

Caracterizando Polarização em Redes Sociais: Uma Análise Comparativa das Discussões no Reddit das Eleições Brasileiras de 2018 e 2022

Gustavo F. Cunha¹, Ana Paula C. da Silva¹

¹Departamento de Ciência da Computação – Instituto de Ciências Exatas –
Universidade Federal de Minas Gerais (UFMG)

{gustavocunha, ana.coutosilva}@dcc.ufmg.br

Abstract. *Given their widespread use, online social media have become increasingly important in the study of social phenomena, including polarization. In this work, we present a methodology for characterizing polarized discussions on social media, which is based on the combination and adaptation of several techniques previously proposed in the literature. As a case study, we collected threads of discussions from Reddit during the Brazilian presidential elections in 2018 and 2022 and conducted an analysis of the degree of political polarization in each period. The results show a trend of increasing polarization and decreasing diversity in the debates conducted by these users between the analyzed periods.*

Resumo. *Dada a popularização de seu uso, as redes sociais on-line têm se tornado cada vez mais importantes no estudo de fenômenos sociais, incluindo a polarização. Neste trabalho, apresentamos uma metodologia baseada na combinação e adaptação de técnicas previamente propostas na literatura para caracterizar discussões polarizadas em redes sociais. Como estudo de caso, foram coletadas threads de discussão no Reddit durante as eleições presidenciais brasileiras de 2018 e 2022 e conduzidas análises do grau de polarização política em cada período. Os resultados mostraram uma tendência de crescimento da polarização e uma diminuição da diversidade nos debates conduzidos pelos usuários nos períodos analisados.*

1. Introdução

Nos últimos anos, as mídias sociais ganharam importância como canal de influência de comportamento e decisão das pessoas, afetando não apenas o mundo on-line, mas também a vida real, no mundo offline [Kim et al. 2013]. Isto é especialmente evidente no Brasil, com destaque no contexto político, onde as mídias sociais se tornaram um meio efetivo para busca de informações e conhecimento sobre candidatos. Os candidatos, por sua vez, começaram também a fazer uso das plataformas digitais objetivando disseminar suas ideias e angariar votos, além de atacar seus oponentes [Guimaraes et al. 2022, Caetano et al. 2022].

O debate político se destaca por seu aspecto controverso e cada vez mais polarizado no Brasil e no mundo, como evidenciado por [Ruediger et al. 2014] e [Heltzel and Laurin 2020], acarretando consequências como perda da diversidade de opiniões e da pluralidade política [Rossini 2020], além do comprometimento de

princípios democráticos e polarização afetiva, causando um aumento da hostilidade e divisão social [Sunstein 2018, Levin et al. 2021]. Este ciclo vicioso é fortalecido constantemente, dado que conteúdos com alto nível de polarização estão frequentemente relacionados a discurso de ódio, propagação de desinformação e informações enviesadas [Vicario et al. 2019]. É sabido que a polarização é contributiva para a democracia, quando tida na medida correta, como atestam [Heltzel and Laurin 2020, Mutz 2002, Dahlberg 2007], mas pode trazer mefeícios ao coletivo ao atingir níveis perigosos.

Sob esta perspectiva, a caracterização do debate político nas mídias sociais é primordial para a proposta de plataformas mais plurais, com regras eficazes de moderação para casos onde a discussão política ultrapasse os limites democráticos. Vários trabalhos na literatura apresentam este tipo de caracterização em plataformas de interação on-line, em diferentes contextos. No cenário político brasileiro, trabalhos recentes [Recuero et al. 2020, Brum et al. 2022, Franco and Pound 2022], delinham a polarização em diferentes redes sociais que, por possuírem diferentes características e particularidades, têm metodologias, bem como resultados, mais ou menos acurados.

A principal contribuição da pesquisa aqui conduzida é a proposição de uma metodologia para caracterização de discussões polarizadas em mídias sociais, que se baseia na combinação e adaptação de algumas técnicas previamente propostas na literatura [Garimella et al. 2018, Alsinet et al. 2021a, de Arruda et al. 2022]. Para isso, como estudo de caso, foram analisadas as discussões realizadas nas comunidades do Reddit¹, uma rede social organizada em subcomunidades por áreas de interesse (*subreddits*). Nesta plataforma, os usuários discutem diferentes assuntos, através de interações do tipo postagem-comentários, chamadas de *threads*. Nas postagens e comentários também é possível que os internautas dêem votos positivos - *upvotes* - e votos negativos, chamados de *downvotes*. Esta rede social foi classificada como o 19º website mais acessado do mundo em 2023 e é composta por milhares de comunidades [Britt et al. 2023]. Além disso, por possuir um limite insignificante de caracteres nas postagens (40.000), a plataforma é rica em discussões políticas, por permitir que os usuários expressem suas opiniões de maneira mais ampla. O conjunto de dados utilizado consiste em atividades de usuários (postagens e comentários) que ocorreram entre os meses de setembro e novembro de 2018 e 2022, período que engloba pré e pós eleições no Brasil.

Os resultados das análises iniciais se mostraram promissores para a detecção de polarização no conteúdo-alvo da caracterização, independentemente do tópico abordado. Os resultados evidenciam uma tendência de aumento da polarização e diminuição do caráter de diversidade dos debates realizados por estes usuários entre os períodos analisados, corroborando discussões abordadas em trabalhos de diferentes campos de pesquisa, como [Ortellado et al. 2022, Corrêa 2023, Silva 2023, Lima 2021].

