

Detecção de Posicionamento e Rotulação Automática de Usuários do Twitter: o caso da CPI da Covid-19

Stance Detection and Twitter Users Automatic Labelling: the case of Covid-19's CPI

Patricia Dias dos Santos¹ , Denise Hideko Goya^{1,2} 

¹Pós-Graduação em Ciência da Computação – Universidade Federal do ABC (UFABC)

²Centro de Matemática, Computação e Cognição – Universidade Federal do ABC (UFABC)

Santo André, SP – Brasil

{patricia.santos, denise.goya}@ufabc.edu.br

Abstract. *With the increasing influence of social media on public opinion, automated identification of political positions has become a crucial challenge for the fields of information systems and political science. In this context, this article aims to detect the stance and label a large number of Twitter users in Brazil on a politically controversial topic, in an automated and non-dependent way on annotated databases. For this purpose, an automated method is proposed to detect and label the stance of Twitter users on a politically controversial and polarized topic, using the Covid-19 Parliamentary Inquiry Committee (CPI) as a case study. The classification is based on a spectrum that encompasses both favorable and unfavorable positions towards the CPI, through the calculation of a sentiment score and two complementary metrics: balance degree and engagement. Unsupervised computational approaches, such as dimensionality reduction methods and clustering algorithms, were combined with minimally supervised techniques such as topic modeling and contextualized embeddings. This approach, along with social factors such as homophily and network structure, allowed for the automatic labeling of approximately 98% of the users in the studied databases with minimal supervision. This strategy may have significant implications for analyzing public opinion on controversial issues, providing insights into the distribution of political positions and the engagement strategies that social media users may employ.*

Keywords. *Stance Detection; User Automatic Labelling; Polarization; Twitter.*

Resumo. *Com o aumento da influência das mídias sociais na opinião pública, a identificação automatizada de posicionamentos políticos tornou-se um desafio crucial para as áreas de sistemas de informação e ciência política. Nesse contexto, esse artigo tem por objetivo detectar o posicionamento e rotular uma*

grande massa de usuários do Twitter no Brasil em um tema politicamente controverso, de forma automatizada e não dependente de bases anotadas. Para isso é proposto um método automatizado para detectar e rotular o posicionamento de usuários do Twitter em relação a um tema politicamente controverso e polarizado, utilizando a CPI da Covid-19 como estudo de caso. A classificação é feita com base em um espectro que engloba posições favoráveis e contrárias à CPI, por meio do cálculo de uma pontuação de valência de posicionamento e duas métricas complementares: grau de equilíbrio e engajamento. Foram combinadas abordagens computacionais não supervisionadas, tais como métodos de redução de dimensionalidade e algoritmos de agrupamento, com técnicas minimamente supervisionadas como modelagem de tópicos e embeddings contextualizados. Essa abordagem, em conjunto com fatores sociais, tais como homofilia e estrutura de rede, permitiram rotular automaticamente aproximadamente 98% dos usuários presentes nas bases de dados estudadas, de forma minimamente supervisionada. Essa estratégia pode ter implicações significativas para a análise de opinião pública em questões controversas, fornecendo insights sobre a distribuição dos posicionamentos políticos e as estratégias de engajamento que podem ser usadas pelos usuários das mídias sociais.

Palavras-Chave. *Deteção de Posicionamento; Rotulação Automática de Usuários; Polarização; Twitter.*

1. Introdução

A detecção de posicionamento tem uma longa tradição no domínio dos debates políticos e ideológicos online [Mohammad et al. 2016] e a sua aplicabilidade prática permite avaliar a opinião pública, identificar grupos que se cruzam e divergem, e compreender questões de interesse para diferentes comunidades de usuários [Magdy et al. 2016]. Estudos recentes sugerem que os usuários da web tendem a polarizar sua opinião [Conover et al. 2011] e a formar comunidades a partir de afinidades políticas [Figeac and Favre 2021], seguindo o princípio da homofilia [McPherson et al. 2001].

Apesar da natureza ruidosa, desestruturada e fragmentada dos *tweets*, eles têm sido analisados com sucesso por meio de técnicas de mineração de texto, processamento de linguagem natural e aprendizado de máquina [D’Andrea et al. 2019]. Atualmente, uma das suas principais aplicações é o monitoramento de questões políticas, uma vez que as informações extraídas da web produzem um meio relevante de, juntamente com técnicas mais tradicionais como pesquisa de opinião [Jungherr et al. 2017], estimar o posicionamento médio de uma população em relação a um determinado tópico.

As conversas nas mídias sociais entre 2020 e 2021 no Brasil foram amplamente dominadas por discussões sobre as ações do governo federal no combate à pandemia [Maia et al. 2021] e a vacinação contra a Covid-19 [Malagoli et al. 2021], com discurso altamente politizado. Em maio de 2021 foram iniciados os trabalhos de uma Comissão Parlamentar de Inquérito, intitulada CPI da Pandemia¹, também conhecida por CPI da

¹<http://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=464162&ori=1>

Covid-19, criada para investigar possíveis omissões do governo no enfrentamento à pandemia. Uma base de dados composta por conversações sobre a CPI foi coletada e utilizada como estudo de caso nesse artigo, por se tratar de um tema controverso, polarizado e atual ao ter a política e a ciência como temas centrais.

Controvérsias políticas são muito complexas e não raramente opiniões são expressas de forma sutil, o que dificulta a determinação de posicionamentos contrários/favoráveis [Awadallah et al. 2012]. A tarefa de aferir o posicionamento de um usuário referente a um tema controverso torna-se desafiadora, assim como o processo de apresentação de uma resposta “correta” a essa aferição. Mas como definir controvérsia? Neste trabalho será utilizada a definição de [Dori-Hacohen et al. 2015], de que “tópicos controversos são aqueles que geram forte desacordo entre grandes grupos de pessoas”.

No estudo de [AlDayel and Magdy 2021], são propostos quatro conjuntos de características para modelar a posição do usuário no Twitter. Esses conjuntos incluem conteúdo, que modela o texto do tweet; rede de interação, que modela a rede com a qual o usuário interage em suas postagens; rede de preferência, que modela a rede de preferência do usuário a partir dos tweets que ele curte; e rede de conexão, que modela as conexões sociais online entre os usuários. Para este trabalho, será utilizada a rede de interação do usuário. Isso significa que a posição do usuário será inferida com base em vários textos retuitados pelo usuário, em vez de se basear em um único texto anônimo.

Embora haja um grande interesse em identificar o posicionamento de um usuário em relação a um determinado assunto, a escassez de dados anotados de forma adequada é um obstáculo significativo para o desenvolvimento de modelos precisos de detecção de posicionamento [Li et al. 2021]. Também é importante notar que a maioria das pesquisas sobre detecção de postura enfoca o idioma inglês, enquanto poucos estudos se dedicam a explorar outras línguas [Alturayef et al. 2023]. A principal **Questão de Pesquisa (QP)** deste trabalho é: **QP:** *Como detectar o posicionamento e rotular uma grande massa de usuários do Twitter no Brasil em um tema politicamente controverso, de forma automatizada e não dependente de bases anotadas?* Tendo como motivação a questão de pesquisa, o **objetivo geral** desse trabalho é desenvolver um método automatizado para estimar o posicionamento e rotular usuários do Twitter em um tema politicamente controverso e com indícios de polarização. O **objetivo secundário** consiste em criar um conjunto de dados de usuários rotulados como a favor ou contra um determinado assunto, com alta qualidade, a fim de reduzir os riscos de vieses e erros comuns associados à rotulação manual.

Este estudo é uma versão ampliada de um trabalho anterior publicado no XI Workshop Brasileiro de Análise e Mineração de Redes Sociais (BRASNAM) [Santos and Goya 2022]. Nesta versão ampliada, foram realizadas as seguintes adições: (i) ampliação dos dados analisados no estudo de caso, com a inclusão do conjunto de bases de dados referentes à segunda fase da CPI, destacando diferenças e semelhanças entre as duas fases, e explorando a temporalidade dos dados; (ii) inclusão de discussões sobre a qualidade dos resultados obtidos após aplicação do método proposto.

As principais **contribuições** deste trabalho são descritas a seguir:

1. Adaptação de uma técnica não supervisionada de detecção de posicionamento para determinar o posicionamento de grupos de usuários do Twitter em relação

- a um tópico controverso e polarizado sobre postagens em português.
2. Rotulação automática individualizada de centenas de milhares de usuários empregando uma combinação de atribuição de rótulos e cálculo de pontuação de valência.
 3. Mensuração dos graus de engajamento e equilíbrio para caracterizar o comportamento dos usuários rotulados nos eventos.
 4. Método de classificação que obteve êxito na avaliação em um estudo de caso envolvendo uma quantidade expressiva de dados e que tem potencial para ser generalizado e aplicado na detecção automática de posicionamento em variados temas que apresentem sinais de polarização e contexto político.

O artigo está organizado da seguinte forma: a Seção 2 discute os fundamentos e trabalhos relacionados, a Seção 3 descreve materiais e métodos, a Seção 4 analisa os resultados e a Seção 5 fornece uma avaliação da qualidade dos resultados. Por fim, a Seção 6 discute as ameaças à validade da abordagem proposta e a Seção 7 conclui o trabalho e apresenta sugestões para pesquisas futuras.

2. Fundamentos e Trabalhos Relacionados

A detecção de posicionamento tem como objetivo determinar automaticamente se o autor de um texto é a favor, contra ou neutro em relação a um tópico ou proposição [Mohammad et al. 2017]. Sua aplicação é particularmente relevante nas mídias sociais, pois oferece a oportunidade de identificar o posicionamento de um grande número de usuários sobre diferentes questões.

[Darwish et al. 2020] apontam que as abordagens atuais apresentam vantagens e desvantagens. As abordagens supervisionadas são fáceis de implementar, mas dependem de dados de treinamento rotulados manualmente e sua precisão pode variar de acordo com a técnica de classificação e a quantidade dos dados de teste e treinamento. Já as abordagens semi e não supervisionadas requerem menos interferência humana, mas podem ter dificuldade em classificar usuários que raramente expressam suas opiniões, produzindo resultados abaixo do ideal nessas situações. Por outro lado, essas abordagens tendem a ser mais eficazes na classificação de usuários altamente vocais, que expressam suas opiniões com mais frequência.

