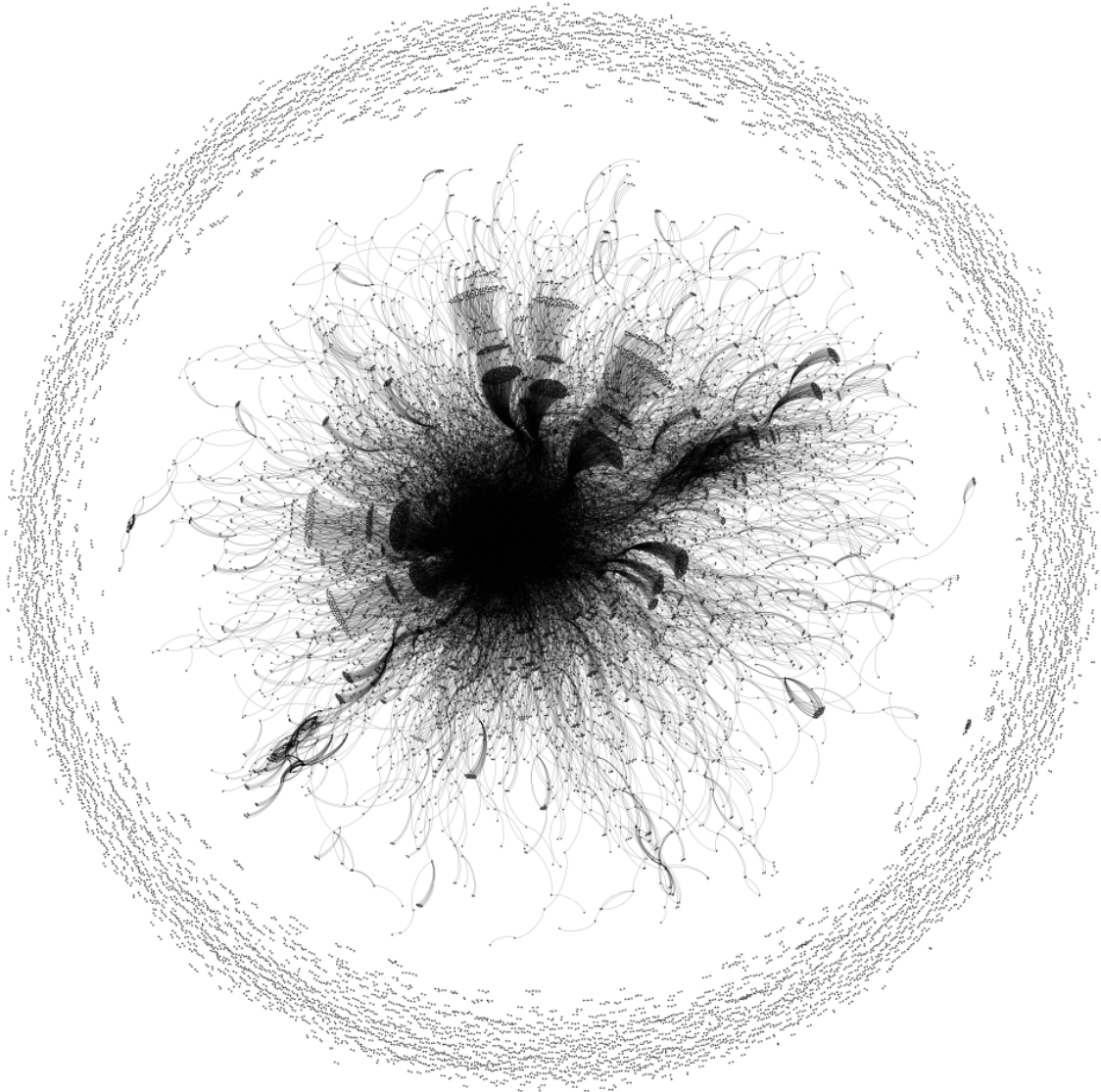


Figura 6 – Grafo representando as conexões entre páginas no Facebook

Rede de conexões entre as páginas do Facebook que postaram conteúdos de um mesmo domínio

do grafo e os valores foram comparados entre as comunidades e inclinação política das páginas. Dentro da rede analisada, a possibilidade de conexão entre páginas de inclinação política diferente é praticamente nula, enquanto a possibilidade de conexão entre páginas da mesma comunidade é maior do que 88%. Os resultados podem ser vistos na Figura 8.

