

# Um *Framework* para Análise Bidimensional de Disseminação de Informações em Plataformas de Mídias Sociais

Geovana S. Oliveira<sup>1</sup>, Otávio Venâncio<sup>3</sup>, Vinícius Vieira<sup>4</sup>, Jussara Almeida<sup>3</sup>

Ana P. C. Silva<sup>3</sup>, Ronan Ferreira<sup>2</sup>, Carlos H. G. Ferreira<sup>1</sup>

geovana.so@aluno.ufop.edu.br, otavio.venancio@dcc.ufmg.br, vinicius@ufsj.edu.br, jussara@dcc.ufmg.br

ana.coutosilva@dcc.ufmg.br, ronan.ferreira@ufop.edu.br, chgferreira@ufop.edu.br

<sup>1</sup>Departamento de Computação e Sistemas, Universidade Federal de Ouro Preto, Brasil

<sup>2</sup>Departamento de Ciências Exatas e Aplicadas, Universidade Federal de Ouro Preto, Brasil

<sup>3</sup>Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Brasil

<sup>4</sup>Departamento de Ciência da Computação, Universidade Federal de São João Del Rey, Brasil

## ABSTRACT

Social media platforms have revolutionized how people connect and share information, but they also bring challenges when it comes to information dissemination. Many studies in the literature examine this phenomenon using network models. However, they often focus on a unidimensional analysis, considering only the volume of interactions on the edges, which does not fully capture the different aspects of this phenomenon, especially concerning the speed of dissemination. In this work, we propose a framework that allows for the bidimensional analysis of information dissemination on social media platforms, taking into account both the volume and the speed of interactions. Our framework is based on backbone extraction techniques to identify the most salient edges in both dimensions and classifies the edges into different dissemination profiles, allowing a detailed analysis of the topology and community presence for each profile. We applied it to two case studies covering critical information dissemination scenarios, notably on Twitter/X and Telegram. Our results show that the proposed framework is able to uncover different patterns of information dissemination. This emphasizes the importance of considering multiple dimensions simultaneously for a deeper understanding of the phenomenon.

## KEYWORDS

Mídias Sociais, Disseminação de Informação, Modelagem de Rede

## 1 INTRODUÇÃO

As plataformas de mídias sociais revolucionaram a maneira como as pessoas se conectam, comunicam e obtêm informações. Dessa forma, esses ambientes têm sido amplamente estudados para entender fenômenos de comportamento coletivo, como a disseminação de desinformação e a manipulação da opinião pública por meio de ações coordenadas [23, 37, 44]. Uma abordagem frequentemente usada para modelar e analisar esses fenômenos é o uso de grafos que representam redes de interações [17, 25]. Em termos simples, um grafo é uma representação matemática de um conjunto de objetos (nós) e das relações (arestas) entre eles [2]. Nesse contexto, é possível modelar redes que capturem diferentes tipos de interação,

como, por exemplo, *co-retweets* no Twitter/X, onde os usuários são representados por nós e as arestas representam o número de *tweets* em comum que esses usuários compartilharam [21].

Sabe-se que essas redes têm um grande conjunto de arestas que podem ser provenientes de relações fracas e esporádicas, que podem representar relações irrelevantes para a compreensão do fenômeno estudado [17, 24]. Portanto, a extração de *backbone* tem sido uma técnica crucial para realçar as arestas mais salientes (relevantes), que representam relações mais fortes e permitem o entendimento dos aspectos mais importantes com relação a um dado estudo [12, 14, 28–31]. Assim, a partir da seleção das arestas salientes para um determinado problema ou fenômeno, obtém-se uma versão reduzida, porém mais representativa, da rede original, que é chamada de *backbone* da rede [8]. Diversos métodos de extração de *backbone* existem com diferentes propósitos e aplicabilidades, muitos deles voltados para investigar aspectos relacionados à disseminação de informação coletiva em plataformas de mídias sociais [12, 17, 21, 24].

Embora esses métodos tenham sido utilizados com sucesso para modelar e analisar a disseminação de informações em plataformas de mídias sociais, eles são constantemente aplicados para a obtenção do *backbone* de forma unidimensional [9, 12, 14, 21, 28, 30, 42, 44]. Em outras palavras, apenas o número de interações de usuários, isto é, o volume, na plataforma é levado em consideração. Por exemplo, pode-se estudar o número de *tweets* em comum no Twitter/X que dois usuários compartilharam formando uma aresta de *co-retweets* [21, 31] ou, similarmente, o número de mensagens semelhantes compartilhadas por dois usuários em grupos distintos no WhatsApp ou Telegram [28, 29, 42]. Após extrair o *backbone* em modelos desse tipo, é possível caracterizar a topologia da rede e a forma como os usuários se organizam em comunidades. Isso permite uma análise dos seus padrões de interação, o que pode revelar comportamentos sincronizados e coordenados entre grupos. No entanto, ao utilizar essa forma unidimensional de modelagem e análise de padrões topológicos, uma importante dimensão é desconsiderada, em particular, a *velocidade* com que a propagação de informação ocorre. Essa é uma característica crucial em plataformas de mídias sociais, capaz de trazer grandes impactos à natureza da disseminação do conteúdo nas diferentes plataformas [4, 11, 20, 41].

Neste contexto, este trabalho tem como objetivo estabelecer um *framework* para o estudo de disseminação de informações em plataformas de mídias sociais que considere tanto o volume das co-interações dos usuários, como a velocidade com que eles atuam.