As abordagens supervisionadas são baseadas em aprendizado de máquina e orientadas a dados, utilizando algoritmos tradicionais de aprendizado de máquina ou de aprendizado profundo para treinar um classificador em diversos contextos [Alturayef et al. 2023], porém, elas são altamente dependentes de bases anotadas manualmente. Neste trabalho, o intuito é realizar a rotulação de usuários de forma automática, adotando portanto uma abordagem híbrida entre o que foi proposto por [Darwish et al. 2020] para a classificação de clusters polarizados altamente vocais e uma abordagem minimamente supervisionada similar à adotada por [Sobhani et al. 2015] para rotular os usuários menos vocais, aplicando modelagem de tópicos em textos não anotados compartilhados por cada grupo polarizado para extrair tópicos, e em seguida, clustalizando os dados baseados nesses tópicos.

[Küçük and Can 2020] descrevem trabalhos recentes que utilizaram uma ampla gama de métodos de detecção de posicionamento, incluindo áreas de pesquisa como pro-

cessamento de linguagem natural e ciências sociais computacionais. A detecção de posicionamento em redes sociais pode ser dividida em duas grandes abordagens: na primeira são utilizados algoritmos tradicionais de aprendizado de máquina [Wojatzki and Zesch 2016], com etapas de rotulação manual [Mohammad et al. 2017, Sobhani et al. 2017] enquanto que a segunda abordagem é baseada em técnicas de aprendizado profundo [Zarrella and Marsh 2016], em combinação com modelos BERT (*Bidirectional Encoder Representations from Transformers*) [Popat et al. 2019, Samih and Darwish 2021, Sirrianni et al. 2021]. A aplicação de técnicas de modelagem de tópicos também produz resultados interessantes: [Lin et al. 2017] utilizaram o modelo *Nonnegative Matrix Factorization* (NMF) para a identificação de instâncias múltiplas de posicionamento, com o uso de rotulação automatizada de textos.

Existem poucos trabalhos sobre detecção de posicionamento sobre temas políticos em idiomas distintos do inglês [Vamvas and Sennrich 2020]: [Rashed et al. 2021] coletaram *tweets* relacionados às eleições da Turquia em 2018 e aplicaram um método não supervisionado de detecção de posicionamento, enquanto [Ebeling et al. 2020] propõem um *framework* para analisar e caracterizar o comportamento de grupos com posições opostas e a relação com a polarização política, usando *tweets* sobre a Covid-19 no Brasil. Esses autores empregaram técnicas de modelagem de tópicos, detecção de comunidade e análise de características linguísticas para determinar aspectos psicológicos nos textos. [Christie et al. 2018] caracterizaram, de um amplo conjunto de *tweets* em português, a corrida presidencial brasileira de 2018. [Sáenz and Becker 2021] propõem um modelo de detecção de posicionamento baseado em BERT e mecanismos baseados em atenção para identificar palavras influentes para classificação de posicionamento. Os autores aplicaram o método em *tweets* sobre a vacinação contra a Covid-19.

Em estudo anterior [Santos and Goya 2021] combinaram a contagem de recorrência de palavras para rotular automaticamente o posicionamento de um usuário como contrário ou favorável em relação a um tema, com a classificação supervisionada de BERT para a modelagem de tópicos e a técnica de classificação não supervisionada proposta por [Darwish et al. 2020] para prever o posicionamento de usuários prolíficos do Twitter em relação a assuntos polêmicos em língua portuguesa. Os resultados mostraram que os usuários mais vocais eram responsáveis por cerca de 40 a 50% do volume de *retweets* com forte polarização em dois grupos majoritários, sendo que menos de 1,5% dos usuários apresentava posicionamento divergente ao longo do tempo em relação ao tópico estudado. Já no trabalho [Santos and Goya 2022], foi aumentada significativamente a proporção de perfis rotulados, ao apresentar um método para detectar o posicionamento de usuários menos vocais, além de introduzir medidas que complementam a aferição de posicionamento e fornecem indicativos de engajamento e equilíbrio.

Os trabalhos estudados na literatura rotulam apenas os usuários mais ativos na discussão [Darwish et al. 2020, Santos and Goya 2021], dependem de etapas de rotulação manual de textos [Mohammad et al. 2016, Sobhani et al. 2017] ou posicionamento de usuários [Kamiński et al. 2021, Ebeling et al. 2020]. O método proposto neste trabalho permite rotular a ampla maioria dos usuários participantes da discussão, sem necessidade de etapas de rotulação manual.

3. Materiais e Métodos

Nesta seção, é descrita a metodologia proposta neste estudo. A Figura 1 mostra as principais etapas do processo, posteriores à coleta de dados. A seção 3.2 fornece uma visão geral das seis etapas, que são detalhadas na sequência.

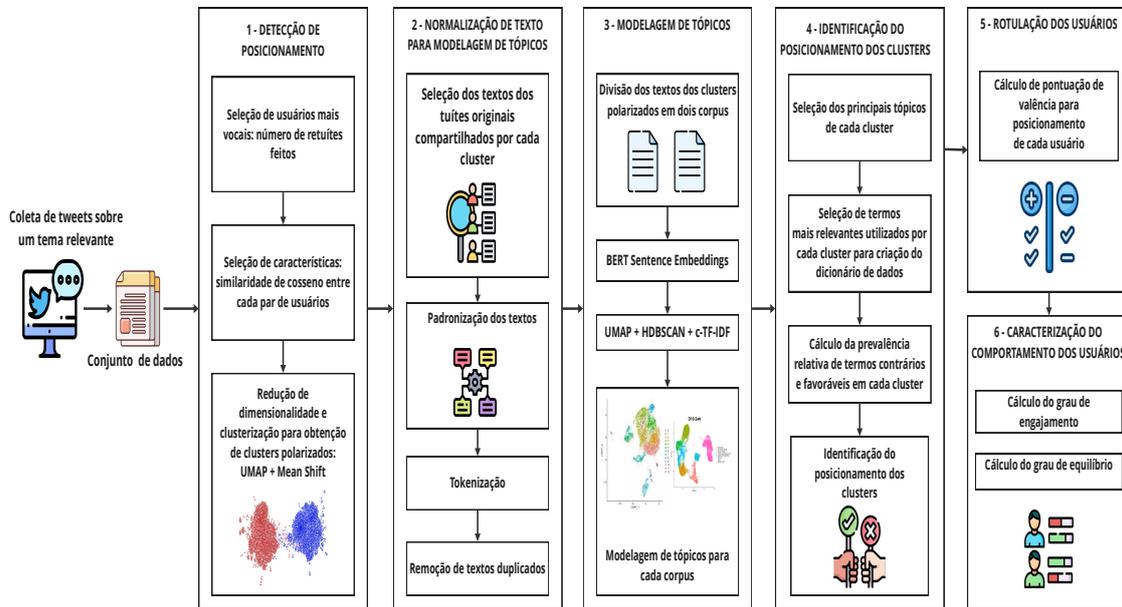


Figura 1. Diagrama descrevendo a metodologia proposta

3.1. Coleta de Dados

A Tabela 1 mostra um resumo dos dados dos *retweets* coletados relacionados à CPI da Covid-19, utilizando-se da API v1.1 do Twitter². As coletas foram realizadas em tempo real cobrindo cerca de 24 horas de repercussão para cada sessão de depoimentos. Várias coletas de dados foram executadas simultaneamente para garantir que todas as sessões fossem coletadas e para maximizar a cobertura dos *tweets* relacionados à CPI durante as sessões. Isso foi feito para assegurar que a amostra coletada representasse uma ampla variedade de opiniões e perspectivas, bem como para capturar o maior número possível de *tweets* relevantes em tempo real. As *strings* de busca usadas na coleta foram compostas por termos e *hashtags* selecionados de acordo com os cronogramas de depoimentos divulgados na imprensa. As bases utilizadas neste artigo estão disponíveis no site do projeto Observa³.

A CPI da Covid, com duração total de 180 dias, foi dividida em duas fases de 90 dias, ou onze semanas cada. A separação das bases de dados em duas fases foi feita levando em conta mudanças na evolução da dinâmica política e diferentes testemunhas e

²<https://developer.twitter.com>

³Projeto CPI da Covid do Observatório de Conflitos da Internet, apoiado por Fapesp/MCTI/MC/CGI 2018/23022-3 <https://observa.pesquisa.ufabc.edu.br/list/cpi-da-covid/>

depoimentos. A análise em uma base unificada pode ser útil em trabalhos futuros, mas a divisão atual possibilita uma avaliação mais aprofundada das particularidades de cada fase e seus impactos na percepção pública.

Tabela 1. Dados dos tweets utilizados neste estudo

| Tipo | Descrição | Dados | Tamanho |
|-----------------|--|---|---------|
| <i>Retweets</i> | 1ª fase: 4/05 a 16/07 de 2021 2ª fase: 3/08 a 26/10 de 2021 | Id, Data, Texto, Usuário que fez o <i>retweet</i> , Usuário retuitado | 13,3 mi |
| Usuários | Dados dos usuários que fizeram <i>retweets</i> | Nome de usuário | 757 mil |

3.2. Abordagem Metodológica

A abordagem proposta consiste em um método não supervisionado organizado em 6 etapas. Os resultados preliminares das primeiras quatro etapas foram discutidos no estudo [Santos and Goya 2021]:

1. Aplicação de um método de detecção de posicionamento não supervisionado que busca agrupar os usuários em dois clusters polarizados levando em conta os usuários que eles retuitam (ver Seção 3.3).
2. Recuperação do conteúdo retuitado em cada cluster e normalização dos dados textuais para aplicação de técnica de modelagem de tópicos (ver Seção 3.4).
3. Modelagem de tópicos em cada cluster utilizando *embeddings* (ver Seção 3.5).
4. Identificação do posicionamento dos clusters utilizando um dicionário de termos recorrentes utilizados por cada grupo polarizado, com confirmação humana (ver Seção 3.6).
5. Rotulação dos usuários, utilizando uma técnica de atribuição de rótulos e cálculo de pontuação de valência (ver Seções 3.7 e 3.8).
6. Caracterização do comportamento dos usuários rotulados (ver Seção 3.9).

