

Câmaras de eco e mis/desinformação sobre vacinas contra a COVID-19 no Twitter: abordagem baseada em machine learning e análise de redes sociais

Echo chambers and vaccines against COVID-19 mis/disinformation on Twitter: machine learning and network analysis-based approach

Câmaras de eco y desinformación sobre vacunas contra COVID-19 en Twitter: enfoque basado en machine learning y análisis de redes

Recebido: 17/01/2023 | Revisado: 29/01/2023 | Aceitado: 02/02/2023 | Publicado: 09/02/2023

Arthur da Silva Lopes

ORCID: <https://orcid.org/0000-0001-9137-3184>

Universidade Federal da Bahia, Brasil

E-mail: arthur.lopes@ufba.br

Antonio Brotas

ORCID: <https://orcid.org/0000-0001-8438-2445>

Fundação Oswaldo Cruz, Brasil

E-mail: brotas@bahia.fiocruz.br

Resumo

Haja visto a importância infodemiológica das câmaras de eco na disseminação de mis/desinformação, objetivou-se analisar as redes de interação dos usuários mais expostos à mis/desinformação ou controvérsia sobre vacinas no âmbito da pandemia COVID-19. Para tal, é proposto nesta investigação uma metodologia baseada em machine learning e Análise de Redes Sociais para detecção automatizada de conteúdo controverso e mis/desinformativo sobre vacinas, através da qual chegou-se a um modelo com 92% de acurácia. Dos quase 24 milhões de tweets coletados, 12.4 milhões (52%) foram assinalados como controversos e/ou potenciais à mis/desinformação, sendo os meses de janeiro e junho de 2021 aqueles de maior atividade, sendo analisados através de uma coorte. Diferentemente dos trabalhos anteriores, analisou-se a rede de todas as maneiras de interagir no Twitter, e de toda a estrutura textual dos tweets – não apenas links ou hashtags -. No que concerne à conversação sobre vacinas contra COVID-19, os achados foram diferentes daqueles associados à discussão político-partidária anteriormente descritos na literatura, uma vez que a rede de menções e respostas privilegia relações heterofílicas, e que conformações de “eco” não foram observáveis. Finalmente, faz-se necessário novos estudos para melhor compreender a disseminação de desinformação acerca das vacinas no Twitter.

Palavras-chave: Vacina; Infodemia; Fake news; Câmaras de eco; Twitter; Machine learning.

Abstract

Given the infodemiological importance of echo chambers in the dissemination of mis/disinformation, we aimed to analyze the interaction networks of users most exposed to mis/disinformation or controversy about vaccines in the context of the COVID-19 pandemic. To this end, a methodology based on machine learning and Social Network Analysis is proposed in this research for automated detection of controversial and mis/disinformative content about vaccines, through which a model with 92% accuracy was achieved. Out of the nearly 24 million tweets collected, 12.4 million (52%) were flagged as controversial and/or potential for mis/disinformation, and the months of January and June 2021 were those with the highest activity, being analyzed through a cohort. Unlike previous work, we analyzed the network of all ways of interacting on Twitter, and the entire textual structure of the tweets - not just links or hashtags -. Regarding the conversation about COVID-19 vaccines, the findings were different from those associated with party-political discussion previously described in the literature, since the network of mentions and replies privileges heterophilic relationships, and "echo" conformations were not observable. Finally, further studies are needed to better understand the dissemination of misinformation about vaccines on Twitter.

Keywords: Vaccine; Infodemic; Fake news; Echo chambers; Twitter; Machine learning.

Resumen

Dada la importancia infodemiológica de las cámaras de eco en la difusión de información errónea/desinformación, nos propusimos analizar las redes de interacción de los usuarios más expuestos a la información errónea/desinformación o controversia sobre las vacunas en el contexto de la pandemia COVID-19. Para ello, en esta investigación se propone una metodología basada en aprendizaje de máquina y Análisis de Redes Sociales para la

detección automatizada de contenidos controvertidos y erróneos/desinformativos sobre vacunas, a través de la cual se alcanzó un modelo con un 92% de precisión. De los casi 24 millones de tuits recogidos, 12,4 millones (52%) fueron marcados como controvertidos y/o de potencial desinformación, siendo los meses de enero y junio de 2021 los de mayor actividad, analizándose a través de una cohorte. A diferencia de trabajos anteriores, analizamos la red de todas las formas de interactuar en Twitter, y toda la estructura textual de los tuits -no sólo los enlaces o los hashtags-. En cuanto a la conversación sobre las vacunas COVID-19, los resultados fueron diferentes de los asociados a la discusión partidista descritos anteriormente en la literatura, ya que la red de menciones y respuestas privilegia las relaciones heterófilas, y que no se observaron conformaciones "eco". Por último, se necesitan más estudios para comprender mejor la difusión de información errónea sobre las vacunas en Twitter.

Palabras clave: Vacuna; Infodemia; Fake news; Cámaras de eco; Twitter; Machine learning.

1. Introdução

Ante a pandemia da COVID-19, assiste-se a uma acentuação da desordem informacional, caracterizada pela disseminação de grande quantidade de informações a ponto de se tornar inviável, ou mesmo impossível, discriminar sua relevância ou veracidade; fenômeno nomeado pela Organização Mundial da Saúde enquanto “infodemia” (World Health Organization, 2020), e se agravado muito em função do advento da internet, da democratização dos meios de edição de mídias, e da emergência das plataformas de mídias sociais (Lima et al., 2020; Amaral & Santos 2019).

Fake news, desinformação, informações não-factuais/falsas, boatos, rumores, afinal, qual é exatamente o fenômeno ao qual designa-se tantos termos diferentes? Amaral e Santos (2019) apontam que aquilo que se concebe *fake news* não é o mais ideal para nomeá-lo, uma vez que omite, ou não é capaz de apreender, a dimensão da intencionalidade existente no referido fenômeno.

Isso, pois criar, produzir e compartilhar informações falsas deliberadamente com vistas à satisfação de interesses pessoais não é o mesmo que distribuí-las de maneira não intencional. Com relação ao primeiro, Wardle e Derakhshan (2017) intitulam *disinformation*, enquanto ao último, *misinformation*; daí que o uso do termo mis/desinformação seria um esforço nosso de indicar referência a ambas as dimensões na língua portuguesa.

Acerca da sua disseminação mediada por plataformas de mídia social, autores como Van der Linden (2022) destacam o papel da rede social (conjunto de indivíduos e suas interações nas plataformas de mídia social) onde usuários estão localizados na determinação da exposição à mis/desinformação, uma vez que contribui tanto para a supressão quanto para a viabilização de coexistência e surgimento de novas variantes circulantes de informações não-factuais.

Estudos que se debruçam sobre essas redes são, portanto, de destacada importância; e com efeito, se constituem enquanto frequente unidade de análise no que concerne à compreensão da dinâmica informacional em plataformas de mídia social que possuem maior potencial para ocorrência de infodemia, tal qual o Twitter e o Facebook (van der Linden, 2022).

Trabalhos anteriores

A nível fenomenológico, tem-se como evidência a existência de polarização e o efeito dos conflitos político-partidários na disseminação de desinformação, pois conseguem (re)configurar as redes sociais construídas em plataformas de mídia social e (re)orientar as vias de tráfego de informações (Du & Gregory, 2017). Nesse sentido, Törnberg (2018) demonstrou correlação entre configurações polarizadas de redes e maior “virulência” do conteúdo mis/desinformativo.

Outra evidência é a homofilia, ou seja, a tendência de que indivíduos com opiniões semelhantes possuem maior probabilidade de interagir entre si do que com quem é divergente. Em consequência, outra hipótese diz respeito à presença daquela nomeada de câmara de eco, ou seja, que informações que circulam em um grupo de similares, geralmente, não circula noutro.

Tanto a homofilia quanto as câmaras de eco foram identificadas por Soares e colaboradoras (2021), ao analisarem a

circulação dos 100 links mais relevantes dentre aqueles coletados nos quais se fazem presentes conteúdos desinformativos acerca da hidroxiclороquina no Facebook; e por Cinelli e colaboradores (as) (2020), os quais também analisaram links, mas no Twitter e ligados a outros temas controversos (controle de armas, *Obamacare* e aborto) em um corpus de 92 milhões de tweets.

Através de uma amostra de aproximadamente 60 bilhões de tweets em língua inglesa, e análise da rede de interações da conversação com base nas menções e retweets, Mønsted e Lehmann (2022) concluem que os achados indicam uma conversação sobre vacina altamente polarizada e caracterizada pelo que chamaram de câmaras de eco epistêmicas, onde usuários(as), dependendo da sua atitude vis-à-vis das vacinas (se pró ou anti-vacina), não só interagem nos termos da homofilia, como também são expostos à fontes de informação grupo-específicas, não correspondendo àquelas que se fazem presentes na conversação geral.

No já mencionado estudo sobre a hidroxiclороquina no Twitter, as autoras também apontam para um achado semelhante: “Esse enquadramento criou uma rede polarizada, onde links que poderiam informar melhor sobre o uso da droga simplesmente não circulam onde circula a desinformação” (Soares et al., 2021).