Especificamente, busca-se caracterizar perfis de disseminação de informação, explorando uma lacuna deixada por outros trabalhos na literatura e avaliando o fenômeno de propagação de conteúdo de forma bidimensional. A premissa principal explorada aqui é que os usuários que utilizam esses ambientes podem ser classificados com base no volume do conteúdo com os quais co-interagem e na velocidade com a qual isso é feito. A partir dessa classificação, usuários de uma mesma classe podem ser caracterizados de forma global (distribuição de conectividade) e local (formação de comunidades). Considerando a forma como se organizam, acredita-se que seja possível revelar padrões de disseminação de informação distintos. Para isso, consideramos as seguintes etapas: inicialmente, modelamos redes de co-interações, uma considerando o volume e outra considerando a velocidade dessas interações. Em seguida, aplicamos técnicas de extração de *backbone* para identificar as arestas mais salientes nas duas redes, simultaneamente. Após, classificamos as arestas de acordo com o julgamento dos métodos de *backbone* em diferentes perfis. Essa abordagem nos permite categorizar as arestas que ligam dois dados usuários em quatro perfis de disseminação de informações: (1) usuários que disseminam um baixo volume de informações e de forma lenta; (2) usuários que disseminam um alto volume de informações, mas de forma lenta; (3) usuários que disseminam um baixo volume, porém de forma rápida; e (4) usuários que disseminam um alto volume de informações e de forma rápida. Por fim, caracterizamos a topologia observada em sub-redes formadas por usuários desses quatro perfis de interações, investigando a estrutura de comunidades delas.

Aplicamos nosso *framework* em dois estudos de caso nas plataformas Twitter/X e Telegram durante períodos que contêm fortes evidências de coordenação [1, 15, 42]. Nossos principais resultados mostram que o *framework* proporciona um aprofundamento da compreensão do fenômeno de disseminação de informações a partir de contas de usuários em plataformas de mídias sociais. Observamos diferentes perfis de disseminação de informações, caracterizados tanto pelo volume quanto pela rapidez das interações. Destacamos que no Twitter/X as comunidades são mais coesas e com alto volume de *retweets*, enquanto no Telegram as interações se distribuem em comunidades menores, com um volume menor, mas com maior rapidez de disseminação durante eventos críticos.

O restante do trabalho está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados. A Seção 3 detalha a metodologia, enquanto a Seção 4 descreve os estudos de casos. Em seguida, a Seção 5 apresenta os resultados e, por fim, a Seção 6 traz as conclusões e direções futuras.

## 2 TRABALHOS RELACIONADOS

Vários estudos anteriores empregaram modelos de rede para detectar e analisar a disseminação de informações em plataformas de mídias sociais, como Twitter/X, Telegram, WhatsApp, Instagram, etc [3, 7, 10, 16, 18, 19, 35, 39, 40, 43]. Ainda, um grande número de trabalhos destaca a importância de se identificar e remover arestas que possam refletir relações esporádicas e aleatórias que podem representar ruído ao modelo de redes. Com essa remoção, o modelo concentra-se apenas em arestas relevantes para o fenômeno de disseminação de informações [9, 12–14, 27–31, 42]. Dessa forma, as arestas salientes revelam o que é chamado de *backbone*: uma estrutura subjacente à rede inicialmente modelada que busca fornecer

uma visão mais clara e organizada para identificar padrões topológicos que auxiliem a compreensão da propagação de conteúdo.

Por exemplo, Pacheco et al. [30] investigaram campanhas de desinformação contra a Defesa Civil Síria no Twitter/X, explorando um modelo de rede que conecta usuários que postaram *tweets* similares. Os autores empregaram a extração de *backbone* para remover arestas, cujos pesos estão abaixo de um determinado limiar (limiarização). Em outro trabalho, os mesmos autores propuseram identificar ações coordenadas nas mídias sociais, notadamente no Twitter/X, em diferentes cenários, também empregando a extração de *backbone* por meio de limiarização [31]. Keller et al. [19] exploraram a rede de *retweets* para estudar a coordenação de anúncios durante a eleição presidencial na Coreia do Sul também explorando limiarização.

Linhares et al. [21] e da Rosa et al. [9] investigaram comunidades de usuários no Twitter/X que foram, potencialmente, usadas, para disseminação coordenada de conteúdos relacionados a alegação de fraude durante a eleição dos EUA de 2020, usando métodos como *Disparity Filter* e *Tripartite Backbone Extraction (TriBE)* [12, 17, 34]. Na mesma direção, Araujo et al. [1] estudaram a promoção coordenada de campanhas políticas antecipadas no Twitter/X durante o período pré-eleitoral brasileiro de 2022, utilizando uma rede de *co-retweets* e o *Disparity Filter + Neighborhood Overlapping*. Nobre et al. [28] investigaram a rede de co-compartilhamento de mídia, construída conectando usuários que compartilharam o mesmo conteúdo, também com o uso do *Disparity Filter* para estudar comunidades. Também usando o *Disparity Filter* e *Disparity Filter + Neighborhood Overlapping* para revelar arestas salientes e com evidências de coordenação no Telegram, Venâncio et al. [42] investigaram a disseminação de informações políticas durante as eleições brasileiras de 2022 e os tumultos de 8 de janeiro.

Embora esses esforços explorem a extração de *backbone* para compreender a disseminação de informações de forma estruturada na rede, um aspecto desconsiderado por esses estudos é a distinção das formas como a disseminação pode ocorrer. Especificamente, os esforços supramencionados focam em entender a disseminação apenas de um ponto de vista, do volume. Dessa forma, há uma lacuna na literatura quanto a estratégias para identificar diferentes perfis de disseminação de informações por esses grupos de usuários potencialmente coordenados e compreender o sincronismo considerando duas dimensões: o volume e a velocidade da informação disseminada, o que é explorado neste trabalho.

## 3 FRAMEWORK PROPOSTO

O *framework* proposto neste trabalho consiste em várias etapas, conforme ilustrado na Figura 1. Inicialmente, modelamos duas redes, uma que representa o volume de co-interações e outra que representa o intervalo de tempo entre essas co-interações. Em seguida, aplicamos métodos de extração de *backbone* para identificar as arestas mais relevantes em cada modelo de rede. Então, classificamos as arestas em perfis distintos de disseminação de informações, com base em sua relevância, usando métodos de extração de *backbone*. Finalmente, caracterizamos e analisamos a topologia das redes formada por essas classes incluindo a análise de comunidades.

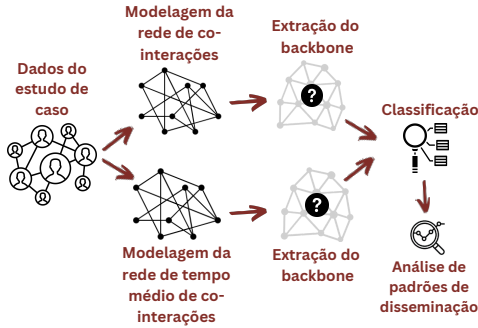


Figura 1: Visão geral do *framework* proposto.