3.3. Detecção de posicionamento baseada na iteração entre usuários

Em cada base semanal, a detecção de posicionamento foi obtida a partir da aplicação dos quatro passos descritos a seguir:

1. **Seleção de usuários mais vocais.** Foi calculado o quartil da quantidade *retweets* por usuário, com o objetivo de selecionar 25% dos valores mais elevados. Desse modo, foi possível selecionar os usuários mais vocais, ou seja, aqueles com maior participação na discussão em cada base. Após isso, foi realizada uma filtragem adicional para remover usuários que fizeram exclusivamente *auto-retweets*, ou usuários cujos *retweets* foram feitos para uma única conta. O objetivo da filtragem foi o de reduzir a incidência *outliers* (usuários não clusterizados).
2. **Seleção de características.** Foi então construído um vetor de categorias para cada usuário selecionado baseado nas contas que ele retuitou e calculada a similaridade do cosseno entre cada par de usuários, utilizando esse vetor.

3. **Redução de dimensionalidade.** Após o cálculo da similaridade, os usuários foram projetados em um espaço bidimensional utilizando o algoritmo *Uniform Manifold Approximation and Projection* - UMAP [McInnes et al. 2018].
4. **Clusterização.** Feita a projeção dos usuários em um espaço bidimensional, eles foram clusterizados utilizando o algoritmo *Mean Shift*. Isso permitiu a obtenção de dois grandes clusters polarizados.

É importante destacar que nesse estudo foram utilizados apenas *retweets* “puros”, isto é, aqueles em que os usuários simplesmente clicam um botão para redistribuir o conteúdo. Os *retweets* com comentários foram excluídos porque considerou-se que um usuário ao compartilhar um retweet com comentário pode expressar endosso ou desacordo, enquanto que o simples ato de retuitar normalmente sinaliza concordância ou apoio por parte dos usuários que compartilham o texto [Wischnewski et al. 2022].

3.4. Normalização de dados textuais para modelagem de tópicos

Essa etapa é necessária para a realização da modelagem de dados, dado que um tweet é composto de dados textuais altamente irregulares. Após a obtenção dos textos dos *retweets* compartilhados pelos usuários de cada cluster, estes foram normalizados utilizando as seguintes técnicas:

1. **Padronização.** Detecção e remoção de *links*, emoticons e *emojis*, *hashtags*, pontuações, espaços e linhas em branco, 'RT' e @ (informações de usuários), números, palavras reservadas do Twitter, aspas simples e duplas e cadeias de caracteres que possuam repetições, conversão de todas as palavras em minúsculas;
2. **Tokenização.** Tokenização dos textos dos *tweets* separados por espaços em branco, removendo também as *stopwords* (verbos auxiliares, preposições, artigos, pronomes e as palavras mais frequentes) e palavras com menos de três caracteres;
3. **Remoção de textos duplicados.** Um filtro foi aplicado para conservar os *tweets* únicos em termos de sua forma padronizada: foi calculada a similaridade entre duas *strings*, e mantidos somente *tweets* não-similares.

3.5. Modelagem de tópicos

Partindo-se do princípio da homofilia [McPherson et al. 2001, Figeac and Favre 2021], considerou-se que os usuários de cada cluster teriam predominantemente a mesma posição (contrária ou a favorável) em relação à CPI. Assim, foi realizada a modelagem de tópicos separadamente nos textos dos *retweets* compartilhados pelos usuários de cada agrupamento majoritário e em cada base coletada. Para a modelagem de tópicos foi utilizada a técnica BERTopic desenvolvida por [Grootendorst 2022]. Esta técnica consiste na extração de *embeddings* de documentos, agrupamento para a criação de grupos de documentos semelhantes com UMAP e HDBSCAN, e na geração de tópicos através da obtenção das palavras mais importantes por meio de agrupamento com TF-IDF baseado em classe (c-TF-IDF). Na extração dos *embeddings* foi utilizado o modelo BERT pré-treinado em português conhecido como BERTimbau [Souza et al. 2020], o qual utiliza um corpus em português brasileiro (BrWaC [Wagner Filho et al. 2018]).

3.6. Identificação de posicionamento dos clusters

A partir da modelagem tópicos, foram identificadas palavras que foram usadas predominantemente por um dos grupos de usuários em cada conjunto de bases. Dentre os termos prevalentes, foram selecionados os mais frequentes ao longo do tempo, para compor um dicionário de termos recorrentes. Após análise, decidiu-se separar o dicionário em dois blocos, já que foi observado que termos que apareciam com muita frequência na primeira frase deixaram de existir ou se tornaram menos importantes na segunda. A Tabela 2 mostra o dicionário dos termos recorrentes selecionados.

Tabela 2. Dicionário de dados de cada grupo por conjunto de base de dados

| Termos recorrentes | Primeira fase | Segunda fase |
|--|--|------------------------------------|
| usados por pessoas favoráveis à CPI | bolsonaris*, genoc*, negacion*, desgovern* | bolsonaris*, genoc*, crim*, medic* |
| usados por pessoas contrárias à CPI | circo, palhaç*, ladr*, pedof* | circo, pedof*, vergonh*, corrup* |

Na sequência, foi realizada uma busca dos termos do dicionário e feita uma contagem de ocorrências de palavras nos conjuntos de tópicos gerados por cada cluster de cada base. Por fim, foi realizada a rotulação automática de cada cluster com base na prevalência relativa desses termos e suas variações, em cada semana da CPI.

3.7. Rotulação automática dos usuários

A rotulação automatizada dos usuários ocorre em três etapas: (i) rotulação dos usuários mais vocais de acordo com o seu cluster e atribuição de rótulo favorável ou contrário às contas que fizeram os *tweets* originais, a partir da porcentagem de *retweets* feitos pelos usuários clusterizados, (ii) rotulação dos usuários menos vocais e re-rotulação dos mais vocais a partir dos *retweets* feitos para cada conta rotulada na etapa anterior e (iii) aplicação de um rótulo numérico de pontuação de valência para usuário. Para a etapa (i), o rótulo do cluster foi definido conforme indicado na Seção 3.6. A etapa (ii), está descrita a seguir. E a etapa (iii) é detalhada na Seção 3.8.

A detecção de posicionamento não supervisionada só funciona de forma eficaz na tarefa de classificar os usuários mais vocais (isto é, que realizam quantidade de *retweets* dentro do limite do quartil superior), os quais constituem a minoria dos usuários [Darwish et al. 2020]. Nos dois conjunto de bases estudados neste trabalho, a porcentagem de usuários mais vocais foi de aproximadamente 27%, em ambas.

Um dos principais objetivos deste trabalho é rotular o posicionamento do maior número possível de usuários adicionais. Para tanto, foram utilizados os usuários rotulados automaticamente em cada cluster na etapa anterior para propagar rótulos aos usuários menos vocais, com base nas contas que cada usuário clusterizado retuitou. Calculou-se então o número total de *retweets* que as contas de usuário receberam em cada cluster, bem como a sua porcentagem. Feito isso, foi atribuído um rótulo (contrário ou favorável) para essas contas a partir da maior porcentagem de *retweets* recebidos por elas. As contas cuja diferença de porcentagem entre os clusters era inferior ao limiar de 20% não foram

consideradas para a rotulação dos menos vocais. Essa escolha foi feita com o intuito de diminuir a incerteza (evitar casos de contas igualmente retuitadas em ambos os clusters) e realizar a atribuição, num primeiro momento, de rótulos dos usuários mais fortemente posicionados. O percentual de contas retuitadas por usuários de ambos os clusters variou entre 3,8 a 7,3% ao longo das bases e consistiu em muitos casos de contas de jornalistas e veículos de imprensa, enquanto que o percentual de contas desconsideradas variou entre 0,21 e 0,73%. Após isso, os usuários menos vocais foram rotulados e os mais vocais receberam novos rótulos de acordo com os *retweets* que eles fizeram para as contas de usuários. Nessa etapa também foi atribuído um rótulo para os *tweets* originais feitos pelas contas de usuário de acordo com o rótulo que elas receberam.

3.8. Pontuação de valência para posicionamento

Para rotular o posicionamento dos usuários menos vocais, foi calculada uma pontuação de valência. Essa pontuação variou entre [-1, 1], onde um valor próximo a 1 indica que o usuário tem posicionamento similar a usuários favoráveis à CPI; -1 indica posicionamento contrário; e 0, uma posição de equilíbrio. Para calcular a pontuação de valência foi utilizada a equação (1), que é uma adaptação da proposta por [Conover et al. 2011].

$$p(u) = 2 * \left(\frac{r(u, F)}{N(F)} \middle/ \left(\frac{r(u, F)}{N(F)} + \frac{r(u, C)}{N(C)} \right) \right) - 1 \quad (1)$$

Onde $r(u, F)$ é o número de *retweets* feitos pelo usuário u para contas de usuários da lista com rótulo favorável e $N(F) = \sum_u r(u, F)$ é o número total de *retweets* recebidos pelas contas da lista com rótulo favorável em cada base. Analogamente, $r(u, C)$ e $N(C)$ são definidos para posicionamento contrário.