3.3.3 Questões políticas e socioeconômicas

A análise exploratória também avaliou o impacto de variáveis políticas e socioeconômicas nos índices de vacinação. Os estados considerados de direita tiveram um impacto mais acentuado nos índices de vacinação em relação aos de esquerda, como pode ser visto no grá-

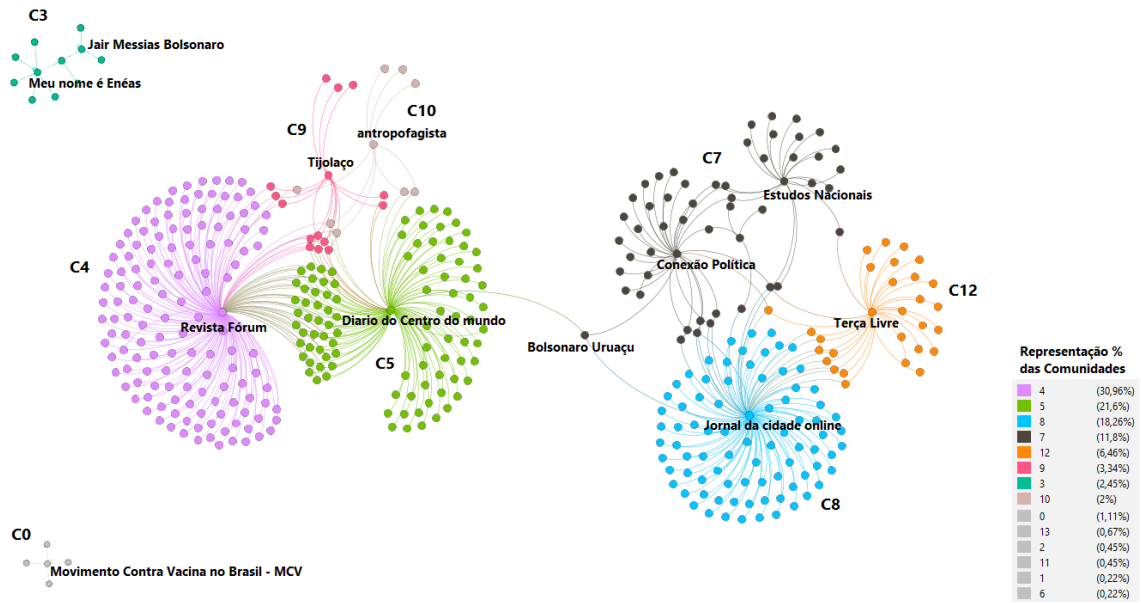


Figura 7 – Comunidades da rede
 Rede de conexões entre as páginas consideradas como questionáveis

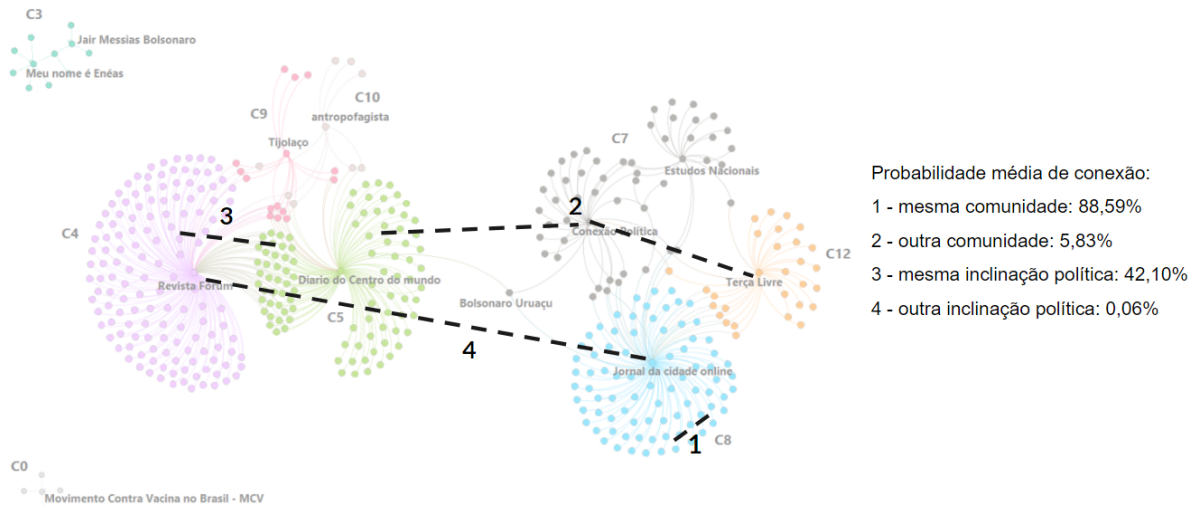


Figura 8 – Possibilidades de novas conexões na rede

fico na Figura 9. Esse maior decaimento da reta de regressão sugere que estados de direita têm a tendência de ter uma menor taxa de vacinação para o mesmo número de interações com as postagens do que os estados de esquerda. Esta é uma observação de correlação e o efeito pode não se confirmar ao se controlar por outras variáveis, como apresentaremos no Capítulo 4.