### 3.1 Modelagem das Redes

A metodologia proposta neste trabalho parte da definição de dois modelos de redes: um que representa o volume de co-interações e outro que representa o intervalo tempo entre essas interações. A primeira rede de co-interações é representada por um grafo  $G_{\text{volume}} = (V, E_{\text{volume}})$ , tal que cada elemento  $v_i \in V$  representa um usuário, e cada elemento  $e_{ij} \in E_{\text{volume}}$  é uma aresta que conecta um par de usuários. O peso da aresta  $w_{\text{volume}}(e_{ij})$  representa o número de co-interações entre os usuários  $v_i$  e  $v_j$ .

Já na rede  $G_{\text{tempo}} = (V, E_{\text{tempo}})$ , as arestas  $e_{ij} \in E_{\text{tempo}}$  armazenam informação sobre o tempo decorrido entre as interações dos usuários representados por  $v_i$  e  $v_j$ . Embora outras funções possam ser utilizadas para agregar os tempos de todas as co-interações de  $v_i$  e  $v_j$ , neste trabalho, o peso  $w_{\text{tempo}}(e_{ij})$  de cada aresta  $e_{ij}$  refere-se ao tempo médio das interações dos usuários representados por  $v_i$  e  $v_j$ . Assim,  $w_{\text{tempo}}(e_{ij}) = \left[ \frac{1}{w_{\text{volume}}(e_{ij})} \sum_{p=1}^{w_{\text{volume}}(e_{ij})} t_{ij}^p \right]$ , em que  $p$  é o tempo decorrido na  $p$ -ésima interação entre  $v_i$  e  $v_j$ .

Os algoritmos utilizados neste trabalho (tanto para extração de *backbone*, como para identificação de comunidades) consideram a noção de que o peso de uma aresta  $e_{ij}$  deve ser diretamente proporcional à importância da relação entre  $v_i$  e  $v_j$ , o que é contra-intuitivo com a definição da função de peso  $w_{\text{tempo}}$  apresentada anteriormente. Um tempo curto entre as ações de  $v_i$  e  $v_j$  de enviarem uma mesma mensagem em uma rede social, por exemplo, deveria sugerir uma forte relação entre  $v_i$  e  $v_j$ , e o contrário é definido por  $w_{\text{tempo}}$ . Para que isso seja alcançado, os pesos são normalizados de acordo com a Equação 1, garantindo que relações mais fortes sejam representadas por valores de peso mais altos. Assim, o peso normalizado  $w'_{\text{tempo}}(e_{ij})$  é dado por:

$$w'_{\text{tempo}}(e_{ij}) = \max(W_{\text{tempo}}) - w_{\text{tempo}}(e_{ij}) + 1 \quad (1)$$

em que  $\max(W_{\text{tempo}})$  é o maior peso encontrado na rede  $G_{\text{tempo}}$ . O termo  $+1$  evita que haja pesos nulos de forma indevida. Para simplificar, o peso das arestas  $w_{\text{tempo}}$  irá referir-se à definição de peso normalizado apresentada na Equação 1 no restante deste trabalho.

### 3.2 Extração dos Backbones

A segunda etapa do *framework* consiste em extrair os *backbones* de ambas as redes modeladas na etapa anterior. Diversos métodos de extração de *backbone* propostos na literatura são aplicados para modelar padrões de disseminação, tipicamente explorando a discrepância e heterogeneidade das redes. Exemplos incluem *Threshold*

[30], *Disparity Filter* [34], *Disparity Filter + Neighborhood Overlap* [21] e *Polya Urn* [22]. Em resumo, esses métodos são usados para explorar a heterogeneidade da rede, identificando quão salientes são as arestas com pesos significativamente maiores, com base em padrões individuais (locais) ou de rede (globais), como representativas de interações persistentes e repetitivas. Com exceção do *Thresholding*, que é um método puramente estrutural e opera de forma global, excluindo arestas com base em um limiar da distribuição dos pesos das arestas da rede, os demais são métodos probabilísticos que constroem modelos nulos para cada nó e são capazes de capturar, do ponto de vista local, o quão saliente uma aresta é de acordo com o peso. Uma revisão do uso destes métodos para este contexto está disponível na literatura [17].

Nosso principal argumento nessa etapa reside na hipótese de que algumas arestas representarão co-interações fortes e consistentes durante o período observado, enquanto outras serão esporádicas, podendo ainda terem sido observadas em um intervalo de tempo excepcionalmente curto ou um intervalo grande. Então, vamos analisar o fenômeno sob a perspectiva do volume de co-interações e do intervalo de tempo em que ocorrem. Para isso, considere novamente os seguintes cenários em uma plataforma de rede social: (1) uma mensagem é disseminada massivamente, compartilhada por muitos usuários pouco ativos na plataforma, formando assim arestas de co-interações com pesos pequenos; e (2) um outro conjunto de mensagens é disseminado de forma consistente e repetitiva por um outro conjunto de usuários, nesse caso formando arestas de pesos maiores. De maneira análoga, podemos considerar o aspecto temporal da disseminação dessas mensagens, buscando compreender se isso ocorre de maneira muito rápida ou não. Ambas as perspectivas, – pelo volume de co-interações e pelo tempo em que ocorrem – já foram mostradas como importantes no processo de disseminação de informações de forma coordenada na literatura. Existem fortes evidências da alocação de contas temporárias, pagas ou não, para disseminarem *tweets* específicos de forma rápida [4, 11, 20, 41] e da atuação de grupos de usuários disseminando informações de forma sistemática e consistente [21, 29, 42] em diversas plataformas, como o Twitter/X. Porém, a incorporação conjugada dessas duas dimensões à análise do fenômeno de disseminação de informação não é utilizada em outros trabalhos na literatura e pode ser apontada como uma novidade trazida pelo *framework* proposto.