3.9. Caracterização do comportamento dos usuários

O comportamento dos usuários rotulados a partir da pontuação de valência de posicionamento é caracterizado por meio das duas métricas complementares descritas a seguir, pois a valência sozinha nem sempre traduz as noções de engajamento, indefinição ou imparcialidade no debate, úteis para melhor compreender os eventos:

1. Grau de **engajamento** de um usuário u , representado pelo máximo valor entre (i) sua participação relativa em *retweets* emitidos e (ii) sua presença relativa ao longo das n semanas. O grau de engajamento é calculado pela equação (2), em que $r(u) = r(u, F) + r(u, C)$ e representa o número total de *retweets* emitidos por u e que puderam ser rotulados, $max(r)$ representa a maior quantidade de total de *retweets* realizados por um único usuário nas n semanas e $s(u)$ é o número de semanas em que o usuário u participou. Usou-se logaritmo para uniformizar as grandezas, pois a distribuição da quantidade de *retweets* seguiu uma função potência. Assim, os valores de engajamento variam entre 0 (baixa participação) e 1 (forte engajamento), capturando ambas as noções de comprometimento: volume de interações e presença no debate ao longo do tempo.

$$eng(u) = max \left(\frac{\log_{10}(r(u))}{\log_{10}(max(r))}, \frac{s(u) - 1}{n - 1} \right) \quad (2)$$

2. Grau de **equilíbrio** de um usuário u , cujo cálculo é feito pela equação (3), representa a diferença entre as proporções de *retweets* a favor e contra realizados por u , considerando a totalidade de *retweets* $R(u)$ (incluindo os *retweets* realizados por u que não puderam ser rotulados). Assim, valores muito próximos dos extremos (-1 ou 1) indicam uma convergência de posicionamento para um dos polos, enquanto valores próximos de 0 podem representar indefinição, imparcialidade ou que o usuário retuitou mais postagens que não puderam ser rotuladas.

$$eq(u) = \frac{r(u, F) - r(u, C)}{R(u)} \quad (3)$$

A pontuação de valência de posicionamento e o grau de equilíbrio podem ser agrupados em 5 faixas com valores de -1 a 1 [Stefanov et al. 2020]:

$$pont = \begin{cases} --, & \text{se } p \in [-1, -0, 6) \\ -, & \text{se } p \in [-0, 6, -0, 2) \\ 0, & \text{se } p \in [-0, 2, 0, 2) \\ +, & \text{se } p \in [0, 2, 0, 6) \\ ++, & \text{se } p \in [0, 6, 1] \end{cases} \quad (4)$$

A escolha de um intervalo exato de 0,4 para cada faixa da equação (4) foi de certa forma arbitrária, mas considerada apropriada pois captura o posicionamento de um usuário mesmo em valores intermediários de pontuação. Além disso, o valor não poderia ser muito próximo de 0, pois a faixa de indefinição ou equilíbrio deve capturar usuários que não mostraram um posicionamento claro com relação aos seus *retweets* [Bechini et al. 2020].

4. Resultados e Discussão

Nesta seção são apresentados e discutidos os resultados obtidos na aplicação do método proposto na Seção 3. É importante destacar que na Seção 5 será descrita uma avaliação dos resultados obtidos com a aplicação do método, a fim de avaliar a qualidade da rotulação sobre uma amostra de usuários.

4.1. Modelagem de Tópicos e Rotulação dos Clusters

Na Tabela 3 é exibida uma amostra do resultado da aplicação da modelagem de tópicos em duas semanas, pertencentes às diferentes fases da CPI, respectivamente para cada um dos clusters. Pelo teor dos termos mais relevantes dos tópicos (na tabela foi exibida apenas uma seleção), já seria possível decidir por um rótulo para cada cluster, contrário ou favorável. No entanto, para reduzir a interferência humana no processo de rotulação dos clusters, foi realizada a identificação descrita na Seção 3.6, que fez uso do dicionário de termos recorrentes. Embora a presença de termos como “pedofilia” possa causar estranheza, eles estão alinhados com as estratégias discursivas de ataques aos adversários observadas na extrema-direita, conforme apontado por [Santos 2022]. Além disso, conforme observado por [de Carvalho Mendes et al. 2022], usuários contrários à CPI frequentemente recorreram à ironia e ao sarcasmo, utilizando expressões como “a CPI do circo” e “a CPI da comédia”.

Tabela 3. Exemplo de modelagem de tópicos, aplicada em cada cluster

| Rótulo favorável | | |
|------------------|------------------------------------|---|
| | tópico | termos mais relevantes |
| semana 1 | 5_cpi genocídio_genocídio | cpi genocídio, tirar foco, foco cpi, genocídio pode, movimento cpi |
| | 88_bolsonaristas_senadores | senadores bolsonaristas, tentando advogar, papel idiotas, impossível obstruir |
| semana 22 | 23_crimes contra_contra humanidade | crimes, contra humanidade, haia, bolsonaro acusado, desumano |
| | 43_povos originários_originários | povos originários, crime genocídio, proposta legislativa, indígenas |
| Rótulo contrário | | |
| | tópico | termos mais relevantes |
| semana 1 | 5_pedofilia_investigado | pedofilia, investigado pedofilia, presidente cpi, acusado, investigado |
| | 6_palhaços_circo | palhaços, cpi circo, cpi circence, covidão cpi, palhaço mandetta |
| semana 22 | 1_dinheiro_cpi_corruptos | dinheiro, cpi, corruptos, senado, circo, bandidos, saúde, vergonha, canalhas |
| | 12_vergonha_cpi vergonha | vergonha, ladrões dinheiro, macabros, credibilidade nenhuma, medo precoce |

A rotulação automatizada de cada cluster com base na prevalência relativa dos termos recorrentes em cada semana da CPI pode ser observada na Figura 2.

| primeira fase | | s1/1 | s1/2 | s2/1 | s2/2 | s3/1 | s3/2 | s4/1 | s4/2 | s5/1 | s5/2 | s6/1 | s6/2 | s7/1 | s7/2 | s8/1 | s8/2 | s9/1 | s9/2 | s10/1 | s10/2 | s11/1 | s11/2 |
|--------------------------|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| termos favoráveis | bolsonaris* | 1,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 0,0 | 0,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 1,0 | 0,0 | 1,0 |
| | genoc* | 0,7 | 0,3 | 0,8 | 0,2 | 0,6 | 0,4 | 0,2 | 0,8 | 0,1 | 0,9 | 0,0 | 1,0 | 0,4 | 0,6 | 0,9 | 0,1 | 0,0 | 0,0 | 0,0 | 1,0 | 0,0 | 1,0 |
| | negacion* | 1,0 | 0,0 | 0,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,7 | 0,3 | 1,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 |
| | desgovern* | 0,0 | 0,0 | 1,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 1,0 | 0,0 | 0,0 | 0,0 |
| termos contrários | circo | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 1,0 | 1,0 | 0,0 | 1,0 | 0,0 | 0,9 | 0,1 | 1,0 | 0,0 | 0,0 | 1,0 | 0,8 | 0,2 | 1,0 | 0,0 | 1,0 | 0,0 |
| | palhaç* | 0,1 | 0,9 | 0,0 | 1,0 | 0,0 | 1,0 | 1,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 0,0 | 1,0 | 0,5 | 0,5 | 0,9 | 0,1 | 0,9 | 0,1 |
| | ladr* | 0,0 | 0,0 | 0,0 | 1,0 | 0,0 | 1,0 | 1,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 0,0 | 1,0 | 1,0 | 0,0 | 1,0 | 0,0 | 0,9 | 0,1 |
| | pedof* | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 0,0 | 1,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 |
| classificação do cluster | | fav | con | fav | con | fav | con | con | fav | con | fav | con | fav | con | fav | fav | con | con | fav | con | fav | con | fav |
| segunda fase | | s12/1 | s12/2 | s13/1 | s13/2 | s14/1 | s14/2 | s15/1 | s15/2 | s16/1 | s16/2 | s17/1 | s17/2 | s18/1 | s18/2 | s19/1 | s19/2 | s20/1 | s20/2 | s21/1 | s21/2 | s22/1 | s22/2 |
| termos favoráveis | bolsonaris* | 0,0 | 0,0 | 0,0 | 1,0 | 0,2 | 0,8 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 0,0 | 0,0 | 1,0 | 0,6 | 0,4 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 1,0 |
| | genoc* | 0,1 | 0,9 | 0,3 | 0,7 | 0,0 | 0,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,7 | 0,3 | 1,0 | 0,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,3 | 0,7 | 0,0 | 1,0 |
| | crim* | 1,0 | 0,0 | 0,5 | 0,5 | 0,7 | 0,3 | 0,5 | 0,5 | 0,2 | 0,8 | 0,7 | 0,3 | 0,4 | 0,6 | 0,7 | 0,3 | 0,1 | 0,9 | 0,2 | 0,8 | 0,2 | 0,8 |
| | medic* | 0,0 | 0,0 | 0,7 | 0,3 | 0,4 | 0,6 | 0,3 | 0,7 | 0,4 | 0,6 | 0,6 | 0,4 | 0,2 | 0,8 | 0,4 | 0,6 | 0,4 | 0,6 | 0,0 | 1,0 | 0,0 | 1,0 |
| termos contrários | circo | 1,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 0,0 | 1,0 | 1,0 | 0,0 | 0,9 | 0,1 | 0,7 | 0,3 | 0,8 | 0,2 | 1,0 | 0,0 |
| | pedof* | 1,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 |
| | vergonh* | 0,8 | 0,2 | 1,0 | 0,0 | 1,0 | 0,0 | 1,0 | 0,0 | 0,9 | 0,1 | 0,0 | 1,0 | 0,9 | 0,1 | 0,8 | 0,2 | 0,5 | 0,5 | 1,0 | 0,0 | 0,9 | 0,1 |
| | corrup* | 1,0 | 0,0 | 0,8 | 0,2 | 1,0 | 0,0 | 0,3 | 0,7 | 0,3 | 0,7 | 0,2 | 0,8 | 0,8 | 0,2 | 1,0 | 0,0 | 0,6 | 0,4 | 0,5 | 0,5 | 1,0 | 0,0 |
| classificação do cluster | | con | fav | fav | con | con | fav |

Figura 2. Prevalência de termos recorrentes favoráveis e contrários em cada cluster (cinza mais escuro determina tendência para classificação)

4.2. Métricas Gerais

As métricas gerais obtidas a partir da aplicação do método proposto são mostradas na Tabela 4. O percentual de *retweets* feitos para contas não rotuladas em ambas as fases foi

de menos de 30%. Apesar disso, é interessante observar que menos de 3% dos usuários não puderam ser rotulados pois não retuitaram contas pertencentes aos dois clusters majoritários polarizados ou retuitaram contas que foram descartadas por serem igualmente retuitadas por usuários de ambos os clusters. O volume de contas não rotuladas pode ser considerado baixo e parece indicar que a quase totalidade dos usuários da base retuitou as mesmas contas de usuário. Isso pode ser um indício da ocorrência do efeito conhecido como câmara de eco (*echo chamber*), no qual os usuários têm suas crenças reforçadas por meio de interações com outros usuários que pensam da mesma forma, devido a padrões de interação homofílicos [Cinelli et al. 2020].