A estrutura deste trabalho está organizada como descrito a seguir. Na Seção 2, são apresentados, de maneira não exaustiva, os principais estudos relacionados a esta pesquisa, sendo detalhados aqueles que serviram de base para a metodologia proposta. Na Seção 3, descreve-se o conjunto de dados utilizados, bem como sua modelagem matemática e como é feita a detecção de posição dos comentários. Em seguida, na Seção 4, caracterizam-se os *insights* obtidos, bem como as limitações do *framework* utilizado. Por

¹<http://www.reddit.com/>

fim, a Seção 5 finaliza a monografia resumindo as contribuições do trabalho e tratando de direcionamentos para dar continuidade ao estudo.

2. Trabalhos Relacionados

Os trabalhos da literatura que abordam o estudo da polarização em discussões em mídias sociais, em sua maioria, apresentam contribuições em três tópicos principais: (i) modelo que represente as interações entre os usuários; (ii) técnicas para a detecção de posicionamento dos usuários (*stance detection*) e; (iii) a caracterização da polarização para uma mídia social de grande interesse. Nesta seção descrevemos os trabalhos em maior consonância com a proposta deste estudo e com a atualidade, bem como aqueles que foram base para a metodologia proposta.

Em [Garimella et al. 2018], os autores apresentaram uma abordagem completa para quantificar a polarização no X (então Twitter). Foram exploradas três modelagens grafos distintas, onde as interações entre usuários representavam *retweets*, ou seguidores ou conteúdo compartilhado. Para cada um dos grafos modelados, os autores realizaram o particionamento em dois grupos opostos de discussão, utilizando o algoritmo METIS [Karypis 1997]. Em seguida, introduziram métricas como *Random Walk Controversy* (RWC) e *Betweenness* para quantificação da controvérsia entre os grupos, tendo como base propriedades matemáticas dos grafos. Para isso, foi verificado o quão provável seria a postagem de um usuário em um grupo chegar até o conhecimento de um usuário do grupo oposto. Complementarmente, foi aplicada uma análise de sentimentos ao conteúdo compartilhado e verificou-se que tópicos mais polarizados possuem uma variância maior nos sentimentos dos conteúdos quando comparados àqueles de menor polarização.

Sob ótica contextual, [Kumar et al. 2022] investigou como se comportam contas consideradas tóxicas ao longo do tempo no Reddit. A análise foi realizada através da caracterização de como esses usuários interagem com os demais, explorando aspectos como frequência e intensidade de comportamentos tóxicos, bem como possíveis mudanças nesses padrões. A metodologia utiliza a API da Google: Perspective² - já utilizada em vários outros trabalhos prévios, como [Hua et al. 2020], [Xia et al. 2020] e [Saveski et al. 2021], para avaliar a probabilidade de toxicidade de um conteúdo textual no intervalo $[0, 1]$. O estudo conclui que apenas uma pequena fração das contas são abusivas (3,1%), mas que essas contas, juntas, são responsáveis por um terço de todos os comentários da plataforma e, com isso, mais da metade dos usuários da rede social já interagiram em pelo menos uma *thread* com conteúdo tóxico.

Já em [Alsinet et al. 2021a], os autores apresentaram um modelo quantitativo baseado no comportamento do usuário para avaliar o nível de polarização em diferentes tópicos de discussão no Reddit, em língua inglesa. Cada debate foi modelado como uma árvore de discussão de dois lados, na qual a raiz é um comentário inicial de uma discussão e os demais vértices são os comentários da mesma *thread*. As arestas da árvore são modeladas com base na cadeia de respostas do debate e cada comentário é submetido a uma análise de sentimentos que retorna um valor no intervalo $[-4, 4]$. A árvore é então dividida em dois lados de posicionamento, através de uma bipartição. Por fim, é proposta uma medida de polarização que passa por avaliações empíricas, cujos resultados indicam que o modelo é capaz de capturar os diferentes níveis de polarização em assuntos diversos.

²<https://perspectiveapi.com>

Em direção complementar, o trabalho apresentado por [de Arruda et al. 2022] utilizou um conjunto de modelos sintéticos de grafos para representar conexões em mídias sociais. Nos grafos gerados, cada usuário (vértice) possui uma opinião em relação a um tópico em discussão. Esta opinião varia no intervalo $[-1, 1]$, sendo -1 um posicionamento extremamente contrário e 1 um posicionamento a favor do debate estabelecido entre os vizinhos no grafo (que representam os relacionamentos na rede social). Os usuários são divididos em dois grupos: (i) os que possuem posicionamento negativo e (ii) os que possuem posicionamento positivo. A partir da definição da métrica *balance*, que é a razão da cardinalidade destes grupos, os autores caracterizam diferentes cenários onde a discussão entre os usuários é mais aberta a ser realizada por usuários de opiniões diversas e os casos onde os usuários se organizam de forma polarizada.

No cenário brasileiro, [Recuero et al. 2020] examinaram o papel do hiperpartidarismo e da polarização no Twitter durante as eleições presidenciais de 2018, mostrando que há uma forte conexão entre polarização, hiperpartidarismo e desinformação. Em [Brum et al. 2022], os autores quantificaram a polarização política do público no contexto da pandemia de COVID-19. Também [Franco and Pound 2022] exploraram, com dados do Instagram, se fatores psicológicos poderiam contribuir para o apelo de uma figura polarizadora como, por exemplo, Jair Bolsonaro.