Dessa forma, sublinha-se aqui um importante, e preocupante, impasse acerca da infodemiologia da desinformação no Twitter concernindo a saúde pública, o jornalismo de *fact-checking* e a divulgação científica: se a mis/desinformação possui caráter viral, e existem grupos específicos na rede atuando enquanto vetores para sua presença e distribuição, expondo outros(as) usuários(as) a esse tipo de conteúdo; onde uma vez “infectados” e integrantes da “rede de mis/desinformação”, passam a ser menos propensos ao acesso à conteúdo que poderia restituir sua saúde informacional, quais as possibilidades de intervenção?

Destarte, o presente trabalho se caracteriza enquanto infodemiológico (Eysenbach, 2009), na medida que visa analisar a ocorrência e distribuição de (mis/des)informações de interesse da saúde pública; e, tem como objetivo geral, analisar a interação existente entre os usuários(as) mais expostos à mis/desinformação no Twitter na conversação acerca das vacinas nos períodos de maior atividade na plataforma, visando melhor compreender de que forma ela se distribui em rede através das seguintes perguntas norteadoras da pesquisa (PP):

PP1: como se configuram as redes sociais provenientes das diferentes formas de interagir no Twitter (retweets, respostas/menções e citações) na conversação sobre vacinas?

PP2: a câmara de eco é um fenômeno por meio do qual caracteriza-se a circulação de informações concernindo vacinas contra COVID-19 no Twitter?

Com relação à PP1, Conover et al. (2011) demonstraram que existem diferenças na estrutura das redes de interação provenientes da atividade de *retweeting* e menção no Twitter, mas ambas se apresentaram homofílicas. Entretanto, generalizar esses achados (que foram decorrentes da análise de uma conversação voltada à política partidária) para outros temas seria extrapolar os limites razoáveis da capacidade dedutiva dos dados analisados por esses autores (amostra de apenas 250 mil tweets, feita por conveniência, com interações contexto-específicas; e unidade textual para teste de similaridade foi apenas a hashtag, e não o texto na íntegra – como realizado no presente trabalho).

Assim, tem-se enquanto principais contribuições da presente pesquisa:

1) Proposta metodológica de captura da conversação onde se tem maior probabilidade de ocorrência de mis/desinformação, ou controvérsia, sobre vacinas usando aprendizado de máquina com resultados superiores àquele apresentado em trabalho anterior (Ferreira et al., 2022);

2) Compreensão não só das redes provenientes das menções e retweets, como nos trabalhos anteriores, mas também das citações. Diferentemente dos estudos sobre política partidária, em uma conversação sobre vacinas, as interações do tipo menção foram heterofílicas;

3) Apresenta evidências advindas da análise de todo o corpo textual dos tweets, diferentemente dos trabalhos

anteriores que analisaram links ou hashtags;

4) Aponta a necessidade de, ao se pretender analisar vias gerais de tráfego de informações em uma dada conversação por meio da Análise de Redes Sociais, faz-se necessário estudar as redes de retweets, menções/respostas e citações; e de maneira individual, sob risco de generalizações distorcidas;

5) Apresenta evidências acerca da dinâmica informacional da conversação analisada onde não se observa conformações de “eco” em nenhuma das interações.

2. Metodologia

Como o objetivo geral dessa investigação é a análise das redes de interação dos usuários mais expostos à mis/desinformação/controvérsia sobre vacinas no âmbito da pandemia COVID-19, faz-se necessário identificá-la na conversação capturada. Contudo, os dados coletados perfazem quase 24 milhões de tweets, e em razão da inviabilidade de adotar qualquer procedimento manual, fez-se uso de técnicas da Mineração de Textos; uma sub-área da Inteligência Artificial que objetiva extrair informações e padrões em conjuntos de dados textuais (Tao; et al., 2020; Hotho; et al., 2005), a qual tem sido empregada, tal qual explicitado por Jung e Lee (2020), em áreas como Educação (Ferreira-Mello et al., 2019), Saúde (Tao; et al., 2020), Ciências Biológicas (Cohen & Hunter, 2008) e Ciências Sociais (Macanovic, 2022).

Assim, foram implementadas múltiplas abordagens nesta investigação, nomeadamente: análise textual *word-level* (por frequência de ocorrência de palavras), análise de associação entre palavras e sequência de palavras (similaridade textual) e técnicas avançadas de análise envolvendo classificação de textos usando aprendizado de máquina. A literatura que se debruça em desenvolver metodologias para detecção automatizada de mis/desinformação é extensa, e se intensificou sobremaneira no âmbito da pandemia COVID-19. Entretanto, a maioria dos recursos provenientes dessas contribuições estão em língua inglesa, como bancos de dados anotados e modelos de *machine learning* pré ou já treinados.

Além disso, a falta de consenso acerca do que seria de fato mis/desinformação e os desafios impostos pelas especificidades linguísticas, subjetivas e contextuais de sua produção e disseminação complexificam as tentativas de detecção automatizada desse tipo de conteúdo, sobretudo quando o lócus de análise são as plataformas de mídia social e o objetivo da classificação é qualificar entre “fato” ou “fake”.

Por essa razão, propomos uma abordagem mais ampla, na qual não se visa empreender um trabalho de detecção de mis/desinformação, mas capturar a conversão onde se tem maior *probabilidade* de ocorrência do que será aqui designado como “elementos associados à mis/desinformação” sobre vacinas, tal qual: “perigosa”, “experimental”, “veneno”, “cobaia”, “efeitos adversos”, “morreu após tomar”, “não vou me vacinar”, “não vou deixar [...] se vacinar”, “não funciona”, “feita às pressas”, “vachina”, “causa mortes”, “morreu após tomar vacina”, “causa AIDS”, “causa autismo”, “plano de controle”, “estratégia política”, “vacina do Doria”, “virar jacaré”, “marca da besta” etc; incluindo-se, portanto, os temas de controvérsia.

A classificação dos tweets do *corpus* de análise foi feita pelo modelo através de aprendizado dito supervisionado, fazendo-se necessário um conjunto de dados de treino já anotado; e a metodologia aqui proposta consiste em utilizar os dados efetivamente checados como falsos/boatos pelas principais agências de checagem brasileiras. Não obstante o bias associado ao uso de produção advinda da atividade jornalística como *ground truth* para o treino do modelo de *machine learning*, trabalhos anteriores têm apontado para a transparência da metodologia de checagem de conteúdo empregada por essas agências (aberta ao público) e sua potencialidade no combate à mis/desinformação (Massarani et al., 2021; Santos & Maurer, 2020).

Coleta de dados

Dos tweets

A etapa da coleta dos tweets se deu através da plataforma de análise de mídia social Netlytic. Através do descritor “vacina OR vacinação” e do período compreendido entre 11 de dezembro 2020 a 13 de novembro de 2021, foram coletados 23 milhões 956 mil e 385 tweets em sistema de coleta contínua a cada 15 minutos, até chegar ao quantitativo de 100 mil tweets, os quais foram salvos, e o processo reiniciado automaticamente.

Dos dados para treino do modelo de mis/desinformação e controvérsia

Define-se *Web Scraping* como sendo o procedimento de coleta automatizada de informações em *websites*. Para tal, foi utilizada a biblioteca Selenium, de forma a capturar as publicações de cinco agências de checagem, nomeadamente: Fato ou Fake (G1) (640 matérias), Boatos.org (787 matérias), Aos Fatos (287 matérias), Projeto Comprova (335 matérias) e Lupa (621 matérias), perfazendo um total de 2.670 matérias cujo conteúdo foi efetivamente checado e classificado como falso ou boato, que pós-filtragem daquelas estritamente relacionadas às vacinas contra COVID-19, tem-se 619.

Como o padrão das agências é a inclusão do conteúdo falso juntamente à sua classificação nos títulos das matérias, como em “#Verificamos: É falso que vacinas transformarão as pessoas em ‘antenas’ de sinal 5G”, foram retiradas as classificações (“#Verificamos: É falso que”) e deixados apenas os conteúdos falsos (“vacinas transformarão as pessoas em ‘antenas’ de sinal 5G”).

Levando em consideração que o estilo de escrita dos títulos por essas agências é mais orientado à variante culta da língua portuguesa, optou-se por selecionar tweets similares ao conteúdo checado usando similaridade de cossenos. Para que o computador possa compreender textos, faz-se necessário traduzir as palavras em representações matemáticas, nomeadamente, em suas representações vetoriais. Existem várias maneiras de fazê-lo, porém aquela adotada neste estudo diz respeito aos *embeddings* do BERT (*Bidirectional Encoder Representations from Transformers*), pois conseguem capturar o contexto de ocorrência das diferentes palavras que compõem um dado texto (Devlin et al., 2018).

Assim, os tweets (provenientes do *corpus*) que satisfizeram o *threshold* de 0.5 de similaridade entre as representações vetoriais dos tweets e do conteúdo checado falso/boato pelas agências de checagem foram selecionados para compor o estudo (12.893 tweets), perfazendo um total de 26.903 dados de treino.