Analisando os estados pela faixa de IDH (gráfico da Figura 10), percebemos que os estados na 3ª faixa apresentaram o maior decaimento da reta, o que sugere que tiveram o maior impacto com o aumento de interações no Facebook.

Inclinação Política	Comunidade	Página de maior centralidade	Exemplo de conexão
Direita	0	Movimento Contra Vacina no Brasil - MCV	contrafatos;
Direita	1	Eduardo Bolsonaro	eduardobolsonaro;
Direita	2	Flavio Bolsonaro	flaviobolsonaro;
Direita	3	Meu nome é Enéas	Jair Messias Bolsonaro
Direita	6	Nando Moura	
Direita	7	Conexão Política	Estudos Nacionais; Bolsonaro Uruaçu; Pacto Contra o PT;
Direita	8	Jornal da Cidade Online	Chega de Corruptos; Brasil com Bolsonaro; Brasil contra o Comunismo; SomosTodosBolsonaro;
Direita	12	Terça Livre	Direita Calibrada; Movimento Brasil; Direita RJ;
Esquerda	4	Revista Fórum	Dilma Resistente; Fernando Haddad;
Esquerda	5	Diário do Centro do Mundo	Maria do Rosário; Galera Vermelha; Militantes de esquerda;
Esquerda	9	Tijolaço	Dr. Rosinha; Solidariedade a Dilma;
Esquerda	10	Antropofagista	Somos milhões de Lula; Partido dos Trabalhadores de Cachoeira Paulista;
Esquerda	11	Direita Opressora	noticiasconcursos;
Esquerda	13	Filósofo Paulo Ghiraldelli	uol; youtube; Ghiraldelli iniciou seu canal no YouTube em 2006. Começou a publicar vídeos com mais regularidade em 2019, no governo Jair Bolsonaro, ao qual o filósofo faz oposição ⁵

Tabela 1 – Classificação política das páginas de maior centralidade em cada comunidade

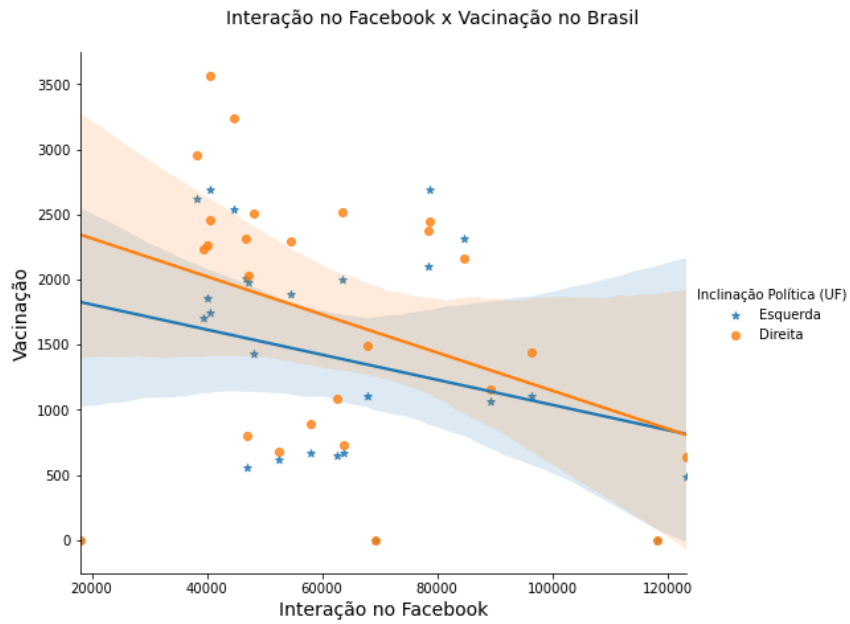


Figura 9 – Reta de regressão mostrando a influência das interações no Facebook por inclinação política do estado (UF)