Neste trabalho, propomos o uso do método de *Polya Urn Filter* [22], devido à sua flexibilidade e capacidade de considerar a importância local das arestas na identificação do *backbone*. No entanto, vários métodos de extração de *backbone* para identificação de atividades relacionadas à coordenação estão na literatura e podem empregados [17]. Em resumo, *Polya Urn* assume que os pesos das arestas surgem do processo agregado das preferências individuais dos nós para interagir entre si ao longo do tempo. Ele também pressupõe que as interações entre os nós são mantidas e reforçadas, de modo que quanto maior o número de interações entre dois nós, maior a probabilidade de eles interagirem novamente. Um modelo de referência é construído para cada aresta, capaz de capturar o reforço das interações existentes, examinando o grau e a força (a soma dos pesos de todas as arestas incidentes ao nó) de cada nó conectado a essa aresta. Esse mecanismo de reforço pode ser regulado e estimado por meio de um processo de ajuste fino. Arestas salientes são aquelas cujos pesos desviam de forma significativa do

modelo de referência com valores maiores, de acordo com um nível de confiança definido [22]. Dessa forma, o *Polya Urn* é aplicado ao grafo  $G_{\text{volume}}$ , gerando o grafo  $B_{\text{volume}}$  que captura os nós e arestas com padrões de co-interações mais altos do que o esperado.

De maneira análoga, o método de *backbone* também deve ser aplicado ao grafo  $G_{\text{tempo}}$ . É importante apontar que os pesos  $w_{\text{tempo}}$  são normalizados de acordo com a Equação 1, para que o método de *backbone* funcione como foi projetado e capture arestas com pesos mais altos do que o esperado [17, 22]. Assim, a aplicação do método *Polya Urn* ao grafo  $G_{\text{tempo}}$ , gera o *backbone*  $B_{\text{tempo}}$ , que captura as arestas com tempos médios de disseminação que fogem ao esperado. Ao extrair os *backbones* das duas redes  $G_{\text{volume}}$  e  $G_{\text{tempo}}$ , podemos classificar as arestas mantidas como representativas de interações fortes e consistentes, enquanto as arestas removidas representam interações esporádicas e menos significativas. Dessa forma, é possível extrair da rede os nós e arestas mais significativos tanto em volume de informação quanto em tempo de disseminação, classificados conforme explicado na próxima seção.

### 3.3 Classificação das Arestas

Ambos os modelos de redes considerados pela metodologia proposta (representados por  $G_{\text{volume}}$  e  $G_{\text{tempo}}$ ) correspondem ao mesmo conjunto de arestas, associando cada par de vértices que representam usuários que co-interagiram nas redes sociais. É possível, então, definir quatro classes de arestas, com base na sua presença ou ausência em  $B_{\text{volume}}$  e  $B_{\text{tempo}}$ , após a identificação dos *backbones* a partir de  $G_{\text{volume}}$  e  $G_{\text{tempo}}$ , respectivamente. Formalmente, cada classe de arestas é definida como:

- **Classe 1** (baixo volume e baixa velocidade):

$$C1 = (E_{\text{volume}} - E_{\text{volume}}^B) \cap (E_{\text{tempo}} - E_{\text{tempo}}^B) \quad (2)$$

Essas arestas não são mantidas em nenhum dos *backbones*.

- **Classe 2** (alto volume e baixa velocidade):

$$C2 = E_{\text{volume}}^B \cap (E_{\text{tempo}} - E_{\text{tempo}}^B) \quad (3)$$

Essas arestas são mantidas em  $B_{\text{volume}}$  mas não em  $B_{\text{tempo}}$ .

- **Classe 3** (baixo volume e alta velocidade):

$$C3 = (E_{\text{volume}} - E_{\text{volume}}^B) \cap E_{\text{tempo}}^B \quad (4)$$

Essas arestas são mantidas em  $B_{\text{tempo}}$  mas não em  $B_{\text{volume}}$ .

- **Classe 4** (alto volume e alta velocidade):

$$C4 = E_{\text{volume}}^B \cap E_{\text{tempo}}^B \quad (5)$$

Essas arestas são mantidas em ambos os *backbones*.

Como ilustração, no contexto do Twitter/X, as arestas classificadas podem ser interpretadas da seguinte forma: Classe 1 indica que dois usuários não retuitaram um alto volume de *tweets* e quando isso acontece, é feito em um grande intervalo de tempo médio, sugerindo baixo volume e velocidade de disseminação. Classe 2 indica que dois usuários retuitaram muitos *tweets*, mas com um grande intervalo de tempo entre eles, sugerindo um volume alto, porém pouco veloz de disseminação. Classe 3 indica que dois usuários retuitaram poucos *tweets* e com um intervalo muito curto de tempo entre eles, sugerindo alta velocidade de disseminação, mas com baixo volume. Classe 4 indica que dois usuários retuitaram muitos *tweets* em um curto intervalo de tempo, sugerindo uma forte coordenação entre eles. Em resumo, a abordagem proposta aqui difere das

metodologias aplicadas na literatura por permitir a diferenciação de perfis e estruturas além de uma visão unidimensional.

### 3.4 Análise de Padrões de Disseminação

Considerando a classificação das arestas descrita na seção anterior, quatro sub-redes são identificadas, cada uma relacionada a uma das classes identificadas com base no volume e na velocidade das interações entre os vértices. Assim, podemos definir perfis de disseminação tomando como ponto de partida a análise dos padrões observados nas diferentes classes, investigando a organização topológica que emerge de cada sub-rede, assim como semelhanças e diferenças entre elas. Em particular, como se dá a organização dos nós dessas sub-redes em estruturas de comunidades, isto é, subconjuntos de nós com conexões mais densas entre si do que com o restante da rede. A investigação da organização de nós em comunidades tem sido aplicada com sucesso no domínio de plataformas de mídias sociais para a identificação de ideologias específicas, caracterização de campanhas políticas antecipadas, detecção de desinformação ou notícias falsas [1, 14, 28, 29]. Embora diferentes métodos possam ser empregados para identificação de comunidades de nós em redes com as características dos modelos aqui propostos [36], neste trabalho é utilizado o algoritmo de Louvain [5], amplamente utilizado em outros trabalhos na literatura.