Dos dados para treino do modelo para análise de sentimentos

Inserido no contexto das técnicas de mineração de opinião (LIU, 2020), a análise de sentimentos consiste em tarefa de classificação supervisionada através de modelos de *machine learning* ou em *Lexicon*. A classificação por sentimento foi implementada devido ao fenômeno da pós-verdade, em que crenças e emoções passam a ter mais valor que a factualidade (Amaral; Santos, 2019). Como será posteriormente apresentado, a inclusão dos sentimentos – juntamente ao texto dos tweets – melhoraram a performance do modelo de classificação em mis/desinformação ou controvérsia.

Em razão da escassez de *dataset* de treino em língua portuguesa fez-se uso do banco de dados proveniente da plataforma de programação Kaggle, disponibilizado para realização da competição de classificação intitulada *Sentiment Analysis of COVID-19 related Tweets*, lançada em 2021. Trata-se de um *dataset* contendo 5 mil tweets classificados em 11 sentimentos (traduzidos para português através da API do Google Tradutor). Aqueles escolhidos para treino do modelo foram adaptados e são: tristeza, pessimismo, piada, relatório oficial, gratidão, indignação (anteriormente raiva) e apreensão (anteriormente ansioso).

Para aumentar o tamanho desse conjunto de dados – e balanceá-lo -, fez-se um dicionário de termos associados à cada um dos sentimentos usando nuvem de palavras do conjunto de tweets assinalados, e a partir dos termos mais frequentes – e

exclusivos -, fez-se uma coleta de tweets relacionados usando a biblioteca Snsrape, posteriormente selecionados usando similaridade de cossenos seguindo o mesmo procedimento supramencionado (*threshold* de 0.4). No final, tem-se um *dataset* de treino contendo 15.862 tweets, 2.266 para cada sentimento.

Seleção, treino e validação dos modelos de machine learning

Para selecionar o modelo que melhor se adequaria aos dados de treino (mis/desinformação ou controvérsia), levou-se em consideração o *benchmarking* realizado por Khan et al. (2021), mostrando que modelos de aprendizado profundo baseados no BERT são os que apresentam melhor acurácia.

Entretanto, o modelo Naive Bayes (NB) apresentou resultados próximos, 0.93 contra 0.96 do RoBERTa; tendo sido, portanto, o de escolha nesta pesquisa pelo melhor *trade-off* entre performance e custo. Assim, o modelo Naive Bayes implementado foi o *Multinomial*, do *framework* Scikit Learn (sklearn), uma simplificação do teorema de Bayes, em que o *naive* representa a presunção de independência entre variáveis.

Para ambos os modelos, os atributos (dados de entrada para o treinamento do modelo) foram provenientes da vetorização das sentenças do *dataset* de treino através do cálculo do *Term Frequency-Inverse Document Frequency* (TF-IDF), implementado por meio do sklearn usando a regressão Lasso como mecanismo de normalização dos dados para redução do número de atributos do corpus - e, assim, do custo computacional -, para que fosse possível fazer uso tanto de palavras quanto de sentenças de duas palavras de maior relevância no *dataset* de treino como *input* para o modelo.

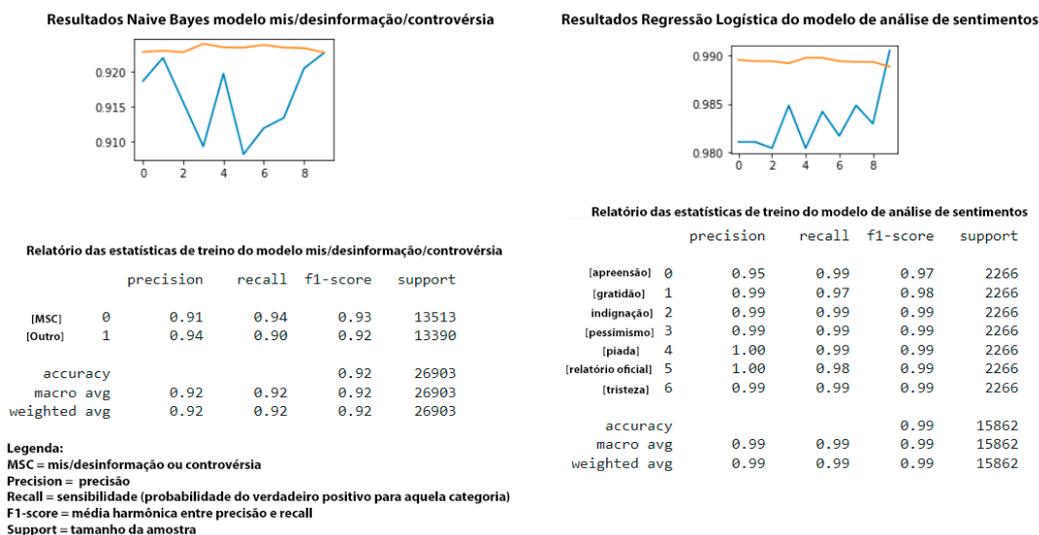
A seleção de hiperparâmetros foi feita através de *grid-search*; e a validação dos modelos por meio de validação cruzada estratificada, com 10 *folds*, tal qual evidenciado na literatura como sendo o procedimento de validação mais adequado não só quando o *dataset* de treino possui uma quantidade não tão compreensiva de dados, como também apresenta mecanismo capaz de superar limitações relacionadas à divisão do *dataset* em partições de treino e teste (como saber se o percentual escolhido para tal é o mais apropriado? Ou se os dados foram repartidos de forma que sejam suas melhores representações?) (Géron, 2017, p. 54).

Isso se dá, porque por meio da validação cruzada o conjunto de dados de treino é repartido em uma dada quantidade (k) de partes (*folds*), que funcionarão alternativamente como dados de validação (não fazendo-se mais necessário, portanto, um conjunto de dados específico para esse fim), e finalmente, testado em partição específica (dados ainda não vistos pelo modelo) para cálculo de estatísticas como a acurácia.

O melhor modelo para classificação em mis/desinformação ou controvérsia foi aquele com alfa de 0.1, com o qual chegou-se a uma acurácia de 0.92 (0.86 sem os sentimentos como *input*) nos dados de teste e treino (0.007 de diferença); enquanto para a análise de sentimentos, foi a Regressão Logística (C= 5, *max_iter* = 100, *penalty* = 12, *solver* = saga), com 0.98 de acurácia (contra 0.96 do *Naive Bayes*, também testado).

A análise do treinamento foi feita com base nos valores comparados de acurácia nos dados de treino e teste, dispostos na Figura 1. Quando o modelo de *machine learning* se ajusta demasiado nos dados de treino (*overfitting*), ele acaba tendo uma performance inferior ao classificar dados ainda não vistos (dados de teste). Como é possível perceber nesta figura, houve convergência dos modelos (nos gráficos, as acurácias de teste ao longo do treinamento atingiram aquelas de treino no final do processo) e ambos apresentaram capacidade ótima de generalização (baixa diferença entre performance treino-teste). Além disso, são apresentadas também as estatísticas *Recall*, precisão e *F1-score* cujo cálculo foi implementado pelo módulo *Classification Report* do sklearn.

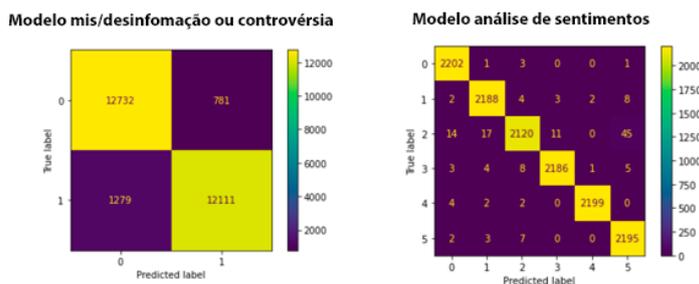
Figura 1 – Resultados da validação cruzada e relatórios contendo as estatísticas do treinamento dos modelos de machine learning.



Fonte: Elaborada pelos autores.

Os resultados da matriz de confusão possibilitam a análise dos erros de classificação do modelo com base nas predições feitas nos dados de teste após o treinamento. Privilegiou-se maior especificidade às ocorrências potencialmente falsas, pois deixar de analisá-las implicaria em maior prejuízo analítico do que a presença de algum tweet com pouco potencial à mis/desinformação no corpus de análise, e como podem ser observados na Figura 2, espera-se que aproximadamente 6% de conteúdo “potencial/controverso” (0) seja classificado erroneamente como “outro” (1); não sendo, portanto, analisado.

Figura 2 – Matriz de confusão dos modelos de machine learning.



Fonte: Elaborada pelos autores.