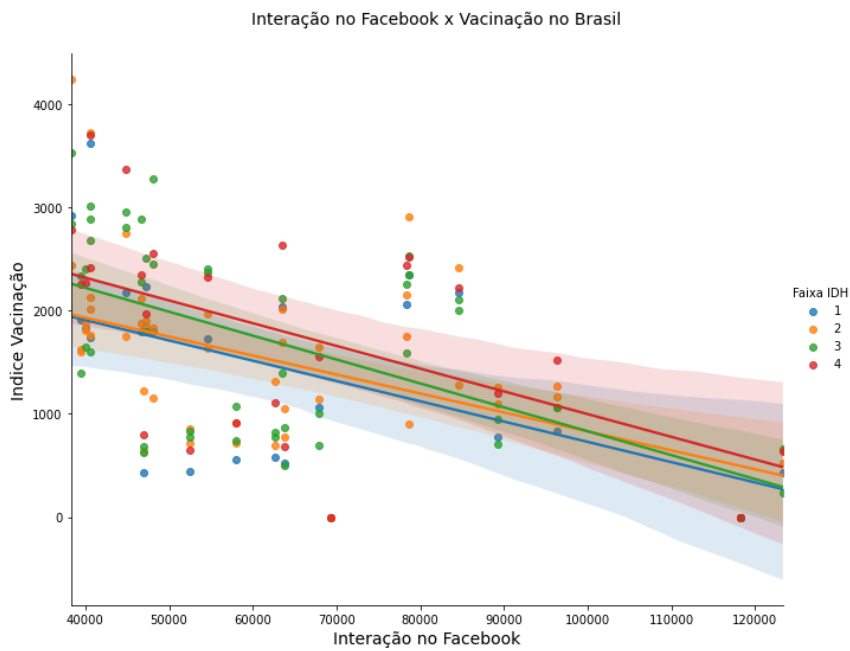


Figura 10 – Reta de regressão mostrando a influência das interações no Facebook por faixa de IDH

4 RESULTADOS

Na Seção 4.1 desse capítulo são apresentados os resultados obtidos com o modelo de regressão linear proposto na Seção 3.2 e as análises que podem ser realizadas a partir deles. Também é demonstrado o resultado da análise geográfica, na Seção 4.2.

4.1 Regressão Linear

Os resultados do modelo de regressão com Fixed Effects introduzido na Seção 3.2 são apresentados na Tabela 2. Executamos o modelo em duas versões, sendo que na versão (1) a inclinação política da fonte da postagem não foi considerada e na versão (2) dividimos as variáveis oriundas do Facebook em direita e esquerda, de acordo com a inclinação política da página emissora, conforme a metodologia descrita na Seção 3.3.2. O modelo (2) considera a inclinação política, onde o prefixo “dir” significa que a página possui uma inclinação de direita e “esq” de esquerda.

O resultado da versão (1) do modelo indica que as postagens e compartilhamentos de conteúdos elencados como questionáveis produziram um impacto negativo nos índices de vacinação, principalmente considerando o lag de 1 semana nas postagens e os compartilhamentos realizados na mesma semana da vacinação. As interações dos usuários com as postagens tiveram um índice positivo e estatisticamente significativos ($p\text{-value} < 0,01$), tanto para a mesma semana da vacinação como para as variáveis com lag de 1 e 2 semanas. Os compartilhamentos realizados na mesma semana da vacinação produziram um impacto negativo, porém menor do que as postagens. Os compartilhamentos com lag de 1 e 2 semanas não tiveram índices estatisticamente significativos ($p\text{-value} > 0,1$). As variáveis IDH, Renda per capita e inclinação política do estado tiveram índices positivos e significativos ($p\text{-value} < 0,01$) nessa versão do modelo.

Considerando a inclinação política das páginas, para as variáveis de interação, as postagens de direita apresentaram associação em geral positiva, enquanto os de esquerda não atingem significância estatística. Para o total de postagens, o grupo de direita apresentou uma associação em geral negativa e o grupo de esquerda não atingiu significância estatística. Para os compartilhamentos, tanto o grupo de esquerda quanto o grupo de direita tiveram resultados com significância estatística, onde o grupo de direita possui uma associação em geral positiva e o grupo de esquerda em geral negativa.

As variáveis de controle óbitos, IDH e inclinação política do estado não atingiram significância estatística nessa versão do modelo. A variável renda per capita do estado obteve significância estatística com uma associação positiva em relação à vacinação.