Os resultados apresentados neste trabalho se diferenciam dos demais nos aspectos a seguir. Foram combinados e adaptados os modelos de interação apresentados em [Garimella et al. 2018, Alsinet et al. 2021a, de Arruda et al. 2022]. Para a detecção de posicionamento, substituiu-se a análise de sentimentos que, segundo alguns trabalhos, não é a melhor técnica a ser utilizada [ALDayel and Magdy 2021], pelo uso de *Large Language Models* (LLMs), mais precisamente o *ChatGPT*, que se mostrou adequado para a tarefa no contexto da língua inglesa [Zhang et al. 2022, Zhang et al. 2023]. Para língua portuguesa, por exemplo, o trabalho de [Seno et al. 2024], utilizou-se desta mesma LLM para uma tarefa diferente, com resultados promissores. Uma vez que a maioria dos trabalhos no contexto brasileiro utilizam dados do Instagram e X, com enfoque nas eleições presidenciais de 2018, a contribuição do presente estudo é analisar as interações no Reddit, comparando os cenários de possível polarização nas eleições presidenciais de 2018 e 2022.

3. Metodologia

A seguir descrevemos a coleta do conjunto de dados, a modelagem matemática das discussões e o detalhamento do método utilizado para mensurar o nível de polarização nos debates analisados.

3.1. Conjunto de Dados

Reddit é uma mídia social on-line multilíngue, fundada em 2005. Nesta rede, subdividida em comunidades, debates nos mais diversos contextos e dos mais variados tipos têm lugar. Seleccionamos postagens e comentários entre os meses de setembro e novembro de 2018 e 2022, no contexto das discussões sobre as eleições presidenciais brasileiras.

A seleção dos *subreddits* mais relevantes, em total de publicações para o contexto analisado, foi realizada utilizando a pesquisa por relevância na API Python do Reddit³.

³<https://praw.readthedocs.io/>

A partir dos *subreddits* de maior relevância, foram coletadas 95.933 postagens datadas de setembro a novembro de 2018 e 2022. Um subconjunto destas postagens foi selecionado, com a presença pelo menos de uma das palavras-chave relacionadas ao contexto das eleições tais como *eleições, política, Lula, Bolsonaro, Haddad, esquerda, direita, PT, PSL, PL*, resultando em 20.515 postagens. Por fim, retiramos também as postagens com conteúdo deletado/removido, escritas em outra língua e/ou constituídas somente por *links*. Finalmente, os comentários realizados nas 10.578 postagens restantes foram coletados. A Tabela 1 apresenta os *subreddits* selecionados, a quantidade de postagens e comentários coletados e o número de usuários envolvidos nos debates analisados.

Subreddit	#Postagens	#Comentários	#Usuários
r/brasil	5.895	139.651	16.273
r/BrasilDoB	828	3.357	1.207
r/brasilivre	3.855	61.288	6.346
Total	10.578	204.296	23.826

Tabela 1. Estatísticas dos subreddits selecionados.

3.2. Modelo Matemático

As discussões realizadas pelos usuários foram modeladas através de árvores, i.e., grafos sem ciclos. Cada postagem inicial, denominada p_j , com $0 \leq j \leq P$ é considerada um contexto de discussão e pode receber R comentários como resposta.

Cada resposta (comentário) r_i , com $0 \leq i \leq R$, associada à uma postagem p_j , gera uma árvore $T = (V, E)^4$, onde V é o conjunto da cadeia de comentários que sucedem r_i (inclusive) e E é o conjunto de arestas direcionadas (v_k, v_l) que conectam um comentário v_k (origem) à sua resposta v_l (destino). A cada uma das arestas direcionadas, associamos um peso $\sigma_{kl} \in [-1, 1]$, calculado a partir da técnica de detecção de posicionamento descrita na Seção 3.3. Resumidamente, se v_l é o comentário resposta à v_k , σ_{kl} assume valores negativos se v_l possui um posicionamento contra ao de v_k e valores positivos se o posicionamento é a favor ao de v_k . Os casos com $\sigma_{kl} = 0$ são os de posicionamento neutro, mas aqui considerados como levemente negativos.

Com os pesos definidos, é realizada a bipartição de cada árvore em dois subgrupos disjuntos de posicionamento A e B , analisando a polarização em torno de dois polos de pensamento. A raiz r_i é atribuída a um dos lados, que convencionamos, sem perda de generalidade, ser o lado A . Para os vértices v_k de nível 1, e para $\sigma_{i,k} > 0$, v_k é adicionado ao lado A e para $\sigma_{i,k} < 0$ ele é adicionado ao lado B . Para o caso de $\sigma_{i,k} = 0$, v_k é interpretado como levemente mais negativo do que positivo, assim, sendo assinalado também ao lado B . Isso porque, segundo a hipótese de [Agrawal et al. 2003], as pessoas tendem a responder mais quando discordam do que quando concordam. Também, em [Alsinet et al. 2021b], é dito que no Reddit, por exemplo, quando o usuário concorda com um comentário, ele pode simplesmente deixar um *upvote* sem responder, mas, por outro lado, quando discorda, é mais provável que exponha suas razões. Este processo se repete a cada novo nível da árvore, considerando pares pai-filho de comentários.

⁴Para facilitar a leitura da notação, iremos desconsiderar o índice que identifica cada árvore associada a cada resposta.

3.3. Análise da Polarização

Um dos pontos mais importantes para a análise da polarização em discussões on-line é definir o posicionamento de um usuário com base na sua resposta a um comentário. Ou seja, o objetivo é medir o quanto uma resposta concorda ou discorda de um comentário. A partir dos posicionamentos de todas as respostas, construímos uma árvore bipartida, com dois subgrupos distintos de posicionamento.