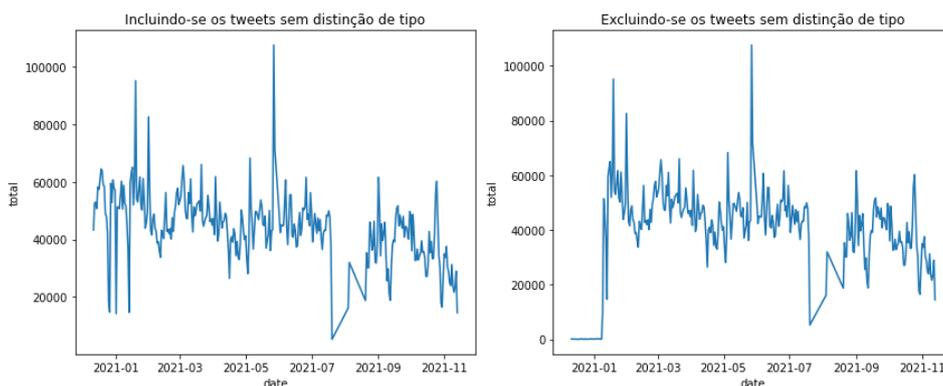
Acerca do resultado das etapas anteriores, dos quase 24 milhões de tweets coletados, 12 milhões 462 mil e 390 tweets foram assinalados como potenciais à maior exposição à mis/desinformação ou controvérsia acerca das vacinas contra COVID-19 (52%), compartilhados/produzidos por 2.400.908 milhões de usuários únicos, os quais estiveram, portanto, em exposição direta; tornando-se o *dataset* de base para a investigação.

A proporção de conteúdo mis/desinformativo ou controverso na conversação coaduna com o caráter polarizado da conversação sobre vacinas no Twitter, como apresentados nos trabalhos anteriores. Para que fosse possível responder às perguntas norteadoras da pesquisa, optou-se por selecionar uma coorte de acompanhamento.

Seleção da coorte

Dessa forma, para operacionalizar essa seleção, a linha de tendência da atividade de conversação no Twitter advinda da amostra obtida pós-classificação (12.4 milhões) foi analisada, e se deu tal qual ilustrado na Figura 3. Como é possível observar, foram identificados 1.517.258 milhões de tweets com dados faltantes no que se refere ao tipo de interação (metadado *tweet_type* presente no corpus) - o que inviabilizaria responder à PPI -, sendo, portanto, excluídos.

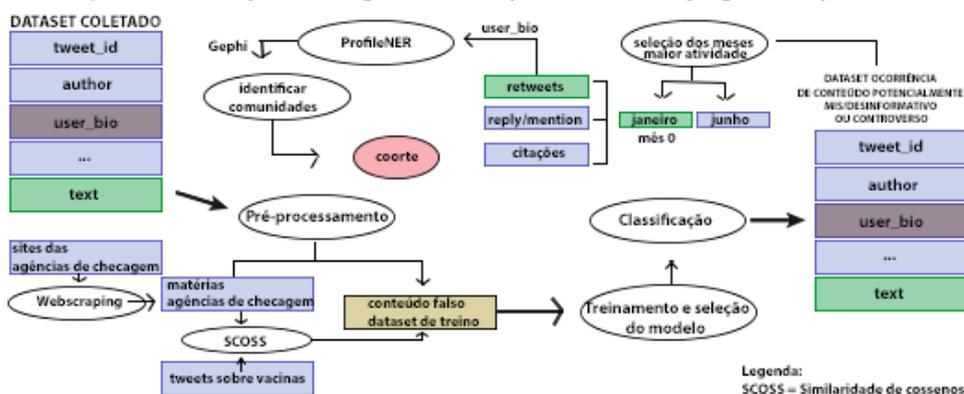
Figura 3 - Linha de tendência da atividade na conversação classificada enquanto sendo mis/desinformação/controvérsia.



Fonte: Elaborada pelos autores.

Tal qual indicado na Figura 3, os meses de janeiro e junho de 2021 foram aqueles com maior intensidade de atividade no Twitter concernindo o tema vacina em meio a conteúdo potencialmente mis/desinformativo/controverso. Como janeiro marca o início da série temporal, ele foi considerado o mês 0, e usado para selecionar a coorte, composta inicialmente por um total de 331.158 usuários únicos que participaram da conversação através do *retweeting*, conforme procedimento ilustrado na Figura 4.

Figura 4 - Fluxograma da operacionalização da metodologia para seleção da coorte de análise.



Fonte: Elaborada pelos autores.

Mas, porque o *retweeting*? Essa escolha foi feita em função da sua dimensão de influência e referência (Recuero & Zago, 2021). Ou seja, quando um usuário reproduz (retweet) o tweet de outro, ele o faz, pois: (i) considera que aquele conteúdo é interessante para ser exposto em sua rede social, (ii) acredita que o usuário-autor seja relevante, ou (iii) concorda com o conteúdo do tweet.

Para responder às perguntas de pesquisa faz-se necessário a utilização de uma metodologia que possibilite analisar a

estrutura da rede social, quer dizer, a conformação topológica das diferentes redes entre os usuários partícipes da conversação capturada em função das diferentes maneiras de interagir no Twitter, ou seja: retweets, respostas/menções e citações. Para tal, fez-se uso do *software* Gephi, um programa para Análise de Redes Sociais (ARS).

A ARS é baseada na Sociometria e na Teoria dos Grafos, e com ela, é possível realizar análises da estrutura social de diferentes grupos, isto é, da disposição dos indivíduos (nós) e suas interações (arestas) (Recuero, 2017, p. 9). Uma das métricas passíveis de serem mensuradas nessa abordagem é a modularidade, que possibilita a detecção de comunidades na rede usando as conexões como medida de correspondência entre nós (Recuero, 2017, p. 43).

Assim, uma vez selecionada a conversação com maior prevalência de conteúdo mis/desinformativo/controverso acerca das vacinas, os meses de maior intensidade de atividade nesta conversação, e os usuários que retuitaram no mês de janeiro (mês 0), a rede de interações desses usuários (*retweeting*) foi analisada e as comunidades nas quais estiveram alocados foram detectadas usando o Gephi, através do cálculo da modularidade, como supracitado.

Como resultado, 5 comunidades foram mapeadas, nomeadamente: Grupo A (GA, laranja; 41,24% dos nós, 49,18% das arestas), Grupo B (GB, verde; 38,86% dos nós, 22,48% das arestas), Grupo C (GC, vermelho-salmão; 12,49% dos nós, 20,69% das arestas), e Grupo D (GD, azul, 7,41% dos nós e 7,65% das arestas), ilustrados na Figura 5a; assim como se visualiza as arestas (linhas de conexão), e os nós (cada um dos 331.158 usuários), posicionados enquanto aqueles que retuitaram (grau de saída) e os que foram retuidados (grau de entrada, nós-alvo). A comunidade GD foi desconsiderada pela baixa representação destes no número de nós e arestas, restando 304.809 usuários na coorte final.

Figura 5 - Representações topográficas de todas as interações no mês de janeiro de 2021.



Fonte: Elaborada pelos autores.

Caracterização da coorte

Para tal, foi usado o ProfileNER-classifier, um algoritmo que, por meio de expressões regulares, busca na biografia feita pelos usuários na plataforma (*user_bio*), termos associados a certas categorias, a saber: Outros (não classificados), Profissionais da Saúde (enfermeiro(a), médico(a), etc.), *Scientific Background* (mestrado, doutorado, divulgador(a) científico(a), etc.), Jornalista (repórter, jornalista, colunista, etc.), Veículos de informação (jornal, portal de notícias, revista, etc.), Profissionais da Educação (professor(a) de física, biologia, etc.), Organização (instituto, ministério, secretaria, hospital, etc.), Políticos (prefeito(a), governador(a), senador(a), etc.) e Conservadores (que se nomeiam como tal).

Análise de Redes Sociais para os diferentes tipos de interação no Twitter

A amostra das interações da coorte foi dividida com base no tipo de interação (metadado disponível no *dataset*), um para retweets, outro para respostas/menções, e outro para citações, de forma que cada análise topológica da rede fosse

individual (ARS independentes para cada interação); procedimento replicado para o mês de junho, aquele em que a coorte foi reconstituída (os mesmos usuários analisados em janeiro) e suas interações estudadas: quantos remanesceram na conversação? Houve deslocamentos? Pessoas que outrora estavam alocadas no Grupo C passaram para outros grupos? E aquelas susceptíveis, mas em comunidades onde se tem maior efeito protetor (GA e GB), foram cooptadas pelo Grupo C? A dinâmica informacional ordenadora das interações em janeiro, manteve sua natureza em junho, 5 meses depois?

Análise textual *word-level*

Para responder às PP, faz-se necessário a análise das ocorrências textuais, e a ARS por si só não dá conta da compreensão dessa dimensão, fazendo-se necessária a adoção complementar de abordagens qualitativas voltadas ao conteúdo daquilo que circula na conversação. Neste trabalho, o procedimento adotado foi a manipulação de *DataFrames* (dados tabulados em estrutura semelhante às planilhas) em linguagem de programação Python, através da biblioteca Pandas, para verificação de circulação (presença) de conteúdo “estrangeiro” (tweets e links que não foram produzidos por nós pertencentes à dada comunidade) nos diferentes grupos. Além disso, fez-se uso do *bigram*, por meio da biblioteca *Natural Language Toolkit*, técnica baseada na contagem da co-ocorrência de duas palavras num dado texto, tal qual apresentado por Suen (1979), implementada para análise comparativa e exploratória dos temas latentes abordados por cada um dos grupos analisados na coorte.