Tabela 4. Métricas Gerais

| | Primeira fase | | Segunda fase | |
|----------------------------------|---------------|------|--------------|------|
| Retweets | Quantidade | % | Quantidade | % |
| Total coletado | 10,1 mi | - | 3,2 mi | - |
| Rotulados | 7,5 mi | 74,1 | 2,5 mi | 78,0 |
| Rotulados a favor | 3,9 mi | 52,4 | 1,2 mi | 46,4 |
| Rotulados contra | 3,6 mi | 47,6 | 1,3 mi | 53,6 |
| Usuários envolvidos em RT | Quantidade | % | Quantidade | % |
| Total coletado | 664.793 | - | 276.441 | - |
| Rotulados | 653.981 | 98,4 | 269.943 | 97,6 |
| Menos vocais | 471.147 | 72,5 | 196.039 | 72,6 |
| Mais vocais | 179.834 | 27,5 | 73.904 | 27,3 |

O gráfico de *retweets* rotulados da Figura 3 mostra a distribuição dos *tweets* compartilhados pelos usuários ao longo das semanas. É interessante notar a alternância de uma maioria simples tendendo para um dos posicionamentos ao longo das semanas, com exceção das bases das semanas 2, 3 e 17, em que a concentração de *tweets* favoráveis foi consideravelmente maior. Isso ocorreu porque ao longo da CPI, vários episódios controversos envolvendo casos de suspeitas de corrupção e inépcia das autoridades foram expostos: a revelação da recusa de vacinas da Pfizer/BioNTech ganhou destaque considerável, pois envolvia uma solução que poderia ter contribuído para um controle mais rápido da doença no Brasil [Barros and do Vale 2021]. Tornando a situação mais crítica, a CPI deparou-se também com um escandaloso episódio de suspeita de corrupção relacionado à aquisição da vacina indiana Covaxin [Dantas and Nippes 2022]. Já nas semanas 7, 12 e 20 houve prevalência de *tweets* contrários à CPI. A repercussão negativa da CPI afetou significativamente a rede bolsonarista, que ao longo das semanas tentou mobilizar o público em apoio ao governo de Jair Bolsonaro. Para tanto, compartilharam *tweets* com o intuito de atacar e desacreditar a CPI, seja criticando os senadores envolvidos ou alegando que a comissão tinha apenas propósitos políticos e eleitoreiros [Morosini 2022].

No entanto, quando se observa o gráfico da pontuação da valência de posicionamento dos usuários ao longo das semanas (Figura 4), percebe-se que os usuários rotulados como muito contrários à CPI estiveram em menor número (19,7% na média de todas as 22 semanas), mas fizeram proporcionalmente mais *retweets* (49,1%) que os usuários muito favoráveis, podendo ser considerados uma espécie de “minoria barulhenta”. Porém, quando se observam as variações ao longo das semanas, é interessante

notar que a proporção de usuários contrários aumentou consideravelmente na segunda fase da CPI (semanas 12 a 22), o que pode indicar a perda do interesse no tópico por parte dos usuários favoráveis após o final da primeira fase.

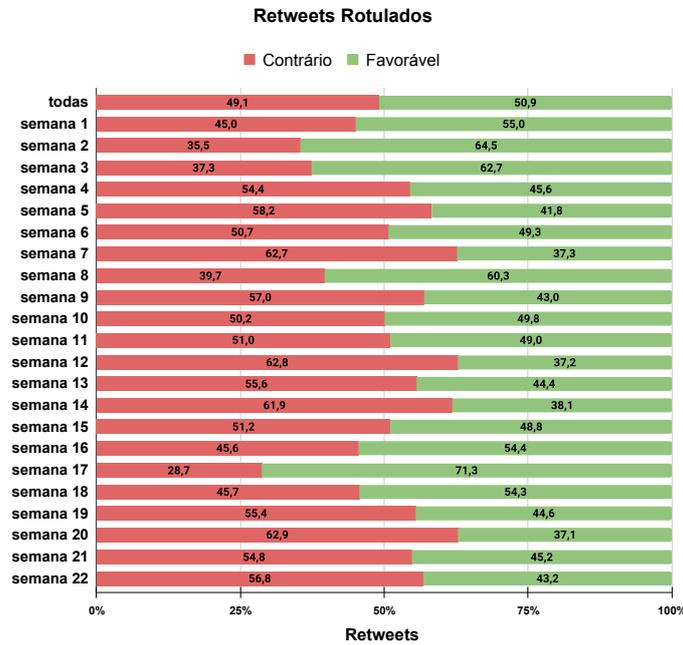


Figura 3. Distribuição dos *retweets* rotulados, por posicionamento e por semana

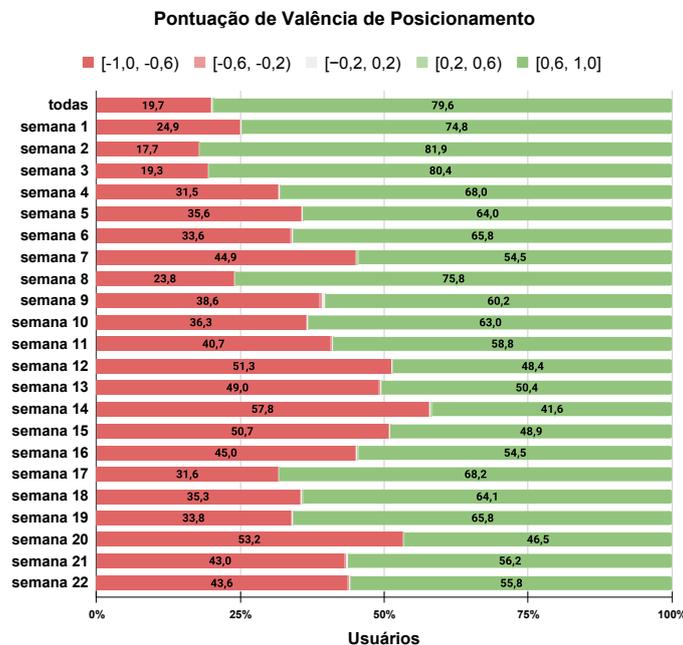


Figura 4. Distribuição dos usuários rotulados, por faixa de pontuação de valência de posicionamento e por semana

Verificou-se que 180.787 usuários participaram de ambas as fases da CPI (24% do total). Porém, não houve uma mudança significativa de posicionamento na grande maioria desses usuários: 89% dos usuários mantiveram a mesma pontuação de valência em ambas as fases. Nos usuários em que houve variação na pontuação de valência entre a primeira e a segunda fase, apenas 1.423 mudaram de faixa, tendo os demais tido uma variação muito pequena na pontuação. Em relação ao grau de engajamento, apenas 3% dos usuários foram mais engajados na segunda fase em relação à primeira.

4.3. Análise do comportamento dos usuários

Nesta subseção são analisadas as diferenças de inclinação e interesse entre os usuários mais e menos vocais usando duas medidas: o grau de equilíbrio e o grau de engajamento.

A Figura 5 mostra o grau de equilíbrio dos usuários rotulados. Um ponto importante diz respeito à frequência de usuários na faixa de posicionamento de grau 0: a distribuição de usuários rotulados nessa faixa é de apenas 0,5% na primeira fase e 0,3% na segunda fase, quando se consideram todos os usuários. Isso indica que maioria dos usuários convergiu para um dos extremos em ambas as fases. Ao analisar os usuários divididos entre mais e menos vocais essas proporções mudam: na primeira fase há uma proporção maior de usuários mais vocais favoráveis à CPI (mais que o dobro de usuários contrários), enquanto que na segunda fase essa diferença entre ambos se reduz para cerca de 1,2%. A diferença na distribuição de usuários menos vocais também muda da primeira para a segunda fase, embora os favoráveis continuem sendo maioria. Como a maioria dos usuários se concentrou em um dos extremos, esses percentuais reforçam a suposição de forte polarização no posicionamento dos usuários rotulados.

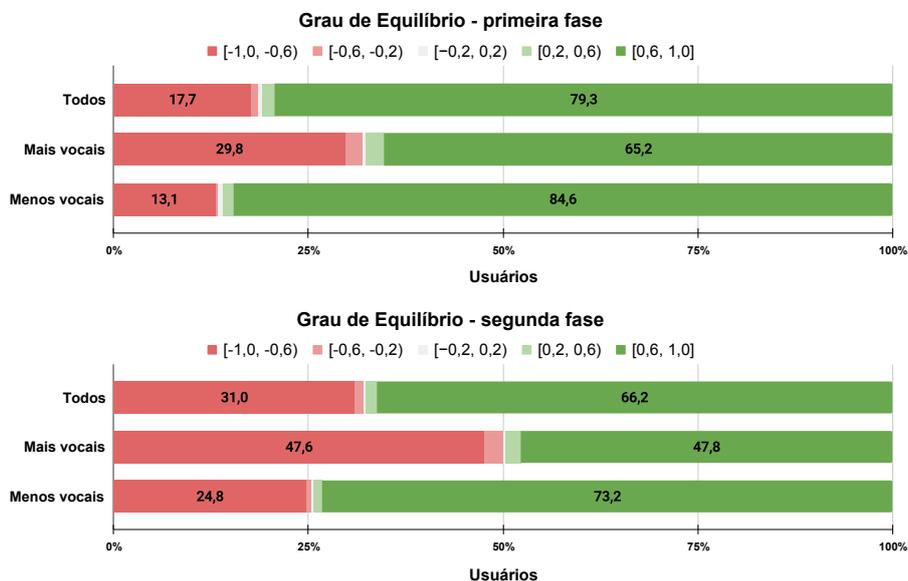


Figura 5. Distribuição do grau de equilíbrio por tipo de usuário

Em relação ao grau de engajamento (Figura 6), não houve variações significativas na primeira e segunda fase da CPI: quando se observa a totalidade dos usuários, 77% se

concentram na faixa com grau mais baixo, proporção que aumenta para 99% quando se consideram os usuários menos vocais. Em relação aos usuários mais vocais, a proporção de usuários é mais bem distribuída (com variação de menos de 1% nas duas fases) nas quatro faixas.