Experimentalmente, testamos a aplicação de uma análise de sentimentos aos comentários, com a biblioteca VADER [Hutto and Gilbert 2014], cuja adaptação para a língua portuguesa foi realizada por [Almeida 2018], que provém um score no intervalo $[-1, 1]$, onde 1 é extremamente positivo e -1 extremamente negativo. Contudo, este método não se mostrou eficaz ao aplicar a bipartição gráfica, uma vez que, analisando manualmente as árvores, percebemos que haviam muitos pares de comentários claramente discordantes que estavam alocados ao mesmo lado do debate e, similarmente, comentários concordantes foram postos em lados distintos na *thread*. Endossando, assim, alguns resultados de estudos prévios que não tiveram sucesso com a análise de sentimentos para este propósito [ALDayel and Magdy 2021, Kobbe et al. 2020, Küçük and Can 2020].

Conforme apresentado em [ALDayel and Magdy 2021] e pelos resultados obtidos com os testes realizados no conjunto de dados, o uso de técnicas de análise de sentimentos não é a melhor escolha para a detecção de posicionamento. Assim, em consonância com os trabalhos em [Zhang et al. 2022, Zhang et al. 2023], aplicamos o modelo GPT-4 [Achiam et al. 2023] para realizar esta tarefa, a partir do uso da API da OpenAI.⁵ A Figura 1 apresenta o *prompt* usado na tarefa. O *score* da resposta que representa o seu posicionamento em relação ao comentário inicial varia também entre $[-1, 1]$, com -1 para os casos extremamente discordantes e 1 para os casos extremamente concordantes. Este valor é aqui denominado *Stance Score*. A Figura 2 apresenta um exemplo de uma árvore bipartida a partir da definição dos posicionamentos. As arestas são coloridas em três cores: vermelho, amarelo e verde, que representam, respectivamente, peso negativo, peso igual à zero e peso positivo. Já a coloração dos vértices representa os dois lados, *A* e *B*, da discussão, de acordo com a bipartição discutida na Seção 3.2. Após a definição do posicionamento para um conjunto inicial de árvores, realizamos uma validação manual dos resultados, que mostraram-se adequados para a execução da tarefa proposta.

Levando em conta que o modelo GPT-4 requer o pagamento⁶ de requisições à sua API e que este trabalho é um estudo inicial da metodologia proposta, realizamos uma amostragem dos dados a serem anotados pela ferramenta. Os dados amostrados mantiveram a representatividade de cada comunidade neste subconjunto. A Tabela 2 apresenta o total de dados analisados nesta fase.

Com a definição das árvores bipartidas de discussão, o próximo passo é verificar a existência e quantificar a polarização nas discussões. Para tal, calculamos as seguintes métricas:

- **Desvio-Padrão e Média do Stance Score.** Seguindo [Garimella et al. 2018], discursos polarizados possuem maior variância de *Stance Score* das arestas da árvore de discussão. Assim, adaptamos o ponto de corte do desvio-padrão proposto no

⁵<https://platform.openai.com/>

⁶<https://openai.com/api/pricing/>

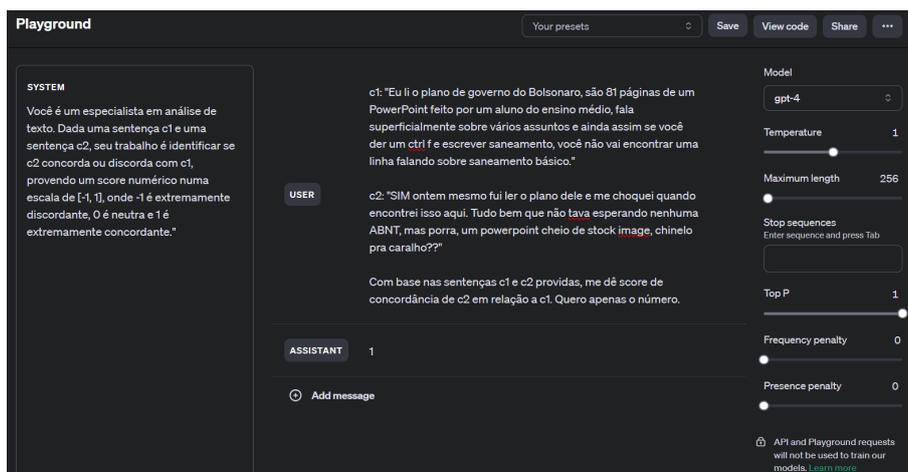


Figura 1. *Prompt* utilizado para detecção de posicionamento usando o GPT-4.

Subreddit	#Postagens (18)	#Postagens (22)	#Comentários (18)	#Comentários (22)	#Árvores (18)	#Árvores (22)
r/brasil	8	7	2.292	2.447	175	237
r/BrasildoB	1	1	3	54	1	5
r/brasiliivre	5	5	377	588	57	73
Total	14	13	2.672	3.089	233	315

Tabela 2. Total de submissões, comentários e árvores utilizadas na análise de polarização nos anos de 2018 (18) e 2022 (22).

mesmo trabalho citado, indicando como polarizadas aquelas árvores com desvio-padrão maior que 0.4375. Já a média do *Stance Score*, representa o teor geral da discussão, indicando e quantificando se as *threads* daquele debate têm aspecto mais concordante (média mais positiva) ou discordante (média mais negativa) par a par.