Na versão (1) do modelo, onde a inclinação política das postagens não foi considerada, o número de novos óbitos semanal teve um impacto positivo e estatisticamente significativo na vacinação. Na versão (2) do modelo, os índices ficaram com $p\text{-value} > 0,1$.

Tabela 2 – Fixed Effects

	<i>Variável Dependente: Índice de vacinação</i>	
	(1)	(2)
Interações	633.001***	
Interações lag1	520.194***	
Interações lag2	155.476*	
Posts	-0.461***	
Posts lag1	-1.223***	
Posts lag2	-0.383***	
Compartilhamentos	-0.008***	
Compartilhamentos lag1	-0.0004	
Compartilhamentos lag2	0.001	
Interações direita		4.095***
Interações dir lag1		1.619**
Interações dir lag2		-1.731*
Interações esquerda		0.447**
Interações esq lag1		0.364
Interações esq lag2		-0.338
Posts direita		-14.962***
Posts dir lag1		-7.201***
Posts dir lag2		9.918***
Posts esquerda		0.283
Posts esq lag1		-1.838*
Posts esq lag2		1.210
Compartilhamentos direita		0.003*
Compartilhamentos dir lag1		0.010***
Compartilhamentos dir lag2		-0.002*
Compartilhamentos esquerda		-0.034***
Compartilhamentos esq lag1		-0.009***
Compartilhamentos esq lag2		-0.006**
Óbitos Novos	0.307***	0.086
Óbitos Novos lag1	-0.003	0.022
Óbitos Novos lag2	-0.089	-0.017
IDH	345.849***	101.470
Renda	531.515***	332.738***
Inclinação Política	268.215**	74.814
Observations	657	657
R ²	0.544	0.679
Adjusted R ²	0.514	0.652
F Statistic	48.921*** (df = 15; 615)	53.321*** (df = 24; 606)

Note:

*p<0.1; **p<0.05; ***p<0.01

4.2 Análise Geográfica

Para identificar possíveis estados mais vulneráveis ao conteúdo questionável, implementamos uma adaptação do modelo completo da regressão apresentado anteriormente. A fórmula deste modelo é apresentada na Equação 4.2. Simplificamos o modelo retirando as variáveis de lag e incluímos variáveis de postagens de direita e dummies representando os estados (E), mantendo as demais variáveis do modelo original. Neste novo modelo, os coeficientes das variáveis de postagens (γ_s) dimensionam o quanto o conteúdo pode ter influenciado em cada estado.

$$I_{s,t} = \alpha + \gamma_s(E_s \times Posts_t) + \sum_{p \in (l,r)} \beta_t^p Int_t^p + \sum_{p \in (l,r)} \theta_t^p Shares_t^p + \sum_{p \in (l,r)} \zeta_t^p Obitos_t^p + \delta Idh_s + \lambda Renda_s + \mu Pol_s + \epsilon$$

As postagens analisadas causaram um impacto negativo em todos os estados e todos os coeficientes ficaram com p-value < 0,01. Os valores dos coeficientes de regressão das postagens em relação a cada estado podem ser vistos na tabela 3.

Os coeficientes das interações resultantes e os p-values de cada estado podem ser observados nos mapas da figura 11. Como é possível perceber no mapa, os coeficientes obtidos no modelo de regressão linear entre as interações e a vacinação ficaram negativos para todos os estados, mas observa-se uma influência maior nos estados de São Paulo, seguido de Minas Gerais e Rio de Janeiro. Os estados que tiveram uma influência menor, porém negativa, foram Espírito Santo, Acre e Mato Grosso do Sul. Todos os estados tiveram um p-value < 0,01, incluindo o Distrito Federal.