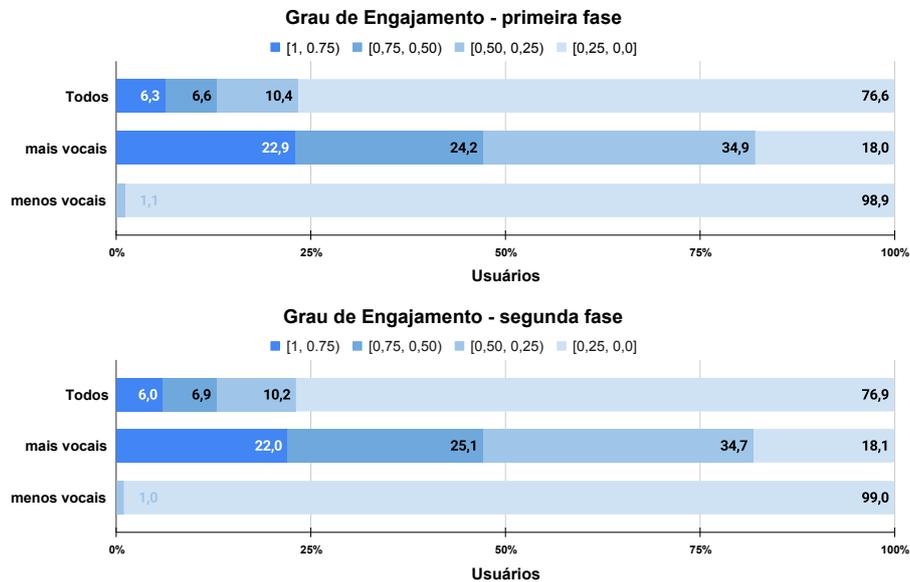


Figura 6. Distribuição do grau de engajamento dos usuários, por faixa

Quando se confrontam os graus de engajamento com os valores dos gráficos na Figura 7, os quais apontam que pouco mais da metade dos usuários foi rotulada em apenas 1 semana e 44% fizeram apenas 1 retweet, é possível afirmar que a maioria dos usuários pouco participou da discussão sobre a CPI, a qual foi movimentada pelos usuários mais vocais e engajados. Isso justifica nossa proposição da métrica engajamento: esses usuários que fizeram apenas um único retweet ou participaram em apenas uma semana possuem métrica engajamento igual a zero, que pode ser adequadamente confrontada com a pontuação de valência, para a realização de análises de posicionamento de usuários.

5. Avaliação da Qualidade dos Resultados Obtidos com o Método

A rotulação de usuários é uma tarefa complexa e subjetiva mesmo quando realizada de forma manual e caso a caso. Para avaliar a qualidade do método foram realizadas duas verificações: (i) checagem manual de pequena amostragem aleatória, realizada por uma pesquisadora, e (ii) comparação com rotulações manuais de outras bases de dados, com validação cruzada entre duas pesquisadoras. Ambas as abordagens são detalhadas nas subseções a seguir.

5.1. Checagem manual de pequena amostragem aleatória

Para analisar se o método proposto estava atribuindo rótulos adequados, foi selecionada uma amostra aleatória de mais de 100 usuários de todas as faixas de posicionamento, os

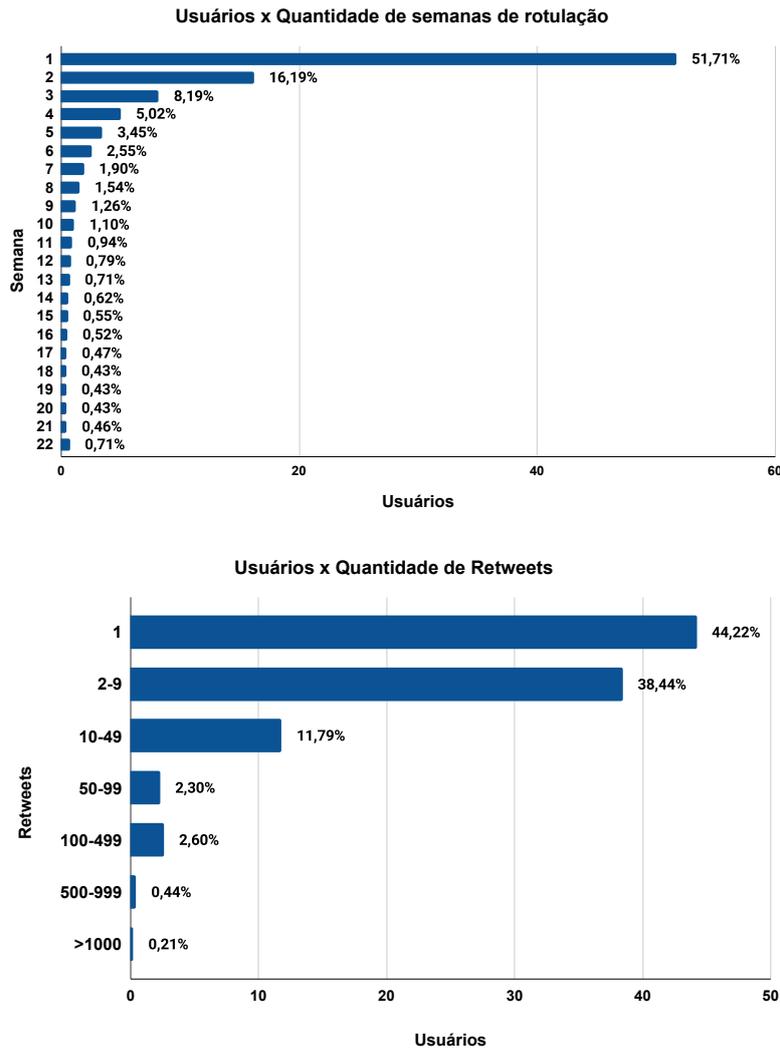


Figura 7. Distribuição de usuários de acordo com o número de semanas nas quais ele foi rotulado e com o número total de *retweets* feitos

quais foram verificados pelas autoras em forma de consenso. Para cada usuário da amostra foram analisados em conjunto: valência de posicionamento, grau de engajamento, texto do tweet original e contas retuitadas. Verificou-se que o método teve um excelente resultado, tendo atribuído rótulos adequados a todos os usuários da amostragem analisados. Na Tabela 5 é possível ver um exemplo de três usuários validados na semana 1. O nome dos usuários foi anonimizado na primeira coluna em respeito às políticas de privacidade do Twitter. É essencial ressaltar que houve uma fase subsequente de análise entre as duas pesquisadoras, com o objetivo de verificar se existia concordância nos julgamentos das verificações realizadas previamente.

Essa verificação manual permitiu observar empiricamente que quanto maior o engajamento, maior a probabilidade do método ter acertado (fato esperado em teoria). Outro ponto observado é que muitos dos usuários com pontuações de valência mais centrais fi-

zeram *retweets* para contas que foram muito retuitadas por ambos os clusters (aquelas próximas do intervalo de 20% que não foram descartadas, conforme explicado na etapa (ii) da seção 3.7).

Tabela 5. Amostra de usuários validados na semana 1 e retuites realizados

| Usuário | Conteúdo do tweet original retuitado pelo usuário | Conta |
|--|---|------------------|
| Usuário1 valência 1,0 engajamento 0,67 | Bolsonaro queria incluir covid na bula da cloroquina mesmo sabendo que era ineficaz? Por que Teich foi demitido em menos de um mês? Quantas vacinas foram descartadas? Hoje começa a CPI que pode derrubar o covarde genocida. | @GuilhermeBoulos |
| | Citado por Mandetta, Carlos Bolsonaro pode ser convocado à CPI do Genocídio | @revistaforum |
| Usuário2 valência -0,17 engajamento 0,75 | Presidente da CPI da Covid-19 já foi investigado até por pedofilia Em 2005, uma movimentação no Legislativo amazonense protegeu o então vice-governador Omar Aziz da acusação de exploração sexual infantil ! Vergonha ! #SenadoDaVergonha | @gouveia_kita |
| | Que a cloroquina não tem eficácia no tratamento da covid-19 todos com neurônios ativos sabem.O que não sabíamos, é que o presidente queria que a ANVISA alterasse a bula do remédio pra tratamento da doença. Quem declarou isso foi Mandetta, primeiro depoente na CPI da Covid. | @updatechartuc |
| Usuário3 valência -1,0 engajamento 0,89 | Só se surpreende com a CPI quem esperava algo diferente. Reforço o que disse antes: as CPIs só existem apenas para atrapalhar o governo e travar o Brasil. Nunca investigam, trata-se de um balcão de negócios, palanque de demagogia, circo de chantagens e muita corrupção. | @RafaelFontana |
| | As críticas contra a CPI da Covid-19 são muitas e devastadoras, a começar pela acusação de falta de imparcialidade atribuída a alguns dos seus integrantes. | @JornalDaCidadeO |

5.2. Comparação com rotulações manuais de outras bases de dados

Para validar a qualidade do método foi feita uma comparação entre os usuários rotulados automaticamente neste trabalho com rotulações feitas manualmente em [Kamiński et al. 2021] e [Kamiński et al. 2022]. Esses trabalhos foram escolhidos porque realizaram rotulações manuais de perfis do Twitter que se envolveram em discussões em contexto de polarização política no Brasil no ano de 2020 (período da pandemia pela Covid-19). O primeiro trabalho rotulou 147 usuários centrais em debates de 21 eventos, indicando suas inclinações entre Direita, Esquerda e Indefinida; o segundo considerou 33 eventos e rotulou 1850 usuários, com posicionamentos alinhados à Direita, Esquerda ou Indefinido. Os dois conjuntos de dados somaram 1863 usuários rotulados manualmente, dos quais 913 participaram das discussões da CPI da Covid-19 e puderam também ser rotulados automaticamente. Desses 913, 19 perfis se referiam a contas indisponíveis (suspensas, protegidas ou encerradas) e, portanto, foram removidas da comparação.