- **Balance Score e Balance Side.** Estas métricas avaliam o quão balanceados estão os dois lados (bipartições) da discussão. A primeira, proposta em [de Arruda et al. 2022], calcula a razão entre o mínimo e o máximo do total dos *Stance Scores* positivos e negativos, ou seja, os pesos das arestas positivas e negativas das árvores. Na segunda, proposta neste trabalho, a cardinalidade dos conjuntos gerados pela bipartição é o total de comentários em cada lado da discussão. As medidas têm intervalo de variação $[0, 1]$, sendo 0 indicativo da predominância total de apenas um lado da discussão, indicando pouca ou nenhuma polarização e 1 evidenciando uma discussão polarizada. A Figura 2, por exemplo, apresenta uma árvore de discussão em que o *Balance Score* é igual a 0.5 (6 arestas vermelhas e 3 verdes) e o *Balance Side* é igual a 0.66 (60% azul claro e 40% azul escuro).
- **Média do Controversy Level.** Esta métrica, proposta neste trabalho, calcula, primeiramente, o Desvio-Padrão do *Stance Score* das arestas de saída de cada comentário. A média é calculada considerando todos os comentários da árvore em análise. Caso o comentário tenha uma única resposta, o *Controversy Level* é igual ao módulo do *Stance Score* desta resposta. No caso de uma folha, o *Stance Score* é o módulo referente à resposta dada ao comentário de nível superior. Varia no intervalo $[0, 1]$, e valores maiores apontam discussões controversas, com a presença de polarização.

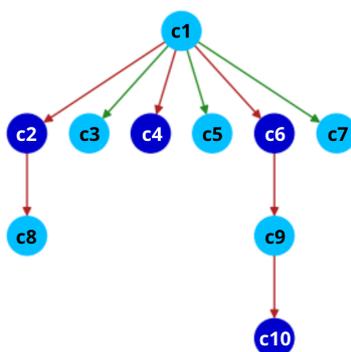


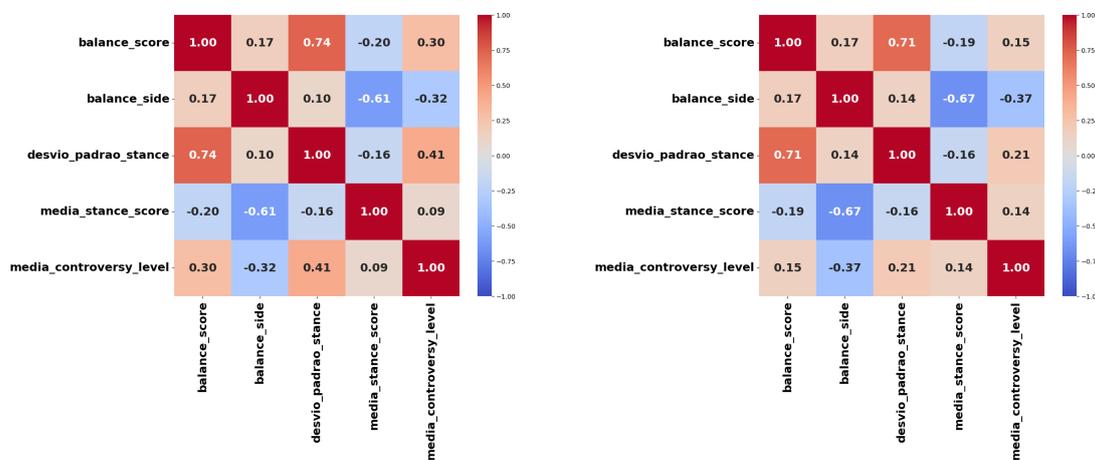
Figura 2. Exemplo de uma árvore bipartida de discussão.

4. Resultados

Nesta seção apresentamos os resultados obtidos com a metodologia proposta aplicada ao contexto das *threads* das eleições brasileiras de 2018 e 2022 no Reddit.

4.1. Correlação e Acurácia das Métricas

Para avaliar a acurácia e a coerência das métricas de quantificação da polarização nas discussões, verificamos como elas se correlacionam nos dois anos de análise, através dos mapas de calor das Figuras 3(a) e 3(b).



(a) Eleições de 2018.

(b) Eleições de 2022.

Figura 3. Matriz de Correlação entre as métricas de polarização.

Nota-se que o *Balance Score* e o desvio-padrão do *Stance Score* apresentam uma correlação positiva elevada, de 0.74 em 2018 e 0.71 em 2022, corroborando a conclusão de [Garimella et al. 2018], que afirma que tópicos controversos possuem variância mais elevada com relação a tópicos não controversos. Também destaca-se a correlação altamente negativa entre o *Balance Side* e a média do *Stance Score* e *Controversy Level*, evidenciando que, quando a discussão atinge teor de maior discordância (*Stance Score* negativo), os dois lados opostos no debate tendem a estar mais bem formados e balanceados, isto é, a discussão têm participação mais igualitária entre os indivíduos de uma primeira e segunda opiniões.

4.2. Caracterização das Árvores de Discussão em 2018 e 2022

Os gráficos das Figuras 4(a) e 4(b) sugerem que o intervalo entre 0.2 e 0.6 dos valores de *Balance Side* e *Balance Score* contém mais pontos em 2018, o que indica maior frequência de discussões com polarização moderada e caráter mais diverso. Isso é endossado na Figura 4(c), que mostra que, em média, o desvio-padrão do *Stance Score* foi maior em 2018, indicando maior variabilidade nas opiniões expressas.