3. Resultados

Destarte, a Tabela 1 mostra a relação de categorias de perfil dos nós que compõem a coorte por comunidade, a partir da qual faz-se possível perceber que aquela com maior incidência de pesquisadores(as) (incluindo-se divulgadores científicos em saúde como Ethel Maciel e Atila Lamarino) é a de usuários do Grupo A, assim como de profissionais da saúde, jornalistas e veículos de informação. Destaca-se na comunidade de usuários do Grupo B a maior frequência de profissionais da saúde, artistas e profissionais da educação; e no Grupo C, de conservadores.

Tabela 1 - Caracterização da coorte em função da presença relativa de categorias de perfis por comunidade.

Categorias*	GA	GA-alvo***	GB	GB-alvo	GC	GC-alvo
Outros	120.831	5.090	124.055	1.202	33.976	1.803
Prof. Saúde	0,28	0,16	0,50	0,15	0,07	0,06
Pesquisador (a)	0,25	0,22	0,13	0,19	0,02	0,05
Jornalista	0,25	0,25	0,18	0,27	0,03	0,10
Veículos de informação	0,21	0,23	0,21	0,25	0,05	0,12
Artistas	0,11	0,07	0,27	0,12	0,01	0,01
Prof. Educação	0,09	0,06	0,13	0,04	0,02	0,02
Organização	0,05	0,04	0,02	0,03	0,01	0,01
Políticos **	0,03	0,08	0,01	0,09	0,01	0,03
Conservadores** (as)	0,02	0,05	0,05	0,04	0,84	0,66
Total	129.676	6.133	126.210	1.430	39.063	2.297

* Categoria “Outros” e “Total” apresentados em valores absolutos. Os valores relativos dos demais são aqueles em que se desconsidera, no denominador, o valor correspondente aos “Outros” do cálculo; ou seja, representam percentualmente a presença daquelas categorias quando foi possível atribuir pelo menos uma delas aos usuários.

** Baixo percentual de políticos é consequente da tendência daqueles enquadrados nesse perfil de se designarem, também, conservadores.

*** GA-alvo, GB-alvo e GC-alvo referem-se aos nós-alvo, ou seja, aos usuários autores do conteúdo que foi reproduzido por aqueles pertencentes a cada uma dessas comunidades.

Fonte: Elaborada pelos autores.

Já no que concerne aos nós-alvo, ou seja, àqueles que tiveram seus conteúdos retuitados, evidencia-se a relevância dos veículos de informação e jornalistas enquanto receptores nas comunidades GA, GB, e até mesmo no Grupo C, uma vez que proporcionalmente, corresponderam a 22% das fontes responsáveis pela produção ou compartilhamento do conteúdo que circulou intragrupo.

Acerca das interações inter/intragrupo, embora a maior parte das interações do GC (maior prevalência de conservadores) tenha sido intragrupo (4.360, 88%), 454 usuários da comunidade GA foram retuitados por nós-GC; e 76 da GB. Da interface entre GA e GC tem-se um contingente de 1.487 tweets, provenientes do primeiro, que entraram no segundo. Ao analisar as categorias de perfil (não-outros) dos usuários da GA com os quais aqueles do GC interagiram, tem-se: veículos de informação (825), políticos (155), profissionais da saúde (95), jornalistas (91), educadores (9), pesquisadores (43), organizações (4), artistas (2) e políticos (2); sendo os mais proeminentes a Gazeta do Povo (207), o ator Bruno Gagliasso (81), o oncologista e pesquisador Bruno Filardi (69), o jornalista Claudio Dantas (45) e o veículo CNN Brasil (31).

Além disso, foi também realizada uma comparação literal entre tweets circulantes em GA, GB e GC através da checagem da presença de ocorrências de conteúdo textual idêntico nas diferentes comunidades. Os achados indicam que 37.5% dos tweets que circularam em GA (razão entre número de tweets idênticos e total de tweets em GA), circularam no GC, 60% em GB. Em contrapartida, 63% do conteúdo que circulou no Grupo C (como numerador, o número de tweets iguais; como denominador, número total de tweets em GC), entrou em GA, e 32%, em GB. Com relação à última, 93% do seu conteúdo esteve em GA, e 71%, em GC.

Em razão da relevante presença de tweets idênticos àqueles em GB, em GA e GC, foi realizada a análise de *bigrams* (frequência de co-ocorrência dupla de palavras) no conjunto de tweets produzidos e circulantes entre usuários do Grupo A e Grupo C. Como resultado, tem-se os seis mais relevantes apresentados na Tabela 2; os quais apontam para a presença de elementos semânticos compartilhados entre tweets circulantes em ambas as comunidades.

Tabela 2 - Relação de *bigrams* representativos dos tweets circulando no Grupo C e de usuários do Grupo A.

Comunidade	<i>Bigrams</i>	Total
GC	(tratamento, precoce)	260
	(governo, federal)	166
	(efeitos, colaterais)	165
	(vai, tomar)	91
	(milhões, doses)	89
	(uso, emergencial)	64
GA	(governo, federal)	646
	(milhões, doses)	371
	(governo, bolsonaro)	370
	(leite, condensado)	331
	(auxílio, emergencial)	321
	(vai, tomar)	276

Fonte: Elaborada pelos autores.

Embora não descritos na tabela, (tratamento, precoce), (*fake, news*), (uso, emergencial), (jair, bolsonaro), (leite, condensado), (vai, ter), (joão, doria) e (ministério, saúde) foram outros *bigrams* compartilhados. Entretanto, as condições de ocorrência de alguns termos foram particulares à cada uma das comunidades. Tratamento precoce, por exemplo, foi um tema objeto de diferentes enquadramentos quando em GA ou GC, onde no primeiro tem-se como exemplo: “Tratamento precoce: Governo Bolsonaro gasta quase R\$ 90 milhões com remédios ineficazes, mas ainda não pagou Butantan por vacinas”, já no segundo: “Tratamento precoce não vale. Mas vacina precoce vale?” (Alexandre Garcia)”.

No que concerne aos links, 21% dos que estiveram no Grupo C também estiveram em GA. Dentre os mais reproduzidos intragrupo-GC, aquele mais relevante em circulação pode ser observado na Figura 6 - uma peça de comunicação publicada pela página no Instagram da Prefeitura de José dos Campos que se propõe a tirar dúvidas sobre a vacina contra a COVID-19 (187). Em seguida, tem-se o link que direciona para uma publicação no Facebook contendo dois vídeos nos quais uma pessoa apresenta espasmos musculares generalizados, descritos como “efeitos colaterais das vacinas” (168), para além do link para o grupo do Telegram do senador Flávio Bolsonaro (161).

Figura 6 - link com maior número de reproduções em circulação no Grupo C em janeiro de 2021.

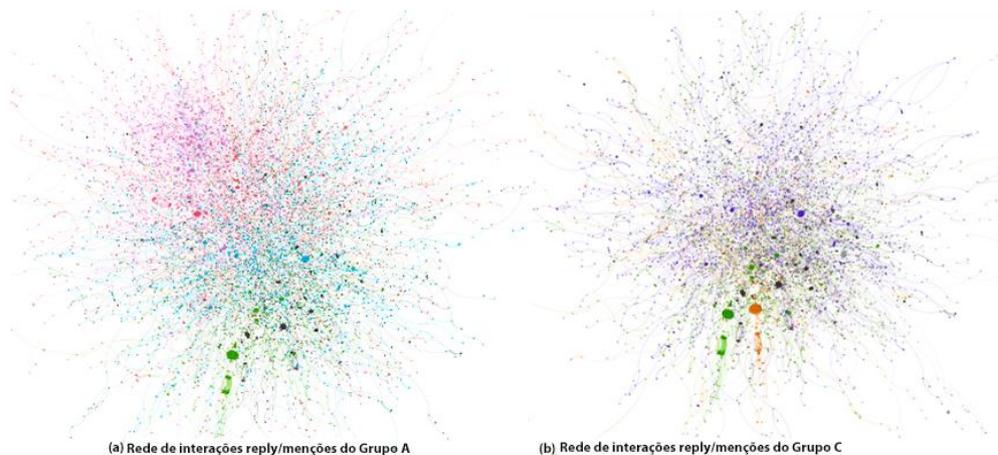


Fonte: Página oficial da prefeitura de São José dos Campos. Disponível em: <https://www.instagram.com/prefeiturasjcamposoficial/>

Agora, no que tange às menções - ilustrada na Figura 5b -, a rede foi composta por 109.970 nós e 124.639 arestas. Tem-se que 31% dos usuários que retuitaram, mencionaram/responderam; em que 8.648 dos nós que constituem o GC também interagiram na plataforma por meio das respostas, contra 20.204 dos GA, e 3.543 dos GB. Em termos relativos, os usuários localizados no Grupo C foram aqueles mais ativos nas menções/respostas, tendo 20% dos seus nós; enquanto GA e GB, 14% e 2%, respectivamente.