A Tabela 6 exemplifica esses dados com uma amostra de 10 usuários rotulados. As colunas Perfil, Posicionamento e Inclinação contêm as rotulações manuais e as colunas Valência e Engajamento, as rotulações automáticas.

Para proceder com a comparação das rotulações dos 894 perfis restantes, foi necessário optar por uma análise de consistência esperada, pois a rotulação manual não fora realizada explicitamente para indicar posicionamento em relação à CPI da Covid-19. Para

Tabela 6. Amostra de usuários rotulados manualmente em outras bases, em ordem decrescente de pontuação de valência de posicionamento e engajamento

| Usuário | [Kamienksi et al. 2022] | | [Kamienksi et al. 2021] | Este trabalho | |
|-----------------|-------------------------|----------------|-------------------------|---------------|-------------|
| | Perfil | Posicionamento | Inclinação | Valência | Engajamento |
| RachelSherazade | Jornalista | Direita | - | 1,0 | 0,67 |
| MarceloFreixo | Político | Esquerda | Esquerda | 1,0 | 0,62 |
| MBLivre | Grupo político | Direita | Direita | 1,0 | 0,48 |
| joicehasselmann | Político | Direita | - | 1,0 | 0,38 |
| SenadoFederal | Instituição | Indefinido | Indefinida | 0,72 | 0,57 |
| DaniloGentili | Celebridade | Direita | - | 0,70 | 0,24 |
| tabataamaralsp | Político | Esquerda | - | 0,49 | 0,15 |
| roxmo | Celebridade | Direita | - | -0,95 | 0,71 |
| kimpaim | Influenciador | Direita | - | -0,99 | 0,90 |
| taoquei1 | Influenciador | Direita | Direita | -1,0 | 0,86 |

tanto, observou-se conjuntamente as métricas obtidas automaticamente para valência, engajamento e equilíbrio, versus rótulos manuais para posicionamento e inclinação. Pelos resultados apresentados na Seção 5 e em especial da modelagem de tópicos (mais exemplos são apresentados em [Santos and Goya 2021]), é possível notar a clara tendência dos usuários contrários à CPI possuírem posicionamento à Direita, anticiência e/ou em apoio ao governo de Bolsonaro. Dessa forma, foram realizados os seguintes passos:

1. Verificação automática considerada consistente:
 - a. para valência de posicionamento contrário à CPI e rótulo Direita;
 - b. para valência de posicionamento favorável à CPI e rótulo Esquerda;
 - c. para engajamento acima de 0,25 e rótulo Indefinido;
2. Verificação manual sobre o restante, com potencial de ser considerado inconsistente:
 - a. dos usuários com rótulos Direita ou Esquerda;
 - b. dos usuários com rótulo Indefinido:
 - i. para valência de posicionamento contrário à CPI e engajamento até o limite inferior de 0,12, foi possível fazer uma avaliação manual;
 - ii. para valência de posicionamento favorável à CPI e engajamento até o limite inferior de 0,19, foi possível fazer uma avaliação manual;
 - iii. dos usuários restantes, optamos por não avaliar manualmente, pois com engajamento baixo na CPI, a incerteza foi alta.

O passo (1) acima cobriu cerca de 85% dos casos. No passo (2a) foi possível identificar quatro erros de rotulação manual e uma inconsistência de rotulação automática; os demais foram considerados consistentes (cerca de 5%). No passo (2b.iii), optamos por não avaliar manualmente 80 contas (8,95% das 894). A Tabela 7 resume esses resultados.

No caso da inconsistência da rotulação automática, o usuário possuía um grau de engajamento próximo de zero e fez apenas 2 *retweets* para uma conta institucional que recebeu muitos *retweets* de ambos os clusters. Esses *retweets* foram mais compartilhados pelo cluster oposto àquele que julgamos esse usuário pertencer, o que comprometeu a pontuação de valência desse usuário específico.

Tabela 7. Avaliação da consistência das rotulações

| Consistência das rotulações | Quantidade | % |
|--|------------|--------|
| Rotulação automática considerada consistente | 809 | 90,49% |
| Rotulações que não foram comparadas, devido ao alto grau de incerteza no julgamento sobre usuários que se manifestaram muito pouco | 80 | 8,95% |
| Erro de rotulação manual | 4 | 0,45% |
| Erro de rotulação automatizada | 1 | 0,11% |
| Total | 894 | 100% |

6. Ameaças à Validade

Esta seção discute alguns pontos a serem observados quanto à validade do estudo apresentado:

1. **A forma de definição dos grupos.** O uso de redes de *retweets* é uma abordagem muito usada para rotulação de grupos em redes sociais, porém pode introduzir vieses. Para mitigar esse problema, excluíram-se *retweets* com comentários das bases estudadas.
2. **A existência de grupos (minoritários) de usuários que não estejam discutindo temas relacionados ao tópico estudado.** Foi possível remover alguns desses grupos durante a etapa de detecção não supervisionada de posicionamento (eliminação de clusters minoritários), mas não é possível avaliar se todos foram removidos, sem uma complexa análise posterior. Além disso, considera-se que o seu impacto não é significativo dado que a maioria dos usuários da base retuita as mesmas contas, conforme verificado na seção 4.2.
3. **As contas que são retuitadas por ambos os lados.** E, portanto, percebidas como neutras foram tratadas igualmente àquelas que estiveram nas outras faixas de pontuação de valência, durante o cálculo das métricas dos usuários que as retuitaram. Dada a significativa polarização verificada no conjunto de bases, foi baixa a proporção percentual de perfis nessas condições, porém usuários com engajamento próximo de zero podem ter suas métricas enviesadas no caso de retuitarem tais contas consideradas neutras, caso elas sejam retuitadas por ambos os lados, mas mais fortemente por um dos polos.
4. **Possível subjetividade na checagem manual de rótulos de usuários.** Esses dados poderiam ser avaliados por mais pesquisadores, utilizando métricas de cálculo de índice de concordância, como o Coeficiente de Concordância de Kappa, por exemplo.
5. Um quinto ponto não se trata exatamente de ameaça à validade do método, mas do **risco do uso das métricas de valência e equilíbrio de usuários com engajamento igual a zero** (ou muito próximo desse valor), sem o devido cuidado. A incerteza sobre a precisão da pontuação de valência nesses casos é alta, porém pretendemos combinar as rotulações obtidas nesse caso CPI da Covid-19 com as de outras bases de nosso projeto (são mais de 100 bases de interesse). Desse modo, ainda que alguns usuários pouco se manifestem em determinados eventos, ao observarmos dados em múltiplos contextos, talvez seja possível obter informações relevantes de posicionamento.

6. **A possível existência de bots.** Foi utilizada a API Botometer⁴ para detectar a existência de *bots* em um recorte de 3.776 usuários que tinham maior probabilidade de ter rótulos divergentes (pontuação de valência entre [-0,2, +0,2]). Verificamos que 1.019 contas foram suspensas ou deletadas pelo próprio usuário, então ao final foi possível analisar um total de 2.757 usuários, dos quais 954 (34,6%) possuía em torno de 70% de probabilidade de serem *bots*. Como esse recorte consistiu apenas de usuários com grau de equilíbrio muito baixo, um estudo em usuários de outras faixas se faz necessário.

7. Conclusão e Trabalhos Futuros

Esse trabalho propôs um método de seis etapas para rotulação automatizada de posicionamento de usuários de forma a caracterizar a opinião pública sobre um tema político e controverso. Para obter a resposta à questão de pesquisa **QP**, esse método foi aplicado em um extenso conjunto de bases de dados coletados em português brasileiro, cujo tema gerou polarização e polêmica na opinião pública.

A análise mostrou que o método é bastante efetivo, tendo sido possível atribuir um rótulo para a quase totalidade dos usuários analisados. É importante destacar a diferença de comportamento dos usuários em relação ao posicionamento aferido: os usuários muito contrários à CPI, apesar de estarem em menor número, compartilharam em média uma quantidade maior de *retweets* do que os usuários favoráveis, cuja proporção é significativamente maior. Além disso, ao analisar a dimensão temporal dos dados coletados e o posicionamento dos usuários ao longo das semanas, é possível inferir prováveis mudanças da opinião pública, possivelmente associadas a eventos relacionados às apurações da CPI.

Este estudo oferece informações úteis para pesquisadores de áreas como computação social e ciências sociais que desejam utilizar a detecção de posicionamento como meio de avaliar a opinião pública a partir de dados do Twitter, mas que não dispõem de tempo ou recursos para realizar rotulações manuais. Adicionalmente, levando em conta que há uma escassez de bases de dados anotadas em português para posicionamento de usuários, um benefício secundário pode ser observado no incremento de recursos linguísticos disponíveis para a comunidade científica atuante nesse campo.

Como trabalho futuro, deseja-se avaliar possíveis refinamentos do método, em particular verificar se o uso de ponderações em contas neutras durante o cálculo de valência gera impactos positivos aos padrões gerais identificados. Almeja-se também ampliar o leque de características analisadas pelo método, como análise das principais narrativas de cada cluster. Além disso, considera-se interessante explorar novas métricas que possam apoiar análises qualitativas sobre os eventos, introduzindo conceitos como características de fluxo de informações entre os clusters e entendimento da dinâmica de rede de cada grupo de interesse. Por fim, pretende-se avaliar o desempenho do método em bases de dados relacionadas a outros contextos de polarização.