Já os gráficos das Figuras 4(d) e 4(e) apontam o aumento das discussões de caráter altamente discordante (mais polarizadas), uma vez que há, no segundo ano, maior número de pontos com média de *Stance Score* muito negativo e média de *Controversy Level* muito positiva. Também, conclui-se que, pela maior ocorrência de *Stance Score* positivamente elevado em 2022, existem discussões com característica fortemente unilateral. Essas tendências também podem ser inferidas nos mapas de calor que ilustram a densidade do *Stance Score* ao longo dos anos, na Figura 5.

No mapa de calor de 2018 (Figura 5(a)), nota-se uma maior concentração horizontal de pontos no primeiro quadrante, que, segundo [de Arruda et al. 2022], mostra que os indivíduos estão conectados a outros com mesmo posicionamento médio, mas também a outros com opiniões diferentes, reforçando, também, as evidências apresentadas nas Figuras 4(a) e 4(b).

Também na Figura 5(a), nota-se que a densidade de pontos no primeiro quadrante não tem uma separação bem demarcada com a densidade mais central (em torno do zero), como ocorre em 2022. Neste caso, infere-se que as discussões no segundo ano são mais "herméticas" do que aquelas realizadas no primeiro, evidenciando uma formação mais forte de *echo chambers*, como nota-se na Figura 5(b)), em que há maior concentração diagonal de pontos nos extremos do primeiro e terceiro quadrante. A maior densidade no terceiro quadrante, por sua vez, aponta a presença de discussões com alto nível de polarização, como mostrado também em [de Arruda et al. 2022].

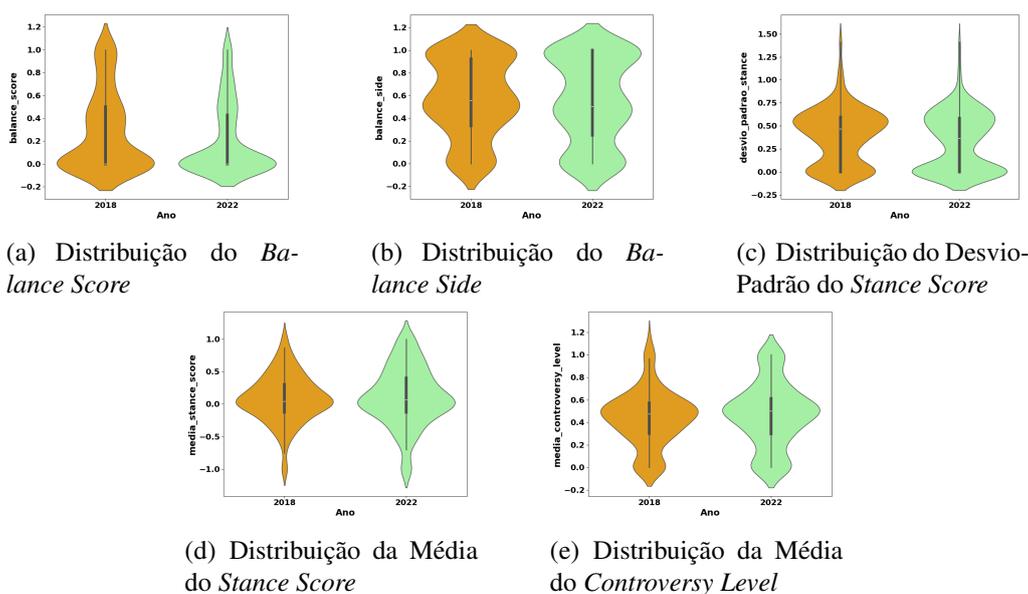


Figura 4. Caracterização das árvores de discussão.

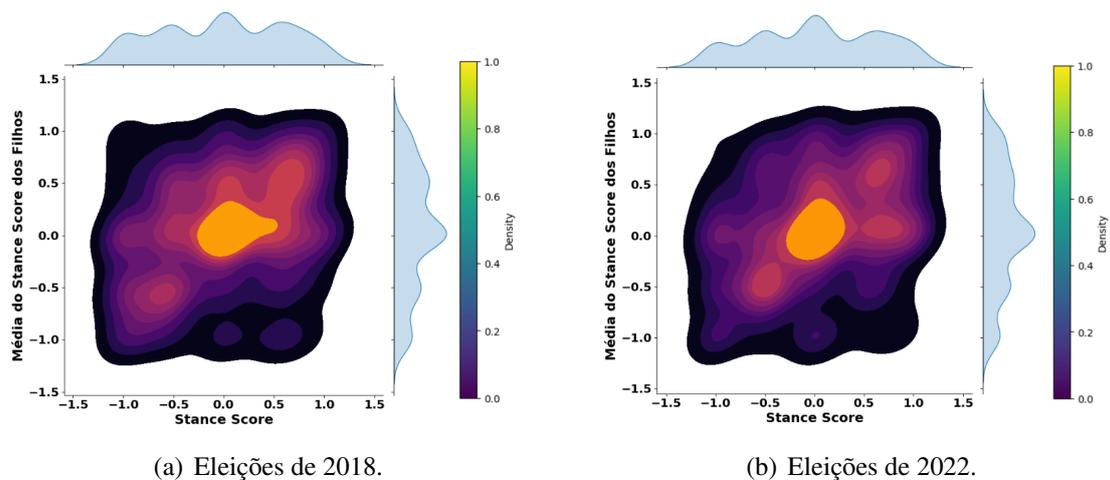


Figura 5. Mapa de Calor da Densidade de Distribuição dos Valores de *Stance Score*.

