

Identificação e Caracterização de Campanhas de Propagandas Eleitorais Antecipadas Brasileiras no Twitter

Marcelo M. R. Araujo¹, Carlos H. G. Ferreira^{1,2},
Julio C. S. Reis³, Ana P. C. Silva¹, Jussara M. Almeida¹

¹Universidade Federal de Minas Gerais (UFMG) – Brasil

²Universidade Federal de Ouro Preto (UFOP) – Brasil

³Universidade Federal de Viçosa (UFV) – Brasil

mmra@ufmg.br, chgferreira@ufop.edu.br,

jreis@ufv.br, {ana.coutosilva, jussara}@dcc.ufmg.br

Abstract. *In this work, we investigate the coordinated promotion of early political campaigns by users on Twitter, focusing on the Brazilian 2022 pre-election period. The explored methodology involves modeling a network based on co-retweets, extracting a backbone of the network, and finally identifying and analyzing communities focused on user characteristics and shared content. The results show a significant number of communities promoting content related to different pre-election candidates from different political spectrums, including right-wing and left-wing. We also found that right-wing communities are much larger than left-wing communities, both in terms of the number of users and the amount of information shared. We believe that our results can provide interesting inputs for understanding the phenomenon in the Brazilian context and, in the future, assist in the formulation of mechanisms to efficiently detect coordinated actions (i.e., pre-election campaigns) on social media platforms.*

Resumo. *Neste trabalho investigamos a promoção coordenada de campanhas de propaganda política antecipadas realizadas por usuários do Twitter, com foco no período pré-eleitoral brasileiro de 2022. A metodologia explorada envolve a modelagem de uma rede baseada em co-retweets, a extração de um backbone da rede, e por fim, a identificação e análise de comunidades, com foco em características dos usuários e do conteúdo compartilhado. Os resultados revelam um número significativo de comunidades que promovem conteúdo relacionado a diversos pré-candidatos de diferentes espectros políticos, dentre eles, direita e esquerda. Ademais, constatamos que as comunidades de direita são muito maiores em comparação com as de esquerda, tanto em relação ao número de usuários quanto no volume de informações compartilhadas. Acreditamos que os resultados identificados fornecem insumos interessantes para o entendimento do fenômeno no contexto brasileiro, bem como, futuramente, para auxiliar a formulação de mecanismos que sejam eficientes na detecção de ações coordenadas (i.e., campanhas eleitorais antecipadas) em plataformas sociais.*

1. Introdução

As plataformas de mídia social mudaram a forma como as pessoas se conectam, se comunicam e consomem informações, especialmente no contexto político

[da Rosa Jr et al. 2022]. Como resultado, novos desafios surgiram, pois elas se tornaram poderosas ferramentas políticas. Por um lado, as pessoas frequentemente usam as mídias sociais para compartilhar ideias, buscar informações e/ou conhecer seus candidatos. Por outro lado, os candidatos começaram a usá-las como plataformas digitais para espalhar suas ideias, atacar seus oponentes e/ou solicitar votos, muitas vezes violando várias leis eleitorais [Guimaraes et al. 2022, Caetano et al. 2022].

Entre os vários problemas que persistem nesse ecossistema está a proliferação de campanhas, particularmente durante o período pré-eleitoral, para promover ilegalmente um conteúdo de cunho político. As campanhas coordenadas no mundo digital tornaram-se uma área cada vez mais importante de estudo devido ao seu potencial para causar polarização política e ameaças à segurança por meio de protestos e tumultos no mundo real [Varol et al. 2017, Tardelli et al. 2022]. No Brasil, a prática de campanha fora de época caracteriza crime eleitoral¹. Há fortes evidências na literatura de que muitas eleições no mundo foram influenciadas por essa prática em várias plataformas de mídia social [Varol et al. 2017, Badawy et al. 2019, da Rosa Jr et al. 2022, Saldanha et al. 2022]. Inspirados por esses eventos, vários estudos examinaram como eleições foram influenciadas por ações coordenadas por usuários no Twitter e, em particular, baseando-se em modelos de rede [Ferrara et al. 2020, Grimminger and Klinger 2021, Linhares et al. 2022]. No entanto, a literatura ainda carece de mecanismos voltados para a compreensão de como as campanhas eleitorais são coordenadas e construídas no contexto brasileiro.

Paralelo a esses esforços, nosso objetivo aqui é investigar a existência de grupos de usuários no Twitter que impulsionaram e disseminaram campanhas políticas antecipadas de maneira coordenada durante o período pré-eleitoral de 2022, especificamente entre 1º de Janeiro e 26 de Agosto. Neste contexto, definimos ação coordenada como *a presença de grupos de usuários que agem para disseminar um conjunto de conteúdos explicitamente relacionados a campanhas durante o período pré-eleitoral de forma persistente e conjunta*. Para isso, coletamos dados que cobrem o período de oito meses antes do período pré-eleitoral (i.e., janeiro a agosto/2022). Em seguida, modelamos uma rede de co-retweets e empregamos a estratégia proposta por [Saldanha et al. 2022], que remove a presença potencialmente grande de arestas ruidosas na rede, ou seja, arestas periféricas que não caracterizam padrões de coordenação e, portanto, não contribuem para o estudo desse fenômeno. Finalmente, identificamos grupos de usuários representativos de comunidades no *backbone* e os caracterizamos em termos de duas dimensões principais, sendo elas: (i) usuários e (ii) análise textual (conteúdo).

Em resumo, nossos resultados revelam várias descobertas interessantes relacionadas ao fenômeno no contexto brasileiro. Primeiramente, apresentamos fortes evidências de coordenação promovendo conteúdo relacionados a diversos pré-candidatos de diferentes espectros políticos, dentre eles, direita e esquerda. Além disso, encontramos um número maior de comunidades pertencentes a direita, e essas comunidades são, de maneira geral, substancialmente maiores do que as comunidades de esquerda, tanto em número de usuários, quanto no volume de informação compartilhada. Esperamos que os resultados identificados sejam úteis para fomentar a proposição de estratégias focadas na detecção de campanhas, especialmente no contexto político brasileiro.

¹http://www.planalto.gov.br/ccivil_03/leis/1950-1969/11164.htm

2. Trabalhos Relacionados

Existem vários estudos na literatura focados na análise de fenômenos no contexto político, a partir de dados oriundos de plataformas de mídia social, como Facebook [Giglietto et al. 2020], WhatsApp [Nobre et al. 2020], Instagram [Ferreira et al. 2021] e Twitter [da Rosa Jr et al. 2022]. Uma parte significativa desses esforços se concentra em analisar a disseminação de conteúdo político no Twitter [Pacheco et al. 2020, Keller et al. 2020, Grimminger and Klinger 2021], identificando e caracterizando ações coordenadas para promover conteúdo político por meio de modelos de rede, conforme discutido neste trabalho. Por exemplo, em [Pacheco et al. 2020] aos autores investigaram campanhas de desinformação contra a Defesa Civil da Síria no Twitter. Ao examinar um modelo de rede que conecta usuários que postaram tweets semelhantes, eles descobriram grupos de usuários que compartilhavam repetidamente o mesmo conteúdo, uma forte indicação de comportamento coordenado. Em [Keller et al. 2020], os autores examinaram a rede de retweets para estudar a coordenação de anúncios durante a eleição presidencial na Coreia do Sul. Já Danaditya et al. (2022) caracterizaram ações coordenadas em discursos do Twitter das eleições indonésias, modelando uma rede de interação no plataforma na qual uma conexão (i.e., link) entre dois usuários significa que houve algum tipo de comunicação entre eles, como por exemplo, um retweet, uma menção, dentre outros [Danaditya et al. 2022].

Por outro lado, há outros esforços se concentraram especificamente em identificar comportamentos coordenados por meio da extração do *backbone* da rede. Por exemplo, Vargas et al. (2020) examinaram a presença de padrões de coordenação em vários modelos de rede baseados no Twitter em questões políticas em muitos países em redes de retweets, co-tweets, co-retweets, co-hashtags, co-comentários e co-URLs. Para remover as arestas consideradas ruidosas, os autores aplicaram limiares baseados em tempo em que um determinado padrão de atividade [Vargas et al. 2020]. Por fim, em uma abordagem similar, uma metodologia para detecção e análise de ações sincronizadas realizadas por grupos coordenados é apresentada em [Ng and Carley 2022].

Ortogonal a esses estudos que empregam modelos de rede e estratégias de extração de *backbone* para análise de disseminação de informação em plataformas de mídia social, nosso estudo foca na disseminação de campanhas antecipadas no Brasil nas eleições de 2022 baseando-se na estratégia proposta em [Saldanha et al. 2022], que conforme explicado nas próximas seções, apresenta uma série de vantagens em relação aquelas aplicados nos estudos supramencionados. Além disso, é importante mencionar que cada evento, como as eleições no Brasil, pode apresentar particularidades que o diferem de todos os outros, o que motiva a realização de estudos focados em contextos específicos.

3. Metodologia

Nesta seção, descrevemos a metodologia proposta para detectar comunidades representativas de grupos de usuários que atuam de forma coordenada com o objetivo de promover propaganda política durante o período pré-eleitoral brasileiro, a saber: 1º de Janeiro e 26 de Agosto de 2022.

3.1. Base de Dados

Com o objetivo de identificar campanhas antecipadas realizadas no período pré-eleitoral referentes às eleições brasileiras de 2022, exploramos um conjunto

pré-definido de palavras-chave definido em trabalhos anteriores [Silva et al. 2021, Guimaraes et al. 2022], que foi elaborado em conjunto com usuários especialistas de instituições fiscalizadoras do processo eleitoral no Brasil, e que foi posteriormente expandido a partir de hashtags que co-ocorrem com as palavras-chave elencadas pelos especialistas. Para fins exemplificação, este conjunto é composto por termos como: ‘*vote em mim*’, ‘*se eu for eleito*’ e ‘*conto com o seu voto*’, que de forma implícita ou explícita apresentam pedidos de voto ao eleitor. Assim, foram coletados tweets em português utilizando a Twitter API Search², restringindo a coleta somente a tweets que mencionam pelo menos uma das palavras-chave presentes nesta lista. No total, coletamos aproximadamente 12 milhões de tweets no período entre 1º de janeiro de 2022 e 26 de agosto de 2022. Os dados do Twitter e detalhes relativos à coleta estão disponíveis em zenodo.org/record/7976332 de acordo com a política de divulgação da plataforma.

3.2. Modelo da Rede

Representamos a rede de compartilhamento de conteúdo entre usuários através de um grafo ponderado não direcionado $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, sendo \mathcal{V} o conjunto de nós que representam os usuários que, em algum instante de tempo do período de coleta, retweetaram pelo menos uma vez um tweet em comum e \mathcal{E} o conjunto de arestas entre pares de nós v_i e v_j que retweetaram um mesmo tweet. Cada aresta é ponderada pelo total de tweets idênticos retweetados por cada par de nós.

3.3. Extração do *Backbone* da Rede de Retweets

Um dos maiores desafios da detecção de ações coordenadas em plataformas de mídias sociais é como, a partir do grafo \mathcal{G} , extrair os relacionamentos que de fato fazem parte da coordenação sendo analisada [Ferreira et al. 2022]. Algumas arestas podem ser resultado, por exemplo, da popularidade de um tweet em particular ou da grande atividade de um determinado usuário. Ou seja, um tweet que viraliza pode ser retweetado por vários usuários pelo simples fato de ser uma informação popular e não em decorrência de uma atividade coordenada. Da mesma forma, um usuário com grande atividade na rede pode constantemente retweetar, meramente por acaso, diversos tweets que sejam difundidos por outros usuários agindo ou não de forma coordenada.

Considerando os efeitos paralelos supramencionados, é esperado que a rede de retweets tenha um grande número de arestas espúrias e aleatórias, isto é, caracterizadas como ruídos que dificultam a identificação das arestas que realmente estão associadas a esforços de coordenação de difusão de conteúdo. Portanto, a partir grafo \mathcal{G} devemos selecionar as arestas que capturem os relacionamentos não aleatórios e que revelem o padrão de coordenação entre usuários. Especificamente, devemos extrair de \mathcal{G} um subconjunto de nós pertencentes a um subgrafo $B_{\mathcal{G}}$ conhecido como *backbone*, formado por arestas salientes que revelam evidências fortes de coordenação para a promoção massiva de conteúdo. Para alcançar tal objetivo, aplicamos a abordagem proposta por [Saldanha et al. 2022], que permite extrair *backbone* representativos de padrões de coordenação como descrito a seguir.

Primeiramente, a abordagem adotada aplica o algoritmo *Disparity Filter* [Ángeles Serrano et al. 2009], que funciona da seguinte maneira. Dada a força de um

²<https://developer.twitter.com/en/docs/twitter-api/tweets/search/quick-start/recent-search>

nó v_i , definida como a soma de todos os pesos das arestas ligados a v_i , o algoritmo considera uma aresta anexada a v_i como saliente para o *backbone* se ela representa uma “grande fração” da força de v_i . Especificamente, cada aresta ligada a v_i é testada contra a hipótese nula de que os pesos de todas as arestas de v_i são *uniformemente* distribuídos. Arestas salientes são aquelas cujos pesos se desviam significativamente dessa hipótese. Observe que uma aresta é testada duas vezes, uma vez para cada nó ao qual incide, e é considerada saliente se for estatisticamente significativa para ambos os vértices quando comparada a um p-valor. Dessa forma, o *backbone* extraído por essa primeira abordagem, consiste de uma rede que revela as conexões mais importantes para cada nós em termos de co-atividade, isto é, co-retweets.

Em seguida, o método *Neighborhood Overlap* [David and Jon 2010] é aplicado no *backbone* resultante da etapa anterior para remover arestas periféricas. A ideia é que com este método, dois nós não somente tenham uma aresta entre si com um padrão de atividade que estatisticamente desvia do esperado, mas também compartilhem um conjunto mínimo de vizinhos que também apresentem o mesmo padrão. Essa etapa é fundamental para mitigar o efeitos negativos da amostragem de dados do Twitter na extração da comunidade coordenadas, considerando que a coleta pode fornecer uma visão parcial do comportamento dos usuários na rede. Dessa forma, o foco fica restrito para as partes mais fortemente conectadas do *backbone*.

3.4. Detecção de Comunidades em B_G

A próxima etapa após a extração do *backbone* B_G é a identificação de comunidades em B_G que promovem conteúdo de campanha pré-eleitoral. Para isso, aplicamos o algoritmo de *Louvain* [Blondel et al. 2008]. O principal objetivo deste algoritmo é maximizar a *modularidade*, uma métrica que captura o quão densamente conectados os nós dentro de uma comunidade estão, em comparação com o quão conectados eles estariam em uma rede aleatória com a mesma sequência de graus. Essa métrica possui valores entre $-1/2$ e $+1$ e valores acima de 0.4 são considerados evidência forte da presença de comunidades bem definidas [Newman and Girvan 2004].

3.5. Caracterização das Comunidades em B_G

As comunidades extraídas do *backbone* B_G foram caracterizadas em termos do (i) perfil dos usuários e; (ii) análise textual do conteúdo compartilhado.

Perfil dos Usuários. Adicionalmente ao total de usuários presentes em cada comunidade, analisamos a distribuição do tempo de vida das suas contas e a distribuição do número de seus seguidores, bem como a probabilidade de um usuário ser uma conta automatizada (ou *bot*)³. Enquanto a presença de um número substancial de contas com alta probabilidade de serem *bot* possa ser, por si só, um indicativo do impulsionamento rápido e automático de conteúdo em uma comunidade, outros fatores, como a presença significativa de contas novas que não possuem muitos seguidores, podem sugerir a utilização de táticas de criação de contas falsas para inflar artificialmente o tamanho da comunidade e amplificar o alcance do conteúdo promovido. Portanto, a análise conjunta dessas métricas podem revelar, por exemplo, se uma campanha antecipada é impulsionada de forma orgânica ou não, e auxiliar na definição do perfil dos usuários que estão difundindo este tipo de conteúdo.

³Para esta análise foi utilizada a ferramenta Botometer (<https://botometer.osome.iu.edu>).

Tabela 1. Propriedades topológicas da rede \mathcal{G} e do *backbone* $B_{\mathcal{G}}$.

Rede	# Vértices	# Arestas	Grau Médio	Densidade	Coef. Clust. Médio	# Componentes Conectados	# Comunidades (>100 users)	Modularidade
\mathcal{G}	155.008	302.557.077	3.903,7	0,025	0,87	702	767 (16)	0,67
$B_{\mathcal{G}}$	12.865	582.302	90,52	0,007	0,65	872	908 (9)	0,59

Análise Textual. Neste segundo conjunto de análises, apresentamos as hashtags, menções e retweets populares das maiores comunidades de quatro diferentes pré-candidatos, com o objetivo de não somente compreender o conteúdo impulsionado, mas também para demonstrar formas sobre como os pré-candidatos mais frequentemente mencionados em cada uma das comunidades analisadas são promovidos. Complementando esta caracterização inicial, aprofundamos a análise textual considerando as seguintes análises complementares: (i) análise de sentimentos, para capturar a percepção dos usuários em relação à um determinado candidato e (ii) análise de toxicidade, para auxiliar na identificação de linguagem ofensiva ou abusiva no conteúdo compartilhado. O *SentiStrength*⁴, método utilizado para análise de sentimentos, fornece uma pontuação inteira variando de -4 (fortemente negativo) a $+4$ (fortemente positivo). A pontuação 0 indica um sentimento neutro. Consideramos aqui como *negativo*, *neutro* e *positivo* tweets e retweets com pontuações menores que 0 , iguais a 0 e maiores que 0 , respectivamente. Considerando a análise de toxicidade, utilizamos a Perspective API⁵, que é uma ferramenta desenvolvida para analisar o texto e avaliar, dentre outros, a probabilidade de o conteúdo ser considerado tóxico ou abusivo. A API analisa o texto fornecido e atribui uma pontuação de 0 a 1 , que indica a probabilidade dos usuários perceberem que o texto é considerado tóxico.

4. Análises e Resultados

Nesta seção, apresentamos e discutimos o conjunto de análises realizadas que visam entender a ocorrência de campanhas coordenadas no período pré-eleitoral no Twitter.

4.1. Propriedades Topológicas de \mathcal{G} e de $B_{\mathcal{G}}$

A Tabela 1 apresenta as principais características topológicas de ambas as redes, a original e do *backbone* extraído. Como parametrização, adotamos um *p-valor* de $0,10$ para o *Disparity Filter* e o 90 percentil do *Neighborhood Overlap*. Observa-se que a extração do *backbone* elimina $91,71\%$ dos nós, $99,8\%$ das arestas e indica, portanto, que a vasta maioria dos nós na rede original não apresentam padrões de coordenação. Devido à eliminação de arestas menos significativas, o grau médio, a densidade da rede e o coeficiente de agrupamento diminuem no *backbone* em comparação com a rede completa. Nota-se também que com a extração do *backbone*, a quantidade total de comunidades aumentou, embora o número de comunidades com uma quantidade mais relevante de usuários (>100 usuários) tenha diminuído. Por fim, observa-se que a *modularidade* em ambos os casos é alta, indicando que o algoritmo de Louvain revela comunidades fortemente estruturadas. No *backbone*, em particular, essas são comunidades potencialmente coordenadas para disseminação de propaganda.

4.2. Perfil das Comunidades

A seguir, apresentamos as principais características das comunidades definidas no *backbone* de interações significativas para o estudo da ação coordenada de promoção de cam-

⁴<http://sentistrength.wlv.ac.uk/>

⁵<https://perspectiveapi.com/>

Tabela 2. Características gerais das comunidades extraídas de B_G .

Ideologia	id	# Usuários	# Retweets	# Tweets Distintos	Média Retweet/Usuário	D.P. Retweet/Usuário	# Menções	# Hashtags	Média de Vida da Conta de Usuário (Dias)	Principal Pré-Candidato
Esquerda	4	886	27.171	1.495	30,67	33,93	27.171	19.502	1.900	C. Gomes
Esquerda	6	858	15.157	2.023	17,67	15,24	15.157	13.681	2.484	Lula
Esquerda	8	155	623	89	4,02	1,39	623	601	2.380	G. Boulos
Esquerda	9	142	2.719	446	19,15	21,98	2.719	2.314	1.911	S. Manzano
Direita	1	2.830	49.898	2.564	17,63	8,84	49.898	45.831	1.855	J. Bolsonaro
Direita	2	2.223	23.766	1.644	10,69	7,85	23.766	22.470	1.894	J. Bolsonaro
Direita	3	1.505	10.157	1.098	6,75	3,2	10.157	9.485	1.793	J. Bolsonaro
Direita	5	869	42.210	2.429	48,57	21,31	42.210	37.934	1.830	J. Bolsonaro
Direita	7	622	14.714	1.485	23,66	23,78	14.714	10.076	1.700	S. Moro

panhas antecipadas. Nossas análises focam em todas as 9 comunidades que possuem um total de usuários maior que 100, como indica a Tabela 1, para entender como a orquestração em torno de campanhas políticas ocorre em maior escala. Para enriquecer as discussões apresentadas, primeiramente realizamos uma classificação manual da ideologia política da comunidade (esquerda ou direita), com base na análise dos retweets mais populares em cada comunidade, assim como as principais hashtags e menções de cada comunidade. Para tal foram recrutados três voluntários independentes, que forneceram uma classificação idêntica para todas as comunidades, resultando em concordância máxima.

Características Gerais. A Tabela 2 mostra as principais estatísticas das 9 comunidades analisadas, apresentando-as de acordo com a ideologia definida pelos voluntários. Para cada uma delas, foi atribuído um identificador único (id) em ordem decrescente de total de usuários, bem como a orientação ideológica de cada uma das comunidades.

Nota-se que, embora o número de comunidades pertencentes à ideologia de direita e à ideologia de esquerda seja equilibrado, as três maiores comunidades em número de usuários (comunidades 1, 2 e 3) são de direita. Além disso, as comunidades de esquerda contém, em média, apenas 30,91% do número de usuários das comunidades de direita. Além do total de usuários, destacamos a diferença expressiva do total de retweets compartilhados em comunidades de diferentes ideologias, onde as comunidades de direita retweetaram 208% a mais que as comunidades de esquerda.

Importante ressaltar que, ao analisarmos o conteúdo da comunidade⁶ com exceção da comunidade 7, de suporte ao pré-candidato Sérgio Moro, as demais comunidades da direita focam na propagação de conteúdo envolvendo o então pré-candidato Jair Bolsonaro. As comunidades de esquerda, por outro lado, são mais heterogêneas, impulsionando conteúdo de quatro pré-candidatos distintos: Lula, Ciro Gomes, Guilherme Boulos e Sofia Manzano (comunidades 6, 4, 8 e 9, respectivamente). Apesar de Guilherme Boulos e Sérgio Moro não terem concorrido à Presidência da República, ambos eram personalidades políticas cotadas para o cargo, e foram lançados como pré-candidatos para outros cargos legislativos por seus partidos, como Deputado Federal e Senador, respectivamente.

Para entendermos melhor como estas comunidades se organizam estruturalmente, a Figura 1 apresenta a topologia do *backbone* com os nós das mesmas comunidades representados pela mesma cor. As comunidades da direita que impulsionam o pré-candidato Jair Bolsonaro (comunidades 1, 2, 3 e 5) se encontram bem próximas, indicando que existem usuários que desempenham o papel de pontes de propagação de conteúdo entre diferentes comunidades. Este fenômeno ocorre em intensidade consideravelmente menor entre as comunidades de esquerda, considerando que os limites destas comunidades são

⁶Foram analisados os principais retweets, hashtags e menções, bem como outras visualizações de conteúdo, dos retweets de cada comunidade.

Figura 1. Representação gráfica das comunidades presentes em B_G .

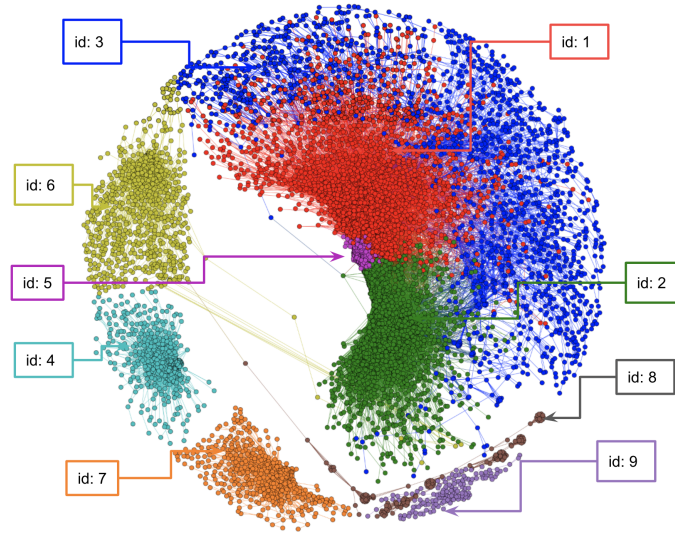
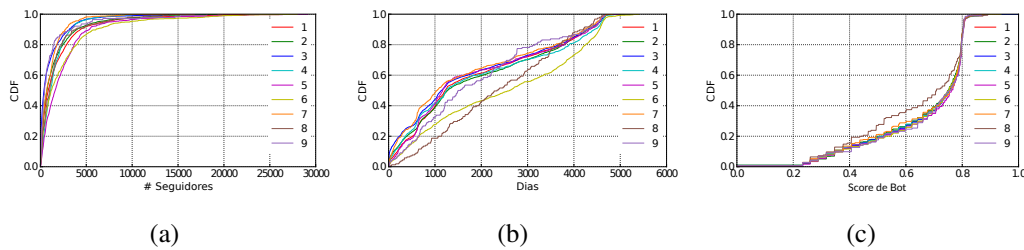


Figura 2. (a) Distribuição do número de seguidores, (b) Distribuição do tempo de vida das contas e (c) Probabilidade do usuário ser um *bot*.



mais bem definidos. Tendo em vista, também, que a coleta fornece uma visão parcial dos dados, ou seja, podem existir interações entre usuários que não foram representadas, isso pode implicar na fragmentação das comunidades de um mesmo pré-candidato. Ou seja, caso tivéssemos a visão completa dos dados, seria possível que a metodologia proposta encontrasse menos comunidades relacionadas à Jair Bolsonaro, mas que seriam maiores.

Perfil dos Usuários. A Figura 2 apresenta a distribuição de probabilidade acumulada (CDF) do número de seguidores (a) e o tempo de vida da conta dos usuários (b), bem como a probabilidade de um usuário ser um *bot*, ou seja, uma conta automatizada (c). Considerando a Figura 2(a), que representa a distribuição do total de seguidores dos usuários em cada comunidade, podemos inferir que, em média, 60% das contas de usuários nas comunidades possuem mais de 100 seguidores. Para a distribuição do tempo de vida das contas Figura 2(b), podemos verificar que algumas das comunidades de esquerda, como a comunidade 6 e 8, tendem a ter contas mais antigas do que as demais comunidades. As distribuições de probabilidade da métrica nas demais comunidades são similares, com pequenos desvios na distribuição entre elas. Por fim, a probabilidade de um usuário ser um *bot* Figura 2(c) mostra que cerca de 20% das contas de cada comunidade possuem um score $\geq 80\%$, o que sugere a presença de alguma atividade automatizada ou com comportamento suspeito nas comunidades.

No entanto, apesar da possibilidade de uma presença não desprezível de engajamento artificial, o fato da maioria das contas terem um tempo de vida longo e um total de seguidores acima de 100 sugere uma organicidade na propagação do conteúdo, indicando

Tabela 3. Top-5 hashtags e menções mais frequentes compartilhadas nas 4 principais comunidades relacionadas à pré-candidatos das eleições de 2022.

Id	Candidato	Top-5 Hashtags	Top-5 Menções
1	J. Bolsonaro	#ColunaCH, #DiariodoPoder, #Bolsonaro2022, #BolsonaroReeleito2022, #DemitaEm2022	@jairbolsonaro, @minfraestrutura, @SF_Moro, @TSEjusbr, @DamaresAlves
4	C. Gomes	#CNNnasEleições, #06deAgostoTemDebate, #CNNBrasil360, #ConvençãoPDT, #TáNaHoraDeVocêOlharProCiro	@cirogomes, @elviscezar_ofc, @PDT_Nacional, @CarlosLupiPDT, @antonionetopdt
6	Lula	#VamosJuntosPeloBrasil, #g1, #BolsonaroVagabundo, #Lula13, #LulaNoPrimeiroTurno	@LulaOficial, @lulaoficial, @jairbolsonaro, @igorgadelham, @ptbrasil
7	S. Moro	#MoroPresidente2022, #MoroOuNulo, #CanalLivre, #Moro2022, #MoroPorUmBrasilMelhor	@SF_Moro, @uniaobrasil44, @bozzellajr, @claudioedantas, @wolffmoro

Tabela 4. Tabela com exemplos de retweets populares nas comunidades que promovem os quatro pré-candidatos mais populares.

Id	Candidato	Retweet Popular da Comunidade
1	J. Bolsonaro	"Essas eleições é pra limpar!! O recado é: NAO VOTE em quem votou contra o voto impresso auditável, em quem votou pela manutenção da prisão de Daniel Silveira e em quem votou pela urgência do PL2630, a PL da censura. #BolsonaroReeleito2022"
4	C. Gomes	"Chega de conversinha fiada. Vote em quem tem projeto pra transformar nosso país em um lugar mais justo. #Ciro2022 #TáNaHoraDeVocêOlharProCiro LINK.OMITIDO"
6	Lula	"Tem quem diga: estude bem os candidatos para votar nas próximas eleições. Eu digo: vote em LULA no 1º turno! Nem precisa estudar, se não quiser, porque o povo brasileiro já conhece a vida pessoal e política de Lula. Então, é só votar #LulaNoPrimeiroTurno #VamosJuntosPeloBrasil LINK.OMITIDO"
7	S. Moro	"Exato! No alvo! Nas ELEIÇÕES/2022, VOTE em CANDIDATOS "FICHA LIMPA"! Valorize o SEU VOTO! CHEGA de banditagem governando o Brasil! ELES usufruem e NÓS pagamos a CONTA! #MoroPresidente2022 OU #MoroOuNulo LINK.OMITIDO"

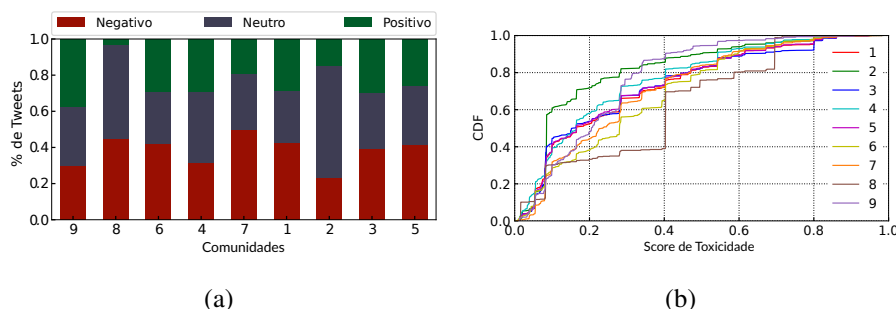
que a maior parte das contas nas comunidades aparentam ser de usuários reais engajados.

Análise Textual. A Tabela 3 apresenta as top-5 hashtags e menções mais populares dos retweets compartilhados nas 4 comunidades com o maior número de usuários e que impulsionam conteúdo de 4 pré-candidatos distintos. A menção mais popular de cada comunidade é a @ do usuário do pré-candidato mais abordado no conteúdo impulsionado. Para o caso das hashtags (por exemplo, #Bolsonaro2022, #LulaNoPrimeiroTurno, #TáNaHoraDeVocêOlharProCiro e #Moro2022) podemos verificar que as mais populares indicam a presença de conteúdo de promoção de pré-candidatos. Para exemplificar o conteúdo de promoção destes pré-candidatos, a Tabela 4 mostra exemplo de retweets populares de cada comunidade, onde, em todos os casos, aparecem conteúdo textual com pedidos de voto explícito para os pré-candidatos, o que, conforme mencionado na introdução deste trabalho, é uma prática vedada pela legislação brasileira vigente.

Com o objetivo de melhor caracterizar o conteúdo impulsionado a partir das comunidades formadas, a Figura 3 apresenta a porcentagem dos retweets classificados como positivos, negativos e neutros (Figura 3(a)) e a distribuição de probabilidade acumulada da pontuação de toxicidade dos retweets (Figura 3(b)).

Em relação ao sentimento expresso pelos usuários das comunidades (Figura 3(a)), os resultados mostram que, apesar das comunidades tenderem a promover conteúdo que expressam, na maioria, sentimento negativo, este sentimento está mais atrelado aos demais pré-candidatos do que ao candidato que a comunidade apoia. Por exemplo, a comunidade 6 de suporte ao Lula, possui 42,51% de retweets com sentimento negativo. No entanto, estes retweets são, em grande parte, conteúdo negativo relacionados à Jair Bolsonaro. O mesmo ocorre na comunidade 7 cujo principal pré-candidato mencionado foi o Sérgio Moro. Cerca de 49,7% dos retweets expressam sentimento negativo, contra 18,81% de sentimentos positivos. Ao analisar alguns retweets, verificamos que vários usuários impulsionaram um conteúdo negativo contra o lançamento da candidatura de Luciano Bivar para Presidência da República pelo União Brasil, em substituição ao pré-candidato Sérgio Moro. Comportamento similar é observado nas comunidades relacionadas à Jair Bolsonaro, principalmente nas comunidades 1, 3 e 5, onde parte do conteúdo compartilhado por estas comunidades são compostos por retweets negativos em relação à Lula. Por fim, a comunidade mais positiva é a comunidade 9 (pré-candidata Sofia Man-

Figura 3. (a) Porcentagem de retweets com sentimento positivo, negativo, neutro. (b) Distribuição de probabilidade acumulada da toxicidade dos retweets.



zano), onde parte do conteúdo impulsionado são de mensagens de suporte para a presidencializável.

A análise de toxicidade do conteúdo dos retweets é mostrada na Figura 3(b), onde é possível notar que as comunidades possuem na faixa de 70% e 90% de retweets com score $\leq 40\%$, indicando que, no geral, as comunidades não possuem um índice alto de elementos semânticos que tornem o conteúdo das comunidades tóxico, como a presença de linguagem ofensiva e abusiva, por exemplo. Entretanto, ainda é possível verificar a presença de mensagens com score mais altos de toxicidade nas comunidades, embora em uma proporção pequena, fato que nós atrelamos, principalmente, à presença de retweets com toxicidade direcionada à pré-candidatos ou personalidades políticas adversárias em relação ao pré-candidato da comunidade, como ocorre à toxicidade direcionada à Lula nas comunidades de Jair Bolsonaro, por exemplo.

4.3. Discussão dos Resultados

A partir do perfil das comunidades, definido a partir dos resultados apresentados, podemos destacar as principais características descritas a seguir.

Diversidade de Comunidades. Encontramos diversas comunidades de usuários no Twitter que promovem campanhas eleitorais em um período anterior ao permitido pela lei brasileira. As comunidades encontradas impulsionam conteúdo de então pré-candidatos à presidência da república tanto da esquerda quanto da direita. As comunidades de direita, em sua maioria, têm o foco mais específico em promover a figura do então pré-candidato Jair Bolsonaro, além de algumas delas não possuírem fronteiras bem definidas entre si, indicando a existência de usuários que servem como pontes entre essas comunidades, difundindo e impulsionando conteúdo em mais de uma das comunidades. Enquanto isso, as comunidades de esquerda se mostraram mais fragmentadas e diversificadas em termos de liderança e ideologia, promovendo conteúdo de uma variedade maior de pré-candidatos.

Tamanho das Comunidades. Constatamos também uma diferença significativa do volume de comunidades de direita e esquerda: comunidades de direita são muito mais expressivas, tanto em termos de número de usuários, quanto em termos de volume de conteúdo difundido. Essa diferença indica um maior engajamento nas comunidades de direita em promover conteúdo, especialmente em relação ao pré-candidato Jair Bolsonaro.

Organicidade dos Usuários. A análise de usuários indica que a maioria das contas nas comunidades aparenta ser de usuários reais engajados, dado que a maior parte (ao menos

50%) das contas de usuários das comunidades possuem tempo de vida maior que 1000 dias, além de terem um número não desprezível de seguidores. Essas características dão indícios que o conteúdo é propagado por meio de interações genuínas entre os usuários, o que é uma característica desejável em comunidades online. No entanto, também constatamos que um número considerável de contas de usuários possuem uma alta probabilidade de serem *bots*, sugerindo a existência de atividade automatizada ou com comportamento suspeito nas comunidades.

5. Conclusão e Trabalhos Futuros

Neste trabalho, utilizamos a metodologia proposta em dados coletados durante o período pré-eleitoral de 2022 para identificar e caracterizar as comunidades de usuários que impulsionam conteúdo de promoção de pré-candidatos no Twitter, antes do período especificado pelo Tribunal Superior Eleitoral (TSE), configurando campanhas antecipadas, prática vedada pela legislação brasileira. Identificamos comunidades que promovem candidatos de diversos posicionamentos políticos e campos ideológicos, dentre eles, direita e esquerda. Além disso, constatamos que as comunidades de direita são, em geral, maiores, tanto quando consideramos o número de usuários que compõe a comunidade quanto em relação ao volume de conteúdo impulsionado pela comunidade. Ademais, foram feitas diversas análises em dois níveis principais, de usuários e de análise textual. O perfil dos usuários apontou indícios de organicidade das comunidades, e a análise textual apresentou evidências da promoção de candidatos por parte das comunidades.

Nosso trabalho fornece indícios que a coordenação visando o impulsionamento de conteúdo pode ocorrer de diversas formas no Twitter. Além da coordenação através de retweets, que foi explorada neste trabalho, outras ações, como usuários retweetando conteúdo utilizando as mesmas hashtags, ou até mesmo compartilhando os mesmos links nos tweets, podem ser exploradas para evidenciar ações coordenadas de usuários. Para trabalhos futuros, vislumbramos a aplicação da metodologia proposta para essas outras ações de coordenação descritas anteriormente, visando observar se existem usuários recorrentes que aparecem em comunidades de diversas ações, o que fortaleceria as evidências de coordenação entre usuários. Além disso, seria interessante investigar como as campanhas antecipadas evidenciadas neste trabalho afetam o comportamento dos eleitores e a dinâmica das eleições em si.

Agradecimentos. CNPq, FAPEMIG e Ministério Público de Minas Gerais (MPMG).

Referências

- Badawy, A., Addawood, A., Lerman, K., and Ferrara, E. (2019). Characterizing the 2016 russian ira influence campaign. *Social Network Analysis and Mining*, 9:1–11.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.
- 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*.
- da Rosa Jr, J. M., Linhares, R. S., Ferreira, C. H. G., Nobre, G. P., Murai, F., and Almeida, J. M. (2022). Uncovering discussion groups on claims of election fraud from twitter. In *Proc. of the SocInfo*.

- Danaditya, A., Ng, L. H. X., and Carley, K. M. (2022). From curious hashtags to polarized effect: profiling coordinated actions in indonesian twitter discourse. *Social Network Analysis and Mining*, 12(1).
- David, E. and Jon, K. (2010). *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, USA.
- Ferrara, E., Chang, H., Chen, E., Muric, G., and Patel, J. (2020). Characterizing social media manipulation in the 2020 us presidential election. *First Monday*.
- Ferreira, C. H., Murai, F., Silva, A. P., Almeida, J. M., Trevisan, M., Vassio, L., Mellia, M., and Drago, I. (2021). On the dynamics of political discussions on instagram: A network perspective. *Online Social Networks and Media*, 25:100155.
- Ferreira, C. H., Murai, F., Silva, A. P., Trevisan, M., Vassio, L., Drago, I., Mellia, M., and Almeida, J. M. (2022). On network backbone extraction for modeling online collective behavior. *Plos one*, 17(9).
- Giglietto, F., Righetti, N., Rossi, L., and Marino, G. (2020). It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 italian elections. *Information, Communication & Society*, 23(6):867–891.
- Grimminger, L. and Klinger, R. (2021). Hate towards the political opponent: A twitter corpus study of the 2020 us elections on the basis of offensive speech and stance detection. In *Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*.
- 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*.
- Keller, F., Schoch, D., Stier, S., and Yang, J. (2020). Political astroturfing on twitter: How to coordinate a disinformation campaign. *Political Communications*.
- Linhares, R. S., da Rosa Jr, J. M., Ferreira, C. H. G., Nobre, G. P., Murai, F., and Almeida, J. M. (2022). Uncovering coordinated communities on twitter during the 2020 us election. In *Proc. of the ASONAM*.
- Newman, M. E. J. and Girvan, M. (2004). Finding and evaluating community structure in networks.
- Ng, L. H. X. and Carley, K. M. (2022). Online coordination: methods and comparative case studies of coordinated groups across four events in the united states. In *Proc. of the WebScience*.
- Nobre, G. P., Ferreira, C. H. G., and Almeida, J. M. (2020). Beyond groups: uncovering dynamic communities on the whatsapp network of information dissemination. In *Proc. of the SocInfo*.
- Pacheco, D., Flammini, A., and Menczer, F. (2020). Unveiling coordinated groups behind white helmets disinformation. In *Proc. of the Web Conference (WWW)*.
- Saldanha, R. S., Rosa, J. M., Ferreira, C. H., Nobre, G., Murai, F., and Almeida, J. (2022). Uncovering coordinated communities on twitter during the 2020 us election. In *Proc. of the ASONAM*.
- Silva, M., Guimaraes, S., Caetano, J., Araújo, M., Santos, J., Reis, J. C., Silva, A., Benevenuto, F., and Almeida, J. (2021). Propaganda eleitoral antecipada: Uma análise de postagens em mídias sociais. In *Anais do BrasNAM*.
- Tardelli, S., Avvenuti, M., Tesconi, M., and Cresci, S. (2022). Detecting inorganic financial campaigns on twitter. *Information Systems*, 103:101769.
- Vargas, L., Emami, P., and Traynor, P. (2020). On the detection of disinformation campaign activity with network analysis. In *Proc. of the SIGSAC*.
- Varol, O., Ferrara, E., Menczer, F., and Flammini, A. (2017). Early detection of promoted campaigns on social media. *EPJ Data Science*, 6:1–19.
- Ángeles Serrano, M., Boguñá, M., and Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proc. of the National Academy of Sciences*, 106(16):6483–6488.