⁴<https://botometer.osome.iu.edu/>

Referências

- [AlDayel and Magdy 2021] AlDayel, A. and Magdy, W. (2021). Stance detection on social media: State of the art and trends. *Information Processing & Management*, 58(4):102597.
- [Alturayeif et al. 2023] Alturayeif, N., Luqman, H., and Ahmed, M. (2023). A systematic review of machine learning techniques for stance detection and its applications. *Neural Computing and Applications*, pages 1–32.
- [Awadallah et al. 2012] Awadallah, R., Ramanath, M., and Weikum, G. (2012). Harmony and dissonance: organizing the people’s voices on political controversies. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 523–532.
- [Barros and do Vale 2021] Barros, C. C. and do Vale, R. P. G. (2021). ”tchau, pfizer!”: Uma análise discursiva de charges publicadas durante a comissão parlamentar de inquérito da covid-19. *Revista de Ciências Humanas*, 3(21).
- [Bechini et al. 2020] Bechini, A., Ducange, P., Marcelloni, F., and Renda, A. (2020). Stance analysis of twitter users: the case of the vaccination topic in italy. *IEEE Intelligent Systems*, 36(5):131–139.
- [Christhie et al. 2018] Christhie, W., Reis, J. C., Moro, F. B. M. M., and Almeida, V. (2018). Detecção de posicionamento em tweets sobre política no contexto brasileiro. In *Anais do VII Brazilian Workshop on Social Network Analysis and Mining*. SBC.
- [Cinelli et al. 2020] Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., and Starnini, M. (2020). Echo chambers on social media: A comparative analysis. *arXiv preprint arXiv:2004.09603*.
- [Conover et al. 2011] Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., and Flammini, A. (2011). Political polarization on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, pages 89–96.
- [D’Andrea et al. 2019] D’Andrea, E., Ducange, P., Bechini, A., Renda, A., and Marcelloni, F. (2019). Monitoring the public opinion about the vaccination topic from tweets analysis. *Expert Systems with Applications*, 116:209–226.
- [Dantas and Nippes 2022] Dantas, A. and Nippes, G. (2022). Quando as máscaras vão cair? *Revista Pet Economia UFES*, 2(2):32–36.
- [Darwish et al. 2020] Darwish, K., Stefanov, P., Aupetit, M., and Nakov, P. (2020). Unsupervised user stance detection on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 141–152.
- [de Carvalho Mendes et al. 2022] de Carvalho Mendes, G. P., Orso, M., and Alves, M. K. F. (2022). A cpi da covid-19 sob a ótica da extrema-direita: análise do perfil brazilfight no twitter. *Tríade: Comunicação, Cultura e Mídia*, 10(23):e022024–e022024.
- [Dori-Hacohen et al. 2015] Dori-Hacohen, S., Yom-Tov, E., and Allan, J. (2015). Navigating controversy as a complex search task. In *SCST@ ECIR*. Citeseer.

- [Ebeling et al. 2020] Ebeling, R., Sáenz, C. A. C., Nobre, J., and Becker, K. (2020). Quaranteners vs. chloroquiners: A framework to analyze how political polarization affects the behavior of groups. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 203–210. IEEE.
- [Figeac and Favre 2021] Figeac, J. and Favre, G. (2021). How behavioral homophily on social media influences the perception of tie-strengthening within young adults’ personal networks. *New Media & Society*, page 14614448211020691.
- [Grootendorst 2022] Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv e-prints*, pages arXiv–2203.
- [Jungherr et al. 2017] Jungherr, A., Schoen, H., Posegga, O., and Jürgens, P. (2017). Digital trace data in the study of public opinion: An indicator of attention toward politics rather than political support. *Social Science Computer Review*, 35(3):336–356.
- [Kamienski et al. 2022] Kamienski, C., de Camargo Penteadó, C. L., Goya, D., Rocha, R. V., de Souza, L. M., di Gênova, D. V. B., Ramos, D. F. S., de França, F. O., Horita, F., and dos Santos, C. d. S. (2022). Measuring network polarization and political sectarianism during the 2020 pandemic. *IEEE Transactions on Computational Social Systems*.
- [Kamienski et al. 2021] Kamienski, C., Mazim, L., Penteadó, C., Goya, D., Di Genova, D., De Franca, F., Ramos, D., and Horita, F. (2021). A polarization approach for understanding online conflicts in times of pandemic: A brazilian case study.
- [Küçük and Can 2020] Küçük, D. and Can, F. (2020). Stance detection: A survey. *ACM Computing Surveys (CSUR)*, 53(1):1–37.
- [Li et al. 2021] Li, Y., Zhao, C., and Caragea, C. (2021). Improving stance detection with multi-dataset learning and knowledge distillation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6332–6345.
- [Lin et al. 2017] Lin, J., Mao, W., and Zhang, Y. (2017). An enhanced topic modeling approach to multiple stance identification. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 2167–2170.
- [Magdy et al. 2016] Magdy, W., Darwish, K., and Weber, I. (2016). # failedrevolutions: Using twitter to study the antecedents of isis support. In *2016 AAAI Spring Symposium Series*.
- [Maia et al. 2021] Maia, M., Oliveira, E., and Gallegos, L. (2021). Covid-19 e tweets no brasil: coleta, tratamento e análise de textos com evidências de estados afetivos alterados em momentos impactantes. In *Anais do X Brazilian Workshop on Social Network Analysis and Mining*, pages 79–90. SBC.
- [Malagoli et al. 2021] Malagoli, L., Stancioli, J., Ferreira, C., Vasconcelos, M., Silva, A. P., and Almeida, J. (2021). Caracterização do debate no twitter sobre a vacinação contra a covid-19 no brasil. In *Anais do X Brazilian Workshop on Social Network Analysis and Mining*, pages 55–66, Porto Alegre, RS, Brasil. SBC.

- [McInnes et al. 2018] McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29).
- [McPherson et al. 2001] McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444.
- [Mohammad et al. 2016] Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., and Cherry, C. (2016). Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 31–41.
- [Mohammad et al. 2017] Mohammad, S. M., Sobhani, P., and Kiritchenko, S. (2017). Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT)*, 17(3):1–23.
- [Morosini 2022] Morosini, C. (2022). Discursos em conflito: estratégias bolsonaristas para deslegitimar a comissão parlamentar de inquérito (cpi) da pandemia através do twitter. *Revista Investigações*, 35(2):1–26.
- [Popat et al. 2019] Popat, K., Mukherjee, S., Yates, A., and Weikum, G. (2019). Stancy: Stance classification based on consistency cues. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6413–6418.
- [Rashed et al. 2021] Rashed, A., Kutlu, M., Darwish, K., Elsayed, T., and Bayrak, C. (2021). Embeddings-based clustering for target specific stances: The case of a polarized turkey. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 537–548.
- [Sáenz and Becker 2021] Sáenz, C. A. C. and Becker, K. (2021). Interpreting bert-based stance classification: a case study about the brazilian covid vaccination. In *Anais do XXXVI Simpósio Brasileiro de Bancos de Dados*, pages 73–84. SBC.
- [Samih and Darwish 2021] Samih, Y. and Darwish, K. (2021). A few topical tweets are enough for effective user stance detection. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2637–2646.
- [Santos 2022] Santos, A. (2022). Reflexões sobre a importância do pânico sexual para a ascensão do bolsonarismo ao poder. *Lumina*, 16(3):92–111.
- [Santos and Goya 2022] Santos, P. and Goya, D. (2022). Detecção de posicionamento e rotulação automática de usuários do twitter: estudo sobre o embate científico-político no contexto da cpi da covid-19. In *Anais do XI Brazilian Workshop on Social Network Analysis and Mining*, pages 49–60, Porto Alegre, RS, Brasil. SBC.
- [Santos and Goya 2021] Santos, P. D. and Goya, D. H. (2021). Automatic twitter stance detection on politically controversial issues: A study on covid-19’s cpi. In *Anais do XVIII Encontro Nacional de Inteligência Artificial e Computacional*, pages 524–535. SBC.

- [Sirrianni et al. 2021] Sirrianni, J. W., Liu, X., and Adams, D. (2021). Predicting stance polarity and intensity in cyber argumentation with deep bidirectional transformers. *IEEE Transactions on Computational Social Systems*, 8(3):655–667.
- [Sobhani et al. 2015] Sobhani, P., Inkpen, D., and Matwin, S. (2015). From argumentation mining to stance classification. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 67–77.
- [Sobhani et al. 2017] Sobhani, P., Inkpen, D., and Zhu, X. (2017). A dataset for multi-target stance detection. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 551–557.
- [Souza et al. 2020] Souza, F., Nogueira, R., and Lotufo, R. (2020). BERTimbau: pretrained BERT models for Brazilian Portuguese. In *9th Brazilian Conference on Intelligent Systems, BRACIS, Rio Grande do Sul, Brazil, October 20-23 (to appear)*.
- [Stefanov et al. 2020] Stefanov, P., Darwish, K., Atanasov, A., and Nakov, P. (2020). Predicting the topical stance and political leaning of media using tweets. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 527–537.
- [Vamvas and Sennrich 2020] Vamvas, J. and Sennrich, R. (2020). X-stance: A multilingual multi-target dataset for stance detection. In *5th SwissText & 16th KONVENS Joint Conference 2020*, page 9. CEUR-WS. org.
- [Wagner Filho et al. 2018] Wagner Filho, J. A., Wilkens, R., Idiart, M., and Villavicencio, A. (2018). The brwac corpus: A new open resource for brazilian portuguese. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*.
- [Wischniewski et al. 2022] Wischniewski, M., Ngo, T., Bernemann, R., Jansen, M., and Krämer, N. (2022). “i agree with you, bot!” how users (dis) engage with social bots on twitter. *New Media & Society*, page 14614448211072307.
- [Wojatzki and Zesch 2016] Wojatzki, M. and Zesch, T. (2016). Itl. uni-due at semeval-2016 task 6: Stance detection in social media using stacked classifiers. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 428–433.
- [Zarrella and Marsh 2016] Zarrella, G. and Marsh, A. (2016). Mitre at semeval-2016 task 6: Transfer learning for stance detection. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 458–463.