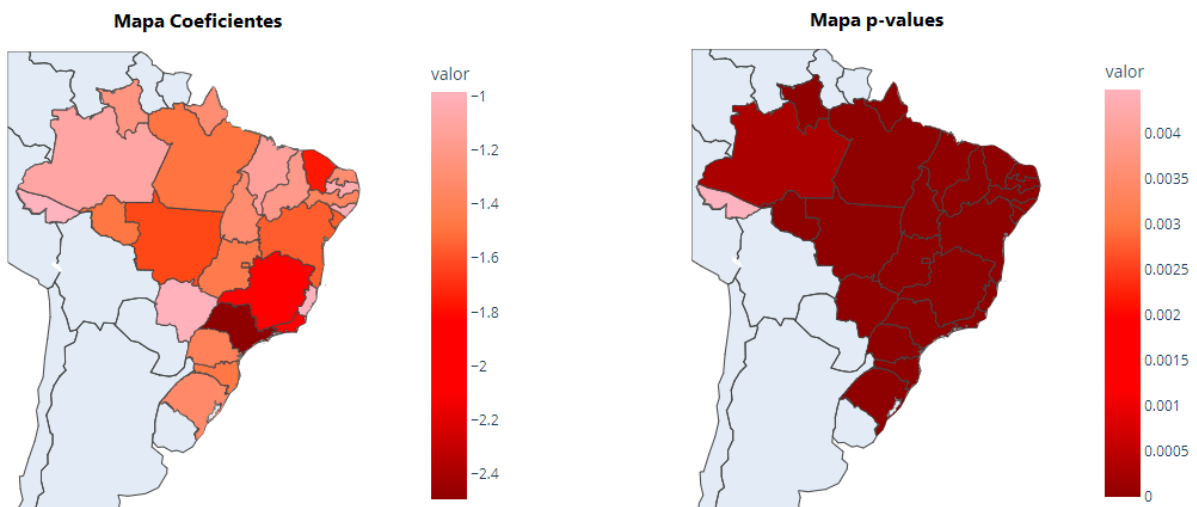


Figura 11 – Mapas Coeficientes de Regressão e p-values

<i>Variável dependente:</i>	
(índice de vacinação)	
Intercept	1711.768***
AC	-0.986***
AL	-1.035***
AM	-1.102***
AP	-1.303***
BA	-1.572***
CE	-1.757***
DF	-1.506***
ES	-0.984***
GO	-1.446***
MA	-1.136***
MG	-1.839***
MS	-0.996***
MT	-1.642***
PA	-1.493***
PB	-1.048***
PE	-1.378***
PI	-1.205***
PR	-1.401***
RJ	-1.829***
RN	-1.315***
RO	-1.488***
RR	-1.243***
RS	-1.339***
SC	-1.476***
SE	-1.578***
SP	-2.492***
TO	-1.317***
Interações	0.004***
Óbitos Novos	0.221***
IDH	-37.485
Inclinação Política	-14.220
Renda	98.319
Compartilhamentos	-0.018***
Observations	659
R^2	0.386
Adjusted R^2	0.353
Residual Std. Error	848.865(df = 625)
F Statistic	11.899*** (df = 33.0; 625.0)

Note: *p<0.1; **p<0.05; ***p<0.01

Tabela 3 – Resultado da adaptação do modelo para capturar influência estadual

Ressaltamos que não foi realizada uma distinção da localidade das postagens no Facebook, sendo consideradas em sua totalidade como de abrangência nacional. A distinção geográfica foi realizada somente sobre os índices de vacinação, considerando o município de aplicação das vacinas.

4.3 Limitações e discussões futuras

Obter dados e modelar as influências de conteúdo questionável sobre a vacinação é um trabalho desafiador. Para oferecer resultados mais conclusivos, é preciso compreender melhor os dados usados e a dinâmica representada pelo modelo. A pesquisa das postagens com desinformação foi limitada a um grupo específico de páginas, considerando a lista de origens presentes no relatório da CPMI Fake News e suas conexões no Facebook. Essa restrição limitou o número de postagens analisadas e pode ter afetado os resultados alcançados. A classificação das postagens para esse estudo considerou somente sua origem, sem analisar o conteúdo das postagens. Por isso, é possível que existam postagens contra e a favor da vacinação, assim como de conteúdo verídico ou inverídico na base de dados, o que impacta diretamente nas conclusões.

Esse estudo não considerou a distribuição regional das vacinas, nem a capacidade de aplicação das doses de cada localidade. A normalização dos dados de vacinação em nível estadual e semanal diminui as possíveis diferenças de resultados no impacto em municípios que por ventura não estivessem aplicando vacinas em determinados dias por falta de insumos.

Baseado nas análises iniciais desse estudo, foi elaborado um artigo ¹ e publicado no WEBMEDIA 2022 ². Após a publicação do artigo e apresentação na conferência, foi verificado um erro de leitura em parte dos dados, afetando cerca de 4% dos dados e consequentemente alguns resultados do modelo. Os dados afetados são referentes às páginas com os maiores números de interações. A percepção desse erro e a análise dos impactos motivaram o tratamento de outliers aplicado nesse estudo, conforme descrito na Seção 3.2. Outro item do artigo retificado nesse estudo foi a origem da lista de páginas consideradas como fonte de desinformação. Durante a elaboração do artigo tínhamos o entendimento de que a lista tinha sido fornecido pelo Google para a CPI da Pandemia, quando na realidade a lista foi utilizada na CPI da Pandemia porém foi elaborada na CPMI da Fake News, conforme metodologia descrita na Seção 3.1.2. Os erros encontrados após a publicação do artigo foram corrigidos, o modelo foi evoluído e as informações foram atualizadas conforme descrito nesse trabalho.