Por fim, objetivando atestar as evidências quantitativas, foram conduzidas análises qualitativas manuais em árvores de discussão de cada quadrante das Figuras 5(a) e 5(b), verificando-se o conteúdo discutido e a distribuição dos lados da discussão entre os vértices. Chamaram atenção as árvores cujos vértices estão localizados majoritariamente no primeiro quadrante que, por se tratar de *echo chambers*, são árvores com caráter quase ou totalmente unilateral, com *balance* muito baixo. Mesmo quando debatem assuntos polêmicos, os comentários dessas *threads* compartilham o mesmo ponto de vista. Também verificamos árvores com vértices no terceiro quadrante que, por outro lado, têm caráter de discussão acalorado, com *balance* elevado, se tratando normalmente de assuntos polêmicos como, por exemplo, "*Sigilo de 100 anos*", "*Carona de Lula até a COP27*", "*Isolamento Político de Bolsonaro às vésperas das Eleições 2022*". Além disso, nesse grupo, as opiniões são divergentes e as árvores possuem vários ramos do tipo resposta-réplica.

4.3. Limitações

Representatividade. Conforme mencionado, devido ao custo de requisição à API da OpenAI, fizemos uma amostragem das árvores de discussão para proceder à aplicação da metodologia. Com isso, não conseguimos verificar o quão representativo é o espaço amostral com base no universo, podendo ter resultados subespecificados. Porém, na amostragem, mantivemos a proporção de postagens por cada *subreddit* e os resultados das métricas mostraram-se coerentes entre si e também com os trabalhos prévios da literatura citados.

Restrição de Domínio. Uma vez que nos aprofundamos nas discussões da plataforma Reddit, o escopo fica restrito, dada a infinidade de redes sociais existentes nos dias atuais, cada qual com diferentes aspectos de modelagem e, possivelmente, diferentes teores de discussão.

Confiança na Detecção de Posição. Como a modelagem do grafo depende do processo de *stance detection* para bipartição, a qualidade do método utilizado é fundamental e torna-se um gargalo para a execução do *framework* proposto. Para tratar disto, em evoluções deste trabalho, pretende-se fazer uso de APIs abertas para *stance detection*,

possibilitando maior volume de aplicação e, assim, uma verificação mais contundente dos resultados.

Bipartição. É sabido que discussões de assuntos controversos não necessariamente têm apenas dois posicionamentos contrapostos. No contexto deste estudo, ao fazer uma bipartição das árvores de discussão, à luz da contraposição política da esquerda e direita no Brasil, consideramos, assim, apenas dois lados.

5. Conclusão e Trabalhos Futuros

Neste trabalho propõe-se uma metodologia para caracterização de discussões polarizadas em mídias sociais, baseada na modelagem matemática de árvores de discussão, na definição de posicionamento através do uso do modelo GPT-4 e no cálculo de métricas de verificação de existência e medição da intensidade de polarização.

Considerando como caso de estudo a comparação entre as discussões realizadas no Reddit, referentes às eleições de 2018 e 2022, os resultados mostraram uma tendência de aumento da polarização e diminuição do caráter de diversidade dos debates realizados por estes usuários entre os períodos analisados, corroborando discussões abordadas em trabalhos de diferentes campos de pesquisa. Assim, os resultados iniciais apresentados neste estudo se mostraram promissores para a aplicação da metodologia de quantificação de polarização em um maior conjunto de dados como também em diferentes contextos e mídias sociais.

Em possíveis extensões deste trabalho, objetivando contornar as atuais limitações, pretende-se fazer o uso de APIs abertas que provenham um bom desempenho para *stance detection* possibilitando, assim, a utilização de um maior conjunto de dados para aplicação da metodologia. Adicionalmente, com o objetivo de viabilizar uma análise mais profunda e qualitativa dos dados, é possível a realização de técnicas de Processamento de Linguagem Natural (PLN) de modo a obter os principais tópicos debatidos e expor de maneira mais clara o conteúdo-alvo de estudo.