Ao analisar os possíveis deslocamentos concernindo as comunidades das atividades de *retweeting* (referência) e menção/resposta, os achados indicam uma tendência de descontinuidade e dispersão sistemática dos grupos, ao passo que novas conexões foram formadas. Isso pode ser percebido tanto através da análise topológica da configuração da rede apresentada na Figura 7, como da correlação entre nós-comunidades, em que apenas 2.9% dos usuários do Grupo C permaneceram em interlocução com atores da extrema direita.

Figura 7 - Representação topográfica das redes de interação menções/respostas particulares à comunidade dos usuários do Grupo A (esq.), e do Grupo C (dir.), em janeiro de 2021.



Fonte: Elaborada pelos autores.

Ainda acerca do caráter das interações, percebe-se que os usuários onde na configuração de referência (retweets) estavam no Grupo C, nas menções/respostas, estabeleceram diálogos entre si em novas e menores conformações comunitárias. Além disso, houve 87% de compartilhamento de comunidades entre GA e GC, onde dentre as 77 comunidades formadas por pelo menos dois nós-GC, em 67, algum nó-GA também esteve.

A interação externa pôde também ser percebida através das relações de nós-alvo, conforme apresentadas na Tabela 3. Evidencia-se, dessa forma, uma abertura maior – ou mesmo preferencial -, para nós-estrangeiros à cada uma das comunidades. Com efeito, tem-se que a comunidade GA respondeu ou mencionou tweets de Jair Bolsonaro (influyente nó do Grupo C) e outros agentes da direita política brasileira; enquanto a comunidade GC, de atores alvo de críticas no seio do seu grupo - como João Doria -, de veículos de informação como a CNN Brasil (mas também, Folha, Estadão e G1) e outros nós não-nativos do grupo, como o Instituto Butantan.

Tabela 3 - Relação de nós-alvo das interações menções/respostas dos nós-GA e nós-GC, em janeiro de 2021.

Comunidade	Nós-alvo	Total
GA	jairbolsonaro	490
	folha	323
	o_antagonista	301
	UOLNoticias	294
	JanainaDoBrasil	246
	g1	245
GC	jdoriajr	329
	claudioedantas	234
	CNNBrasil	159
	LucianoHuck	128
	SigaGazetaBR	123
	revistaoste	109

Fonte: Elaborada pelos autores.

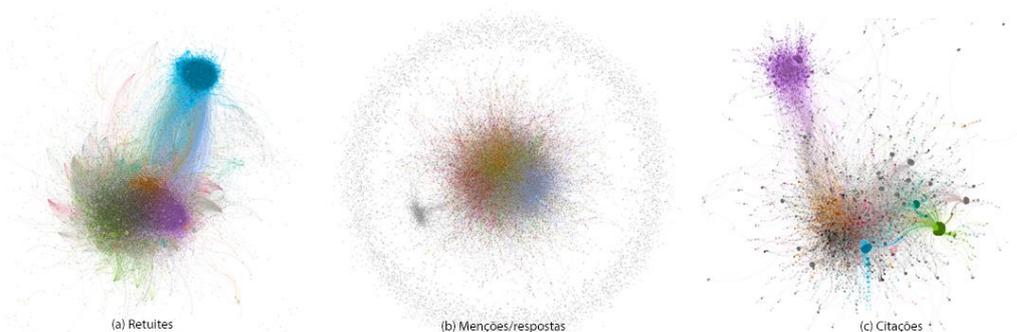
Acerca das citações, houve 40.733 nós e 46.718 arestas, onde 58.5% dos usuários que retuitaram, também citaram. Com relação às comunidades, 11% do Grupo C participou da atividade de *quoting*, 10,6% dos GA, e 3,5% dos GB; além de uma perceptível tendência à aparição de pequenos conglomerados intracomunitários, ilustrados na Figura 5c. Das interações, GC as fez em 77% com a comunidade de maior prevalência de políticos e agentes da direita política, enquanto 23% com outros. Das 26 comunidades em que pelo menos 1 nó-GC se fez presente, em 19, algum nó-GA também esteve; correspondendo a 73% de interações intercomunitárias.

No que se refere aos elementos semânticos circulantes, tem-se que a maior parte das citações feitas pelo Grupo C foram relacionadas ao tópico da eficácia e segurança da CoronaVac: (eficácia, segurança), (hospitalizado, segundo), (estudados, exemplo), (número, idosos); sendo o tweet mais relevante: “Olha aí o Butantã dizendo que a vacina dá “100% de chance de não ser hospitalizado”. O que é isso? Para quem? Segundo a Anvisa o número de idosos estudados por exemplo é insuficiente para qualquer garantia de eficácia e segurança. Tem al”.

Já no que concerne às interações advindas do mês de junho - ilustradas na Figura 8 -, com relação às interações-retweets (referência) da coorte, tem-se que 27% dos nós ativos em janeiro se fizeram presentes neste mês. Em termos relativos, 48% daqueles pertencentes ao GC estiveram nesta conversação, 39% dos GA e 18% dos GB. Ao observar as composições comunitárias, 95% dos nós-GC continuaram no Grupo C (azul), mas os outros 5% (642 usuários) saíram do grupo e passaram a compor as comunidades com maior prevalência da mídia tradicional, de profissionais da saúde, de divulgadores científicos, políticos da esquerda e centro, e influenciadores digitais.

Para além disso, observou-se também um sutil aumento percentual acerca da representação relativa da comunidade GC com relação às demais, passando de 12.49%, para 12.85% dos nós; e 20.69% das arestas para 22.93%. Acerca dos nós GA e GB, tem-se uma maior dispersão de grupos (maior número de comunidades), mas que ainda possuem forte coesão em termos de interações.

Figura 8 - Representações topográficas de todas as interações no mês de junho de 2021.



Fonte: Elaborada pelos autores.

Entretanto, 0,7% dos usuários do Grupo A (387) e 0,9% dos usuários do Grupo B (206) passaram a fazer parte do GC. Além disso, 80% das interações no Grupo C foram intergrupo (redução de 8 pontos percentuais); enquanto em GA e GB, 12% e 17% dos nós-alvos foram aqueles pertencentes ao GC, respectivamente.

A nível de conteúdo textual idêntico circulante nas comunidades GA, GB e GC, tem-se que 39% (1,5% a mais/jan.) dos tweets em GA estiveram em GC *ipse litteris*; 67% (7% a mais/jan.) em GB. Em contrapartida, 62% (1% a menos/jan.) do conteúdo textual em GC esteve em GA; 44,5% (12,5% a mais/jan.) em GB. Acerca dos usuários do Grupo B, 95,3% (2,3% a mais/jan.) dos tweets fizeram-se presentes em GA; 70,3% (0,7% a menos/jan.) em GC.

Como exemplo desse compartilhamento, tem-se o tweet publicado por nós-GA, mas circulante no Grupo C em junho

de 2021: “Gente vacina sempre deu reação adversa. Tétano dor a semana toda, BCG te presenteia com uma cicatriz, Febre amarela e gripe dão febre. Todo mundo tomava vacina e nem sabia o laboratório que era, só sabia da importância vacinal. Onde a gente se perdeu? Sommelieização gourmet??”, com 1.667 reproduções.

Em contrapartida, como representação do conteúdo produzido no Grupo C, mas que entrou na comunidade GA em junho de 2021, tem-se: “Ao se vacinar, Thais Possati teve que assinar isto: “Ainda não há estudos em gestantes/lactantes que assegurem seu uso, ou seja, não foi verificada evidência científica de segurança e eficácia”. Após a morte de Thais, a Anvisa vetou essa vacina p/ grávidas. VACINAS EXPERIMENTAIS”, com 1.553 retweets.

No que concerne aos principais links em circulação, 17% foram de comum circulação entre GA e GC, 6% GB em GC; e aquele com maior número de reproduções (389) no Grupo C foi um vídeo do YouTube de um clipe de Eric Clapton, o qual foi acionado no contexto do seguinte tweet:

"Onde foram parar os rebeldes?" é a pergunta certa de Eric Clapton, um dos poucos artistas a denunciar a tirania disfarçada de proteção sanitária c/ seus lobbies de passaporte sanitário e lockdown. Eric teve reações "desastrosas" à vacina. <https://t.co/HIL5tEZMhB> via @YouTube

Este foi seguido pela matéria publicada em site retirado do ar, com domínio Porão da Mamãe, intitulada “Depois de acusar Bolsonaro de fazer pressão para comprar vacina da Índia, jornalistas acusam Bolsonaro de não fazer pressão para comprar vacina da Índia” (204). Em seguida, tem-se um abaixo assinado em defesa do jornalista Allan dos Santos, co-fundador do veículo hiperpartidário Terça Livre (192); e por fim, um link que redireciona para a página do produto SM-102 da empresa Cayman Chemical (180), acompanhado por outro da agência de checagem Reuters, intitulada “Checagem dos fatos - lipídio SM-102 na vacina Moderna não é perigoso” (180, tradução nossa).