Uma possibilidade de pesquisa futura seria a análise textual do conteúdo das postagens, buscando identificar as postagens que explicitamente foram contrárias à vacinação, com a possibilidade de expandir as buscas para outras plataformas de redes sociais, além do Facebook. Também poderiam ser utilizados outros métodos de análise de regressão, assim como

¹ <https://sol.sbc.org.br/index.php/webmedia/article/view/22087>

² <https://webmedia.org.br/2022/>

a incorporação de novas variáveis ao modelo e a utilização de dados melhores de interação com a desinformação. O estudo do impacto em diferentes níveis sociais, culturais e econômicos também é vislumbrado para trabalhos futuros.

5 CONCLUSÃO

Nesse trabalho avaliamos a influência de postagens oriundas de fontes questionáveis de informação, considerando um relatório de classificação realizado pela CPMI Fake News, sobre os índices de vacinação contra o coronavírus no Brasil. O objetivo principal era avaliar se as postagens, interações dos usuários e compartilhamentos desses conteúdos influenciavam a decisão das pessoas em receber a vacina. Para isso, capturamos as postagens do Facebook entre janeiro e julho de 2021, utilizando a ferramenta Crowdtangle, com base em uma fórmula de consulta que buscava termos relativos à vacina do COVID.

Utilizando o método Louvain, identificamos as comunidades do grafo e classificamos a inclinação política de cada uma delas através da análise das páginas de maior centralidade de cada comunidade. Após esse processamento, utilizamos um modelo de regressão linear com Fixed Effects, onde a variável independente é o índice de vacinação e as variáveis explicativas são o número de postagens, de interações, compartilhamentos e novos óbitos por semana.

Os resultados mostraram que as postagens analisadas causaram em geral um impacto negativo nos índices de vacinação, assim como o número de compartilhamentos desses conteúdos. Levando em conta a inclinação política das páginas emissoras, as postagens realizadas pelo grupo de direita na mesma semana da vacinação causaram um impacto negativo (-14.962). As postagens de esquerda também causaram um impacto negativo (-1.838), porém menor do que as postagens de direita.

Para avaliarmos com mais assertividade o real impacto da desinformação sobre a vacinação, precisaríamos expandir a fonte de páginas consideradas como questionáveis, inclusive com a obtenção de dados de outras plataformas além do Facebook. Outras questões também precisariam ser consideradas no modelo, como a disponibilidade das vacinas em cada região e capacidade de aplicação em cada localidade. Algumas questões pessoais, que interferem diretamente na decisão de tomar ou não uma vacina, são extremamente difíceis de capturar, principalmente em grande escala.

O modelo se mostrou sensível mesmo a pequenas modificações nas variáveis, gerando alterações significativas nos resultados. Essa sensibilidade dificulta as interpretações e deve ser levada em conta nas conclusões baseadas nesses resultados.

Por fim, é importante ressaltar que o consumo de informações da internet deve ser feito com cuidado e avaliação crítica, buscando a confirmação do conteúdo em fontes confiáveis e com o devido zelo antes do compartilhamento.