Referências

- Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.
- Agrawal, R., Rajagopalan, S., Srikant, R., and Xu, Y. (2003). Mining newsgroups using networks arising from social behavior. In *Proceedings of the 12th international conference on World Wide Web*, pages 529–535.
- ALDayel, A. and Magdy, W. (2021). Stance detection on social media: State of the art and trends. *Information Processing Management*, 58(4):102597.
- Almeida, R. J. A. (2018). Leia - léxico para inferência adaptada. <https://github.com/rafjaa/LeIA>.
- Alsinet, T., Argelich, J., Béjar, R., and Martínez, S. (2021a). Measuring polarization in online debates. *Applied Sciences*, 11(24):11879.
- Alsinet, T., Argelich, J., Béjar, R., and Martínez, S. (2021b). Measuring polarization in online debates. *Applied Sciences*, 11(24):11879.
- Britt, R. K., Franco, C. L., and Jones, N. (2023). Trends and challenges within reddit and health communication research: A systematic review. *Communication and the Public*, 8(4):402–417.
- Brum, P., Cândido Teixeira, M., Vimieiro, R., Araújo, E., Meira Jr, W., and Lobo Pappa, G. (2022). Political polarization on twitter during the covid-19 pandemic: a case study in brazil. *Social Network Analysis and Mining*, 12(1):140.
- Caetano, J., Guimarães, S., Araújo, M. M., Silva, M., Reis, J. C., Silva, A. P., Benevenuto, F., and Almeida, J. M. (2022). Characterizing early electoral advertisements on twitter: A brazilian case study. In *Proc. of the SocInfo*.
- Corrêa, E. V. B. (2023). Redes sociais, ódio e polarização política: A psicodinâmica da guerra civil digital brasileira. *POLÍTICA EM FOCO: O melhor embate é o debate–Vol. 3*, page 17.
- Dahlberg, L. (2007). Rethinking the fragmentation of the cyberpublic: from consensus to contestation. *New media & society*, 9(5):827–847.
- de Arruda, H. F., Cardoso, F. M., de Arruda, G. F., Hernández, A. R., da Fontoura Costa, L., and Moreno, Y. (2022). Modelling how social network algorithms can influence opinion polarization. *Information Sciences*, 588:265–278.
- Franco, A. B. and Pound, N. (2022). The foundations of bolsonaro’s support: Exploring the psychological underpinnings of political polarization in brazil. *Journal of Community & Applied Social Psychology*, 32(5):846–859.
- Garimella, K., Morales, G. D. F., Gionis, A., and Mathioudakis, M. (2018). Quantifying controversy on social media. *ACM Transactions on Social Computing*, 1(1):1–27.
- Guimaraes, S., Silva, M., Caetano, J., Araújo, M., Santos, J., Reis, J. C., Silva, A. P., Benevenuto, F., and Almeida, J. M. (2022). Análise de propagandas eleitorais antecipadas no twitter. In *Anais do BrasNAM*.

- Heltzel, G. and Laurin, K. (2020). Polarization in america: Two possible futures. *Current opinion in behavioral sciences*, 34:179–184.
- Hua, Y., Naaman, M., and Ristenpart, T. (2020). Characterizing twitter users who engage in adversarial interactions against political candidates. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–13.
- Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225.
- Karypis, G. (1997). Metis: Unstructured graph partitioning and sparse matrix ordering system. *Technical report*.
- Kim, Y., Hsu, S.-H., and de Zúñiga, H. G. (2013). Influence of social media use on discussion network heterogeneity and civic engagement: The moderating role of personality traits. *Journal of communication*, 63(3):498–516.
- Kobbe, J., Hulpuş, I., and Stuckenschmidt, H. (2020). Unsupervised stance detection for arguments from consequences. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pages 50–60.
- Küçük, D. and Can, F. (2020). Stance detection: A survey. *ACM Computing Surveys (CSUR)*, 53(1):1–37.
- Kumar, D., Hancock, J., Thomas, K., and Durumeric, Z. (2022). Understanding longitudinal behaviors of toxic accounts on reddit. *arXiv preprint arXiv:2209.02533*.
- Levin, S. A., Milner, H. V., and Perrings, C. (2021). The dynamics of political polarization.
- Lima, H. F. d. (2021). Polarização política afetiva: São os valores humanos e os traços de personalidade uma explicação?
- Mutz, D. C. (2002). The consequences of cross-cutting networks for political participation. *American journal of political science*, pages 838–855.
- Ortellado, P., Ribeiro, M. M., and Zeine, L. (2022). Existe polarização política no brasil? análise das evidências em duas séries de pesquisas de opinião. *Opinião Pública*, 28:62–91.
- Recuero, R., Soares, F. B., and Gruzd, A. (2020). Hyperpartisanship, disinformation and political conversations on twitter: The brazilian presidential election of 2018. In *Proceedings of the international AAAI conference on Web and social media*, volume 14, pages 569–578.
- Rossini, P. (2020). Beyond toxicity in the online public sphere: understanding incivility in online political talk. *A research agenda for digital politics*, pages 160–170.
- Ruediger, M. A. et al. (2014). Redes sociais retratam eleição mais polarizada da história recente do brasil.
- Saveski, M., Roy, B., and Roy, D. (2021). The structure of toxic conversations on twitter. In *Proceedings of the Web Conference 2021*, pages 1086–1097.
- Seno, E., Silva, L., Anno, F., Rocha, F., and Caseli, H. (2024). Aspect-based sentiment analysis in comments on political debates in portuguese: evaluating the potential of

- chatgpt. In *Proceedings of the 16th International Conference on Computational Processing of Portuguese*, pages 312–320.
- Silva, A. G. (2023). O insaciável espírito da época: ensaios de psicologia analítica e política, de: Humbertho oliveira, roque tadeu gui e rubens bragarnich. editora vozes, 2021. *Self-Revista do Instituto Junguiano de São Paulo*, 8:e007–e007.
- Sunstein, C. R. (2018). *Republic: Divided Democracy in the Age of Social Media*. Princeton University Press, Princeton.
- Vicario, M. D., Quattrocioni, W., Scala, A., and Zollo, F. (2019). Polarization and fake news: Early warning of potential misinformation targets. *ACM Transactions on the Web (TWEB)*, 13(2):1–22.
- Xia, Y., Zhu, H., Lu, T., Zhang, P., and Gu, N. (2020). Exploring antecedents and consequences of toxicity in online discussions: A case study on reddit. *Proceedings of the ACM on Human-computer Interaction*, 4(CSCW2):1–23.
- Zhang, B., Ding, D., and Jing, L. (2022). How would stance detection techniques evolve after the launch of chatgpt? *arXiv preprint arXiv:2212.14548*.
- Zhang, B., Fu, X., Ding, D., Huang, H., Li, Y., and Jing, L. (2023). Investigating chain-of-thought with chatgpt for stance detection on social media. *arXiv preprint arXiv:2304.03087*.