O algoritmo de Louvain visa maximizar a modularidade, que mede a qualidade da divisão de uma rede em comunidades. Modularidade avalia o quão densamente conectados os nós estão dentro das comunidades, em comparação com uma rede aleatória com a mesma distribuição de conexões. O valor da modularidade varia de -0,5 a 1, onde valores acima de 0,3 indicam comunidades bem definidas. O algoritmo é heurístico e funciona de forma aglomerativa. Ele começa atribuindo cada nó da rede a uma comunidade própria. Em seguida, o algoritmo move nós entre comunidades para aumentar a modularidade. Quando não é possível melhorar mais a modularidade, as comunidades formadas são tratadas como novos nós, e o processo recomeça. O resultado final é a estrutura de comunidades com a maior modularidade encontrada.

Neste trabalho, limitamos o escopo dos experimentos à análise topológica da estrutura das redes. Entretanto, encorajamos uma compreensão mais profunda do fenômeno de disseminação de informação por meio da análise dos conteúdos das mensagens enviadas em cada um dos grupos, que pode evidenciar padrões de comportamento, estratégias de disseminação de desinformação, ou esforços coordenados para promover determinadas informações.

## 4 ESTUDOS DE CASO

Nesta seção, descrevemos os dois estudos de casos explorados à luz do *framework* descrito na seção anterior.

### 4.1 Twitter/X

Nosso primeiro estudo de caso consiste na análise de disseminação de informações no Twitter, recentemente redesenhado como X<sup>1</sup>. O Twitter/X tem sido amplamente usado para disseminação de informações, especialmente em eventos políticos [9, 21, 30, 32, 33]. Aqui, concentramos nossa análise em torno de eventos relacionados às

<sup>1</sup>www.x.com

eleições brasileiras de 2022, em uma coleção de dados disponibilizados em [15]. Essa coleção é composta por dados coletados entre 1<sup>o</sup> de outubro de 2022 e 10 de fevereiro de 2023, período que abrangeu eventos-chave das eleições gerais brasileiras de 2022, como os dois turnos de votação, apuração dos resultados, pesquisas eleitorais, debates e o ataque aos prédios da Câmara dos Deputados, Senado e do Superior Tribunal Federal em 8 de janeiro de 2023. Durante esse período, já se observava a disseminação de informações de conteúdos antidemocráticos no Twitter<sup>2</sup>.

Para analisar as interações dos usuários, consideramos os modelos de redes para duas janelas de tempo diárias  $T = \{01/11/2022, 08/01/2023\}$ , que correspondem aos dois dias de maior atividade observada. Dessa forma, temos os grafos  $G_{\text{volume}}^{01/11/2022}$  e  $G_{\text{tempo}}^{01/11/2022}$ , correspondentes aos usuários com atividade na data de 01/11/2022 e suas respectivas co-interações nessa data, e os grafos  $G_{\text{volume}}^{08/01/2023}$  e  $G_{\text{tempo}}^{08/01/2023}$ , correspondentes aos usuários com atividade na data de 01/11/2023 e suas respectivas co-interações nessa janela de tempo. Os pesos  $w(e_{ij})$  das ligações entre os usuários  $v_i$  e  $v_j$  são calculados de acordo com o modelo definido na Seção 3.1. Os parâmetros do método *Polya Urn* para a extração dos *backbones* neste estudo de caso foram definidos conforme as diretrizes de [17, 22], utilizando o parâmetro de reforço  $a = 0.5$  para  $G_{\text{volume}}$  e para  $G_{\text{tempo}}$ , e  $\alpha = 0.05$  para  $G_{\text{volume}}$  e  $\alpha = 0.1$  para  $G_{\text{tempo}}$ .

## 4.2 Telegram

Nosso segundo estudo de caso diz respeito à disseminação de informações em grupos de Telegram. A plataforma conecta usuários em conversas ponta a ponta e também em grupos, que têm se mostrado eficazes para a grande disseminação de informações [6, 10, 26, 38, 39]. Neste estudo de caso, utilizamos mensagens de grupos politicamente orientados do Telegram no Brasil, publicadas entre 1<sup>o</sup> de setembro de 2022 e 31 de janeiro de 2023, incluindo o período das eleições presidenciais e os tumultos de janeiro de 2023, em uma coleção de dados disponibilizada em [42]. Para analisar a disseminação dos usuários para disseminar conteúdo nos grupos do Telegram, utilizamos os modelos de rede originalmente usados pelos autores, que conecta usuários que compartilharam o mesmo conteúdo textual entre os grupos observados. Além disso, analisamos três janelas de tempo de 15 dias, sendo 7 antes e 7 depois em torno dos seguintes eventos para análise:  $T = \{01/10/2022(1^{\circ}\text{Turno}), 30/10/2022(2^{\circ}\text{Turno}), 08/01/2023(\text{Tumultos})\}$ , representativas dos dois turnos da eleição e o tumulto de 8 de janeiro de 2023.

Consideramos, então, os modelos de redes correspondentes às janelas de tempo:  $G_{\text{volume}}^{1^{\circ}\text{Turno}}$  e  $G_{\text{tempo}}^{1^{\circ}\text{Turno}}$ ,  $G_{\text{volume}}^{2^{\circ}\text{Turno}}$  e  $G_{\text{tempo}}^{2^{\circ}\text{Turno}}$ , e  $G_{\text{volume}}^{\text{Tumultos}}$  e  $G_{\text{tempo}}^{\text{Tumultos}}$ , correspondentes aos usuários que postaram ao menos uma mensagem em um dos grupos do Telegram durante a janela especificada. Cada aresta  $e_{ij}$  indica que os usuários  $v_i$  e  $v_j$  compartilharam a mesma mensagem (conteúdo textual idêntico) pelo menos uma vez dentro da mesma janela de tempo. Os parâmetros do método *Polya Urn* para a extração dos *backbones* foram novamente definidos conforme as diretrizes de [17, 22], utilizando o parâmetro de reforço  $a = 0.5$  e  $\alpha = 0.05$  para  $G_{\text{volume}}$  e  $\alpha = 0.1$  para  $G_{\text{tempo}}$ .