Agora, acerca das menções/respostas, 28.091 (18,5%) usuários-referência utilizaram-na para interagir em junho. Destes, 18% são do Grupo C, 12% dos usuários do Grupo A e 2,7% dos usuários do Grupo B. Houve um aumento da permanência de nós-GC na comunidade onde fazem-se presentes atores da direita política, indo de 2,9% para 20%; apesar da maioria ter se dispersado em outras comunidades, tal qual a tendência observada em janeiro.

Destaca-se também a fragmentação referente aos usuários pertencentes ao grupo GA, e tem-se como resultado um aumento a nível de compartilhamento comunitário, em que nós-GA e nós-GC estiveram ambos alocados em 98,5% das comunidades das quais fizeram parte, em oposição aos 87% anteriores. A presença de nós estrangeiros como nós-alvo de interação via menção/resposta também foi observado no mês de junho, conforme Tabela 4.

Tabela 4 - Relação de nós-alvo das interações menções/respostas dos nós-GA e nós-GC, em junho de 2021

Comunidade	Nós-alvo	Total
GA	jairbolsonaro	223
	folha	158
	UOLNotícias	153
	fabiofaria	150
	BolsonaroSP	145
	mqueiroga2	139
GC	SenadoFederal	176
	CNNBrasil	133
	OmarAzizSenador	133
	AlessandroLoio2	126
	jairbolsonaro	116
	jdoriajr	100

Fonte: Elaborada pelos autores.

No que tange às citações, 59% dos nós-referência de janeiro também interagiram por meio delas em junho. Relativo à cada comunidade, o Grupo C se manteve em um mesmo percentual, 11,5% de atividade por meio das citações, enquanto usuários do Grupo A e B tiveram uma redução; de 10,6% para 8,9% nos primeiros, e de 3,5% para 2,3% para GB. Além disso, 79,5% das interações foram intragrupo, mas das 30 comunidades nas quais um nó-GC esteve, em 23 havia pelo menos um nó-GA (77%).

4. Discussão

Assim, a começar pelos resultados da análise da atividade de *retweeting*, observou-se uma conformação topológica polarizada na rede de interações. Embora GA e GB estivessem próximas, ambas estiveram relativamente distantes do Grupo C, tanto em janeiro, quanto em junho de 2021; o que poderia ser indicativo do efeito da câmara de eco. Porém, existe uma densa região de arestas, ou zona de intercâmbio, entre GA e GC; um canal de fluxo de informações cuja existência já coloca em questão a hipótese da relevância do ordenamento causado pelas câmaras de eco na estrutura de rede da conversação.

Isso se confirma ao se constatar que o conteúdo de usuários provenientes de GA (principalmente veículos de informação) foi reproduzido por GC; mas também, e sobretudo, pois uma quantidade considerável dos tweets que circularam em GA e GB, também circularam em GC. Embora se tenha encontrado resultados semelhantes à Recuero et al. (2021) no que se refere à relevante presença de veículos apócrifos hiperpartidários na conversação – pois são aqueles dentre os mais proeminentes em circulação em GC –, os achados não possibilitam afirmar que links possuindo conteúdo factível não conseguiram entrar no Grupo C.

Em janeiro, o link com maior número de reproduções no seio do GC foi uma peça de comunicação voltada à resolução de dúvidas concernindo as vacinas; em junho, o tweet contendo um link do produto SM-102 foi seguido por outro da agência de checagem Reuters que justamente apontava para a circulação desse link enquanto promotor de mis/desinformação, uma vez que faz passar a suposição de que esse produto é perigoso – quando na verdade o alerta do site faz referência a outra substância - e de que ele é componente da vacina Moderna – quando não o é.

Juntamente aos pontos supramencionados, a análise dos *bigrams* evidencia a participação ativa da função de *gatekeeping* (Recuero & Zago, 2012) dos retweets na plataforma. Embora se perceba, em ambos os meses, empreendimentos

de (re)enquadramentos de informações quando na existência de influxo do contraditório, identificou-se também a presença de significativa reprodução de conteúdo que faz direta oposição à crença hegemônica esperada no Grupo C, onde estão localizados os políticos da extrema direita; indicando que possivelmente seja interessante para a comunidade “informar” seus pares sobre as antíteses às suas crenças, produzidas e circulantes em outras comunidades.

Isso coaduna com a compreensão de que, se o Twitter enquanto plataforma assemelha-se à esfera pública (Dahlgreen, 2009 apud Amaral & Santos, 2020), e suas dependências configuram-se enquanto campo de disputa por legitimação discursiva (Recuero & Soares, 2021), faz-se necessário que haja não só reconhecimento da existência dos polos antagônicos (Shore et al., 2018), mas interação direta entre eles.

Com efeito, os achados apontam que essa interação se dá sobretudo através das menções-respostas; as quais são, de acordo à documentação do Twitter, as únicas mediadas por sistema de notificação. Ou seja, quando um usuário responde/menciona a um tweet/usuário estrangeiro à sua comunidade - pois motivado por intenção contestatória -, essa interação possui certa probabilidade de aparecer na *timeline* das pessoas que o seguem, as quais passam também a ser expostas à conteúdo potencialmente divergente; e ambos, aquele quem respondeu/mencionou e o alvo, passam a ser notificados dessa interação por e-mail e/ou na própria plataforma.

As respostas/menções se mostraram ser, dessa forma, importantes vias de tráfego de conteúdo externo ao Grupo C, pois (re)configuraram a estrutura da rede de tal forma que os nós que se localizam próximos aos políticos da extrema direita, passam a fazer partes de outras comunidades, a grande maioria delas compartilhadas por indivíduos pertencentes ao GA. Essa dinâmica também foi identificada na atividade de citação, ainda que se tenha uma maior predominância de interação homofílica.

Além disso, o que os achados dessa investigação apontam é que ao se objetivar analisar as vias de tráfego de informações no Twitter através da Análise de Redes Sociais, faz-se necessário estudar as diferentes interações de maneira individualizada, pois ao tomar apenas os retweets como parâmetro, tem-se uma maior probabilidade de generalização da dinâmica informacional apenas a partir desse referencial: polarizado, homofílico e com características relacionadas às câmaras de eco.

Mønsted e Lehmann (2022), por exemplo, analisaram uma rede de interações provenientes dos retweets e das menções, só que de maneira agregada. O problema dessa união é que os retweets são, geralmente, aqueles mais frequentes nas conversações, e acabam por distorcer a topologia da rede de interações quando presente. Isso se dá, pois a estrutura da rede é usualmente representada por *layouts* do tipo força de repulsão/atração, como a *Forceatlas 2*, em que a polarização característica dos retweets fará com que os nós se distanciem mesmo quando, na realidade, deveriam estar próximos em função da interação do tipo menção.

Finalmente, observou-se que a sistemática politização associada às vacinas contra COVID-19 amplificou a homofilia e as polarizações assimétrica e afetiva; fenômenos que se fizeram presentes na dinâmica informacional da conversação analisada, uma vez que se observa o Grupo C sempre em oposição aos demais quando nas topologias advindas dos retweets e citações. Contudo, a relevância conferida aos veículos de informação e de jornalistas enquanto fontes – seja de maneira positiva, ou negativa -, evidenciam também a importância da *misinformation*, sobretudo no que concerne aos primeiros.

A recepção à peça de comunicação da prefeitura de São José dos Campos, traduz, inclusive, aquela observada no Twitter; e evidencia que a susceptibilidade, a disseminação e o convencimento relacionados à mis/desinformação compõem um fenômeno complexo e multidimensional. Contudo, os achados sugerem que este parece ser mais da ordem sociocultural e subjetiva (percepção do risco) e socioeconômica, do que meramente informacional (ausência ou acesso dificultado à informação qualificada em função da suposta câmara de eco ou filtro-bolha).

Nesse sentido, a atividade motivada por partidarismo - como aquelas de defesa ao governo federal, e especialmente, à

Jair Bolsonaro -, é de destacada relevância na contribuição para sentimentos hesitantes com relação às vacinas contra COVID-19. Entretanto, sublinha-se que a hesitação e a polarização também se apresentam conseqüentes da falta de uma política nacional de comunicação institucional acerca de temas relacionados à pandemia, fazendo que com órgãos oficiais, ou mesmo entidades públicas, publicassem informações conflitantes; as quais somente alimentavam as já existentes preocupações, impactando na percepção de risco na medida em que suscitavam – e evidenciavam -, mais incertezas que consensos; sobretudo no que compete às medidas não-farmacológicas e à cobertura acerca do processo de desenvolvimento dos imunizantes contra a COVID-19.

5. Considerações Finais

Destarte, a partir das evidências apresentadas, aponta-se que as configurações das redes que constituem o ecossistema social existente no Twitter variam sobremaneira a depender da ferramenta de interação usada para (re)construí-las. Além disso, ao analisar as implicações dessas diferenças nos possíveis trânsitos dentro da estrutura social da plataforma, os resultados não indicaram ordenações decorrentes de conformações de “eco”; uma vez que se identificou um relevante número de tweets e links idênticos em circulação nas diferentes comunidades – incluindo-se aqueles com conteúdo contrário à crença hegemônica intragrupo do Grupo C -, e sistemáticas interações diretas entre nós GA, GB e GC.