REFERÊNCIAS

- AIRES, V.; NAKAMURA, F. Aplicação de medidas de centralidade ao método louvain para detecção de comunidades em redes sociais. **XLIX Simpósio Brasileiro de Pesquisa Operacional**, UEM - Universidade Estadual de Maringá, ago 2017. Disponível em: <http://www.din.uem.br/sbpo/sbpo2017/pdf/169162.pdf>.
- BARABASI, A.-L. **Network Science**. [S.l.]: Cambridge University Press, 2016.
- BARCELOS, T. d. N. de *et al.* Análise de fake news veiculadas durante a pandemia de covid-19 no brasil. **Revista Panamericana de Salud Pública**, SciELO Public Health, v. 45, p. e65, 2021.
- BLONDEL, V. D. *et al.* Fast unfolding of communities in large networks. **Journal of Statistical Mechanics: Theory and Experiment**, IOP Publishing, v. 2008, n. 10, p. P10008, oct 2008. Disponível em: <https://doi.org/10.1088/1742-5468/2008/10/P10008>.
- BRASIL. **Comissão Parlamentar Mista de Inquérito - Fake News**. 2019. SENADO. Disponível em: <https://legis.senado.leg.br/comissoes/comissao?codcol=2292>. Acesso em: 04 mai. 2019.
- CHEIN, F. **Introdução aos modelos de regressão linear: um passo inicial para compreensão da econometria como uma ferramenta de avaliação de políticas públicas**. [S.l.]: Escola Nacional de Administração Pública (Enap), 2019.
- COUTO, M. T.; BARBIERI, C. L. A.; MATOS, C. C. d. S. A. Considerações sobre o impacto da covid-19 na relação indivíduo-sociedade: da hesitação vacinal ao clamor por uma vacina. **Saúde e Sociedade**, SciELO Brasil, v. 30, 2021.
- GU, J. *et al.* The impact of facebook's vaccine misinformation policy on user endorsements of vaccine content: An interrupted time series analysis. **Vaccine**, v. 40, n. 14, p. 2209–2214, 2022. ISSN 0264-410X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0264410X22002092>.
- HANCK MARTIN ARNOLD, A. G. C.; SCHMELZER, M. **Introduction to Econometrics with R**. [S.l.]: University of Duisburg-Essen, 2021.
- I OUAZZANI-TOUHAMI K, M. A. T. N. S. S. T. C. C. g. M. Acceptance of a covid-19 vaccine is associated with ability to detect fake news and health literacy. **J Public Health (Oxf)**, 2021.
- IBGE, A. **IBGE divulga rendimento domiciliar per capita 2020**. 2021. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/30129-ibge-divulga-o-rendimento-domiciliar-per-capita-2020>.
- JR., L. G.; FRIZZON, G. Fake news and brazilian politics – temporal investigation based on semantic annotations and graph analysis. *In*: **Anais do XXXIV Simpósio Brasileiro de Banco de Dados**. Porto Alegre, RS, Brasil: SBC, 2019. p. 169–174. ISSN 2763-8979. Disponível em: <https://sol.sbc.org.br/index.php/sbbd/article/view/8818>.
- LATINO, V. New study: Facebook is primary driver of covid-19 misinformation in the latinx community, fueling vaccine hesitancy. **Internet**, 2021.
- LIMA, E. J. d. F.; ALMEIDA, A. M.; KFOURI, R. d. Á. Vacinas para covid-19-o estado da arte. **Revista Brasileira de Saúde Materno Infantil**, SciELO Brasil, v. 21, p. 13–19, 2021.
- LOOMBA S., d. F. A. P. S. **Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA**. [S.l.]: Nature Human Behavior (2021), 2021.

MACHADO, F. M. Detecção de comunidades em grafos multicamada muito grandes. Comunicação pessoal por e-mail. 2011.

PIERRI, F. *et al.* Online misinformation is linked to early covid-19 vaccination hesitancy and refusal. arXiv, 2021. Disponível em: <https://arxiv.org/abs/2104.10635>.

PNUD IPEA, F. **Radar IDHM: Evolução do IDHM e seus índices componentes no período de 2012 a 2017**. 2019. Disponível em: <http://atlasbrasil.org.br/acervo/biblioteca>.

POVO, G. do. **RESULTADO DAS ELEIÇÕES 2018**. 2018. Disponível em: <https://especiais.gazetadopovo.com.br/eleicoes/2018/resultados/mapa-eleitoral-de-presidente-por-municipios-2turno/>.

SAUDE, M. da. Dados referentes à campanha nacional de vacinação contra covid-19: Campanha nacional de vacinação contra covid-19 - conjuntos de dados - open data. **Internet**, 2021.

TRANSMISSÍVEIS, D. de Análise em Saúde e Vigilância de D. C. N. **Saúde Brasil 2020/2021 : uma análise da situação de saúde diante da pandemia de covid-19, doença causada pelo coronavírus SARS-CoV-2**. [S.l.]: Ministério da Saúde (2022), 2022.

VALENSISE, C. M. *et al.* **Lack of evidence for correlation between COVID-19 infodemic and vaccine acceptance**. 2021. Disponível em: <https://arxiv.org/abs/2107.07946>.

VASCONCELLOS-SILVA, P. R.; CASTIEL, L. D. Covid-19, as fake news e o sono da razão comunicativa gerando monstros: a narrativa dos riscos e os riscos das narrativas. **Cadernos de Saúde Pública**, SciELO Brasil, v. 36, 2020.