**Tabela 1: Topologia das redes do Twitter/X e Telegram.**

Rede	Data	$n$	$m$	$d$	$\hat{k}$	CMA	CC	# Com.	$Q$
Twitter/X	08/01/2023	28,247	86,478,838	0.216	6123,045	0.689	1	5	0.232
Twitter/X	01/11/2022	27,920	72,777,139	0.186	5213,262	0.700	1	5	0.487
Telegram	1 <sup>o</sup> Turno	1,625	81,236	0.062	99,983	0.641	80	110	0.117
Telegram	2 <sup>o</sup> Turno	3,444	161,383	0.027	93,718	0.621	164	193	0.123
Telegram	Tumultos	4,772	71,046	0.006	29,776	0.549	135	155	0.349

## 5 RESULTADOS

Esta seção apresenta nossos resultados. Primeiro, mostramos a análise topológica das redes estudadas. Em seguida, analisamos os padrões de disseminação de acordo com o *framework* proposto.

### 5.1 Análise Topológica das Redes

A Tabela 1 apresenta as características estruturais em cada rede originalmente modelada, isto é, sem a etapa de extração de *backbone*, tanto do Twitter/X como do Telegram.  $n$  é o número de nós em cada rede,  $m$  é o número de arestas,  $d$  é a densidade das redes,  $\hat{k}$  é o grau médio dos nós, CMA é o coeficiente médio de agrupamento dos nós, CC é o número de componentes conectadas, # Com. é o número de comunidades encontradas e  $Q$  é a modularidade da partição.

Ao comparar as redes do Twitter/X e Telegram, observamos características distintas em suas topologias, que é resultado da própria diferença entre os modelos para a construção das redes. Nas redes do Twitter/X, as arestas entre vértices são criadas quando seus respectivos usuários *retweetam* um mesmo conteúdo, o que acontece com uma frequência bastante elevada, fazendo com que o número de arestas (e, consequentemente, a densidade e o grau médio) seja bastante elevado. Nas redes do Telegram, uma aresta só é criada quando um par de usuários publica mensagens de conteúdo idêntico e, ainda assim, observa-se uma alta quantidade de arestas (mesmo que muito menor que o observado nas redes do Twitter/X). Essa alta densidade observada nas redes em ambos os estudos de caso (e especialmente no Twitter/X) representa um forte indicativo da necessidade da aplicação de alguma estratégia para filtragem das relações mais importantes para uma melhor compreensão do fenômeno de disseminação de informação, o que neste trabalho é realizado através da aplicação do método para extração de *backbone*. A presença de um alto número de arestas e, potencialmente, uma grande presença de relações ruidosas, prejudica, inclusive, a identificação da organização dos vértices em comunidades, já que observa-se valores de modularidade que refletem grupos pouco claros. Mesmo nas redes do Telegram, que possuem suas componentes bastante fragmentadas, nota-se valores de modularidade baixos, correspondentes à divisão dos vértices em grupos no máximo fracamente evidentes, como no caso da janela de tempo dos Tumultos.

### 5.2 Extração dos Padrões de Disseminação

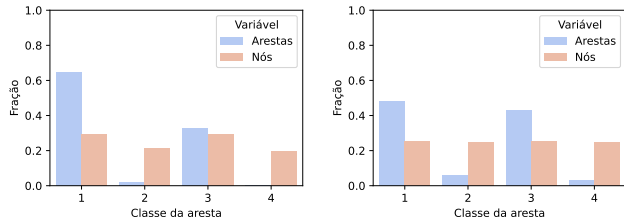
**5.2.1 Twitter.** Movendo para a extração dos perfis de disseminação e caracterização das redes, as características topológicas de rede, bem como da estrutura de comunidades para cada classe das redes do Twitter/X, são apresentadas na Tabela 2.

De imediato, nota-se que há, para as classes nas duas janelas de tempo, um equilíbrio no número de vértices, que não é observado para as arestas das redes. A Figura 2 ilustra a distribuição de vértices

<sup>2</sup><https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=508935&tip=UN>, <https://www.bbc.com/portuguese/brasil-63990040>

**Tabela 2: Topologia das redes do Twitter/X por classe.**

Classe	Data	$n$	$m$	$d$	$\hat{k}$	CMA	CC	# Com.	$Q$
Classe 1	01/11/2022	27,911	47,173,429	0.121	3380,275	0.467	1	6	0.494
Classe 2	01/11/2022	20,363	1,373,833	0.0066	134,934	0.586	17	24	0.434
Classe 3	01/11/2022	27,899	23,980,819	0.061	1719,116	0.356	2	4	0.471
Classe 4	01/11/2022	18,453	249,058	0.0014	26,993	0.271	29	48	0.623
Classe 1	08/01/2023	28,230	41,729,887	0.104	2956,421	0.300	1	5	0.257
Classe 2	08/01/2023	27,401	4,940,636	0.013	360,617	0.397	12	21	0.311
Classe 3	08/01/2023	28,236	37,192,895	0.093	2634,430	0.389	1	6	0.323
Classe 4	08/01/2023	27,578	2,615,420	0.0068	189,674	0.350	8	14	0.342

**Figura 2: Fração de nós e arestas por classe nas redes do Twitter/X.**

e arestas entre as classes e nos ajuda a ter uma melhor compreensão das características topológicas das redes modeladas para o Twitter/X. Ademais, existe um desequilíbrio no número de arestas das redes de diferentes classes. Enquanto a rede para a Classe 1, que representa relações que não são salientes nem para a dimensão volume e nem para a dimensão tempo, apresenta um elevado número de arestas, a rede para a Classe 4, que representa relações que são excepcionalmente relevantes em ambas as dimensões, apresenta um número bastante inferior de arestas, especialmente na janela referente à data de 01/11/2022. Isso se reflete em algumas informações exibidas na Tabela 2, como os graus médios ( $e$ , consequentemente as densidades), que são substancialmente mais altas para a Classe 1 e mais baixas para a Classe 4, mais uma vez revelando a importância da identificação das arestas mais relevantes para a compreensão do fenômeno de disseminação de informação. É interessante também notar uma diferença entre as características das redes da Classe 2, que preserva arestas salientes relacionadas ao volume e da Classe 3, que preserva arestas salientes relacionadas ao tempo. Esses resultados mostram que a dimensão volume apresenta relações muito mais díspares do que a dimensão tempo, fazendo com que as redes da Classe 2 sejam substancialmente menos densas que as redes da Classe 3.

A organização dos vértices em comunidades nas redes com diferentes classes de arestas, mais uma vez, reforça a importância da aplicação do *backbone* de forma bidimensional. Isso fica claro quando observamos os valores de modularidade da Classe 4 em comparação com as outras classes em ambas janelas, mas em especial na relacionada à data de 01/11/2022. Isso mostra que quando são filtradas as arestas ruidosas, provenientes de relações possivelmente esporádicas ou aleatórias, considerando, simultaneamente, as dimensões volume e tempo, uma estrutura de comunidades de usuários pode ser mais claramente observada.

Com o propósito de permitir uma visualização mais aprofundada da forma com que a informação é propagada nas redes com arestas

**Tabela 3: Topologia das redes do Telegram por classe.**

Classe	Data	$n$	$m$	$d$	$\hat{k}$	CMA	CC	# Com.	$Q$
Classe 1	1º Turno	1,406	54,467	0.055	77.478	0.449	118	137	0.126
Classe 2	1º Turno	53	35	0.025	1.321	0.056	19	20	0.883
Classe 3	1º Turno	1,208	26,446	0.036	43.785	0.419	2	7	0.460
Classe 4	1º Turno	120	288	0.040	4.800	0.205	9	15	0.379
Classe 1	2º Turno	3,150	107,218	0.022	68.075	0.488	218	237	0.159
Classe 2	2º Turno	93	78	0.018	1.677	0.063	21	24	0.844
Classe 3	2º Turno	2,141	53,743	0.023	50.204	0.443	3	6	0.462
Classe 4	2º Turno	137	344	0.037	5.022	0.208	5	11	0.414
Classe 1	Tumultos	4,763	70,313	0.006	29.525	0.528	136	158	0.351
Classe 2	Tumultos	254	427	0.013	3.362	0.173	23	33	0.549
Classe 3	Tumultos	263	303	0.009	2.304	0.092	19	26	0.789
Classe 4	Tumultos	5	3	0.300	1.200	0.000	2	2	0.444

de diferentes classes, as Figuras 3 apresentam as funções de distribuição cumulativas (CDFs) do número de *retweets* e do tempo médio de *retweets* de acordo com cada classe. O eixo- $x$  indica, em escala logarítmica, os valores da variável aleatória em questão (seja # *retweets* ou tempo médio), enquanto o eixo- $y$  indica a probabilidade acumulada dessa variável.

É possível observar nas Figuras 3(a) e 3(b), que mostram as curvas de distribuição do volume de *retweets* das arestas por classe, que as Classes 2 e 4, apresentam distribuições notadamente mais à direita no eixo- $x$ , indicando que as interações mais relevantes em relação ao volume, identificadas pelo método de *backbone*, são também aquelas que representam um maior número de *retweets*. Nota-se na Figura 3(a), por exemplo, que para a Classe 1 (Classe 3), aproximadamente 65% (80%) das arestas possuem peso 1. Já para as Classes 2 e 4, que capturam arestas com pesos representando um número maior de co-interações, os valores mínimos observados partem de 3. Uma análise análoga pode ser feita tomando como base a dimensão tempo, a partir das Figuras 3(c) e 3(d), que exibem as CDFs do tempo médio entre *retweets* em minutos para cada classe. Para melhorar o sentido intuitivo da análise, a normalização do tempo definida pela Equação 1 (Seção 3) foi desconsiderada neste experimento. As Classes 3 e 4 apresentam valores de tempo médio entre *retweets* consideravelmente menores, evidenciando que essas arestas mais relevantes em relação à dimensão tempo, capturadas nestas classes, representam disseminações mais rápidas, sejam em baixo ou alto volume. Por exemplo, na Figura 3(d), aproximadamente 90% das arestas das Classes 3 e 4 foram geradas a partir de um tempo médio de co-interação, isto é, *co-retweets*, de aproximadamente 1 minuto<sup>3</sup>, indicando uma probabilidade maior de arestas com disseminação muito mais rápida do que as Classes 1 e 2, em que este número representa cerca de 52%.

**5.2.2 Telegram.** As características topológicas de rede e da estrutura de comunidades para cada classe das redes no Telegram são apresentadas na Tabela 3. Diferentemente das redes observadas para o estudo de caso do Twitter/X, nas redes do Telegram observa-se além do desequilíbrio no número de arestas, observa-se também um grande desequilíbrio no número de vértices entre as classes. A Figura 4, que ilustra graficamente a distribuição dos vértices e arestas entre classes, nos ajuda a compreender melhor as características topológicas das redes. Nas redes do Telegram, a Classe

<sup>3</sup>A função Ceiling (Teto) foi usada para gerar uma melhor visualização e evitar tempo médio igual a zero.

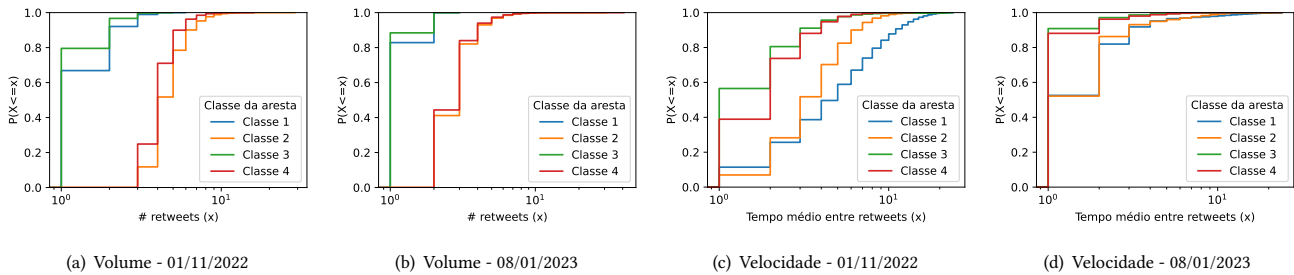


Figura 3: CDF *retweets* e tempo médio entre *retweets* para cada classe nas janelas de tempo 01/11/2022 e 08/01/2023.

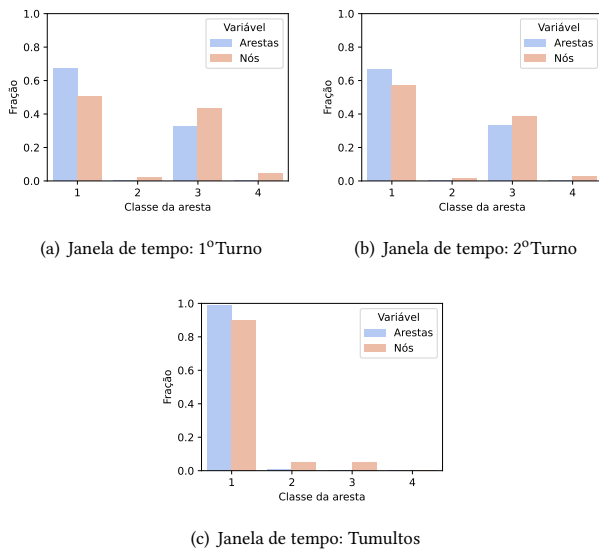
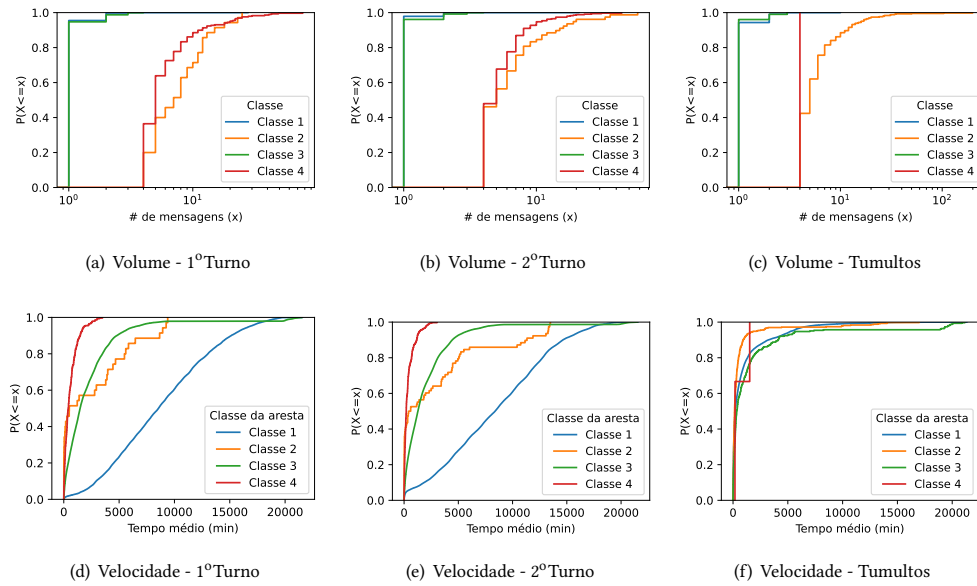


Figura 4: Fração de nós e arestas por classe no Telegram.

1, que corresponde às arestas que não são filtradas pelo método de *backbone* em nenhuma das dimensões, concentra um número consideravelmente maior de nós e arestas, assim como no estudo de caso do Twitter/X. Porém, há outras características topológicas que não se repetem neste estudo de caso. As Classe 2 e 4, que representam relações salientes em relação à dimensão volume, apresentam um número notadamente baixo de arestas e também de vértices, revelando alguns aspectos interessantes sobre o fenômeno de disseminação de informação nessa rede social. Em primeiro lugar, há um número baixo de arestas que se destacam por representarem o compartilhamento de um grande volume do mesmo conteúdo, mostrando que esse tipo de comportamento é bastante concentrado em poucos pares de usuários. Esse fato está fortemente relacionado não só ao fenômeno de disseminação de informação, mas à forma como a rede foi modelada no Telegram, que define como uma aresta o compartilhamento de mensagens com conteúdo textual idêntico, o que é resultado de uma ação menos intuitiva do que um *retweet* no Twitter/X, por exemplo. Além disso, o resultado para as Classe 2 e 4 mostra também que os vértices que estão associados a essas classes concentram relações apenas com essa característica e, nesse sentido, são menos diversos do que aqueles do Twitter/X. Dessa forma, as redes dessas classes apresentam, além do baixo número de vértices e arestas, uma alta fragmentação, o que pode ser observado pelo alto número de componentes conexas observado na Tabela 3.

A fragmentação observada nas redes das diferentes classes no Telegram tem efeito direto na estrutura de comunidades identificada. Em muitos casos observa-se valores de modularidade superiores a 0,8, como na Classe 2 nas janelas de 1º Turno e 2º Turno. Entretanto, deve-se considerar esse resultado com cautela dentro da análise do fenômeno de disseminação de informação, já que a clara definição da estrutura de comunidade pode ser decorrente da divisão dos vértices em componentes distintos. Assim, mesmo que uma informação seja propagada dentro das comunidades, cada componente potencialmente envolverá um baixo número de vértices, dificultando o espalhamento de uma mensagem, o que é reforçado pela barreira imposta pelas diferentes componentes. Isso mostra que considerar arestas no *backbone* identificado apenas na dimensão volume pode não ser uma estratégia adequada para a compreensão do fenômeno de espalhamento de informação no Telegram. Por outro lado, as redes da Classe 3, que consideram arestas identificadas pelo *backbone* na dimensão tempo, parecem mostrar comunidades melhor organizadas, mesmo com valores de modularidade menores. Essas redes são notadamente menos fragmentadas, o que pode ser visto pelo seu número de componentes conectadas, e podem ser estruturadas em comunidades que permitem uma melhor investigação sobre a disseminação de informação. Quando comparamos os resultados observados para o Telegram com aqueles observados para o Twitter/X, temos uma noção da complexidade do fenômeno de disseminação de informação em suas diferentes dimensões e particularidades quanto à plataforma em que acontece. Isso, mais uma vez, reforça a ideia central argumentada neste trabalho sobre a importância do estudo de relações relevantes em redes sociais, considerando não apenas uma, mas mais dimensões simultaneamente.

Para permitir a investigação sobre a forma como ocorre a propagação nas diferentes classes de arestas, a Figura 3 apresenta as distribuições CDF do número de mensagens e o tempo médio de compartilhamento, em minutos, por classe e por janela de tempo. Analisando o volume de informações por meio da distribuição do número de mensagens (Figuras 5(a) e 5(b)) é possível observar que as