Dessa forma, sugere-se aqui que deslocamentos intercomunitários não somente são possíveis, como fazem parte da dinâmica de interações no Twitter; e que existem vias de tráfego de informações através das quais faz-se possível a existência de conteúdo comum compartilhado em diferentes grupos. Para efeito da análise de redes, faz-se uma fotografia das interações num dado tempo e espaço, mas essas relações são dinâmicas; quer dizer, respostas-menções, citações e retweets co-ocorrem de forma que ora se privilegia interações intragrupo, ora intergrupo; ora polarizada, ora imbricada tal qual a topologia observada nas menções-respostas.

Nesse interim, abrem-se brechas para possibilidade de acesso aos indivíduos que possam estar localizados em grupos de maior risco de exposição à conteúdo mis/desinformativo. Ainda que nessa interação-disputa tenha-se potencial de acirramento da polarização, as possibilidades de intervenção parecem ser maiores do que aquelas decorrentes de uma realidade na qual a disseminação de mis/desinformação estaria envolvida em um contexto de defecções quase nulas; pois se na existência de câmaras de eco, uma vez convencido e cooptado pelo grupo de maior prevalência de mis/desinformação, não haveria mais intervenção (informações fundamentadas em evidências científicas) que pudesse efetivamente chegar nesse indivíduo ou coletividades.

Os trabalhos que trazem evidências nesse sentido não são, contudo, colocados em questão. Mas, faz-se necessário melhor investigar as diferenças contextuais entre amostras e empreendimentos metodológicos que levaram a realidades fenomenológicas tão variadas. Os novos estudos podem, por exemplo, se debruçar sobre outros temas que não político-partidários (sobretudo concernindo o campo da saúde) para verificar se os achados coadunam com aqueles encontrados nos resultados da presente pesquisa. Além disso, desenhos metodológicos baseados em coortes poderiam estender a análise para diferentes intervalos temporais de acompanhamento.

Em contrapartida, aqueles(as) que optem por análises sistemáticas poderiam investigar as condições de deslocamentos intercomunitários ao longo do tempo, mapeando padrões e em quais circunstâncias usuários de grupos mais susceptíveis à mis/desinformação passam a ser cooptados por aqueles com maior prevalência de conteúdo crível acerca das vacinas, e de que maneira (discursivamente) as interações entre nós de diferentes comunidades se dá. Neste sentido, trabalhos que busquem estudar qualitativamente a recepção aos conteúdos anti-hegemônicos nas comunidades das quais os usuários fazem parte, a fim de melhor compreender as dinâmicas de reenquadramento de conteúdos pró-vacinas quando conseguem entrar em

comunidades anti-vacinas, são de destacada importância para que seja possível traçar recomendações a nível infodemiológico visando uma atuação mais efetiva no enfrentamento da disseminação de desinformação acerca das vacinas no Twitter.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte financeiro ao projeto Enfrentando a desinformação no Twitter: uma análise da participação de atores da ciência no debate acerca das vacinas contra a COVID-19 (2021-2024), Chamada CNPq/MCTI/FNDCT nº 18/2021, através do processo CNPQ 406611/2021-1, e à Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB).

Referências

- Amaral, I., & Santos, S. J. (2019). Algoritmos e redes sociais: a propagação de fake News na era da pós-. In: Figueira J, Santos S (Ed.), *As fake news e a nova ordem (des)informativa na era da pós-verdade* (pp. 63–85). Portugal: Imprensa da Universidade de Coimbra.
- Cinelli, M., Quattrocioni, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 social media infodemic. *Scientific Reports*, 10(1). <https://doi.org/10.1038/S41598-020-73510-5>
- Cohen, K. B., & Hunter, L. (2008). Getting Started in Text Mining. *PLOS Computational Biology*, 4(1), e20. <https://doi.org/10.1371/JOURNAL.PCBI.0040020>
- Conover, M., Ratkiewicz, J., Francisco, M., Goncalves, B., Menczer, F., & Flammini, A. (2011). Political Polarization on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1), 89–96. <https://doi.org/10.1609/ICWSM.V5I1.14126>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv. <https://doi.org/10.48550/arxiv.1810.04805>
- Du, S., & Gregory, S. (2017). The echo chamber effect in twitter: Does community polarization increase? *Studies in Computational Intelligence*, 693, 373–384. https://doi.org/10.1007/978-3-319-50901-3_30
- Eysenbach, G. (2009). Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *Journal of Medical Internet Research*, 11(1). <https://doi.org/10.2196/JMIR.1157>
- Ferreira, F. V., Varão, R., Boselli, M. A., Santos, L. B., & Moret, M. A. (2022). Uso de Python para detecção de fake news sobre a covid-19: desafios e possibilidades. *Revista Eletrônica de Comunicação, Informação & Inovação Em Saúde*, 16(2). <https://doi.org/10.29397/RECIIS.V16I2.3253>
- Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(6), e1332. <https://doi.org/10.1002/WIDM.1332>
- Geron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. Tokyo: O'Reilly.
- Gomaa, W., & A. Fahmy, A. (2013). A Survey of Text Similarity Approaches. *International Journal of Computer Applications*, 68(13), 13–18. <https://doi.org/10.5120/11638-7118>
- Hotho, A., Nürnberger, A., & Paaß, G. (2005). A Brief Survey of Text Mining. *Journal for Language Technology and Computational Linguistics*, 20(1), 19–62. <https://doi.org/10.21248/JLCL.20.2005.68>
- Jung, H., & Lee, B. G. (2020). Research trends in text mining: Semantic network and main path analysis of selected journals. *Expert Systems with Applications*, 162, 113851. <https://doi.org/10.1016/J.ESWA.2020.113851>
- Khan, J. Y., Khondaker, Md. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2021). A benchmark study of machine learning models for online fake news detection. *Machine Learning with Applications*, 4, 100032. <https://doi.org/10.1016/J.MLWA.2021.100032>
- Lima, C. R. M., Sánchez-Tarragó, N., Moraes, D., Grings, L., & Maia, M. R. (2020). Emergência de saúde pública global por pandemia de Covid-19. *Folha de Rostto*, 6(2), 5–21. <https://doi.org/10.46902/2020N2P5-21>
- Liu, B. (2020). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge, England: Cambridge University Press.
- Macanovic, A. (2022). Text mining for social science – The state and the future of computational text analysis in sociology. *Social Science Research*, 108, 102784. <https://doi.org/10.1016/J.SSRESEARCH.2022.102784>
- Massarani, L., Brotas, A., Costa, M., & Neves, L. F. (2021). Vacinas contra a COVID-19 e o combate à desinformação na cobertura da Folha de S. Paulo. *Fronteiras - Estudos Midiáticos*, 23(2), 29–43. <https://doi.org/10.4013/fem.2021.232.03>
- Mønsted, B., & Lehmann, S. (2022). Characterizing polarization in online vaccine discourse—A large-scale study. *PLOS ONE*, 17(2), e0263746. <https://doi.org/10.1371/JOURNAL.PONE.0263746>

Recuero, R. (2017). *Introdução à análise de redes sociais online*. Salvador: EDUFBA.

Recuero, R., & Soares, F. B. (2021). O Discurso Desinformativo sobre a Cura do COVID-19 no Twitter: Estudo de caso. *E-Compós*, 24, 1–29. <https://doi.org/10.30962/EC.2127>

Recuero, R., & Zago, G. (2021). “RT, por favor”: considerações sobre a difusão de informações no Twitter. *Fronteiras - Estudos Midiáticos*, 12(2), 69–81. <https://doi.org/10.4013/4668>

Santos, C. R. P., & Maurer, C. (2020). Potencialidades e limites do fact-checking no combate à desinformação. *Comunicação & Informação*, 23. <https://doi.org/10.5216/CI.V23I.57839>

Shore, J., Baek, J., & Dellarocas, C. (2018, October 22). Twitter Is Not the Echo Chamber We Think It Is. *MITSloan Management Review*. <https://sloanreview.mit.edu/article/twitter-is-not-the-echo-chamber-we-think-it-is/>

Soares, F. B., Viegas, P., Bonoto, C., & Recuero, R. (2021). Covid-19, desinformação e Facebook: circulação de URLs sobre a hidroxicloroquina em páginas e grupos públicos. *Galáxia (São Paulo)*, 46. <https://doi.org/10.1590/1982-2553202151423>

Suen, C. Y. (1979). n-Gram Statistics for Natural Language Understanding and Text Processing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 164–172. <https://doi.org/10.1109/TPAMI.1979.4766902>

Tao, D., Yang, P., & Feng, H. (2020). Utilization of text mining as a big data analysis tool for food science and nutrition. *Comprehensive Reviews in Food Science and Food Safety*, 19(2), 875–894. <https://doi.org/10.1111/1541-4337.12540>

Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS ONE*, 13(9). <https://doi.org/10.1371/JOURNAL.PONE.0203958>

van der Linden, S. (2022). Misinformation: susceptibility, spread, and interventions to immunize the public. *Nature Medicine*, 28(3), 460–467. <https://doi.org/10.1038/s41591-022-01713-6>

World Health Organization. (2020). *Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation*. <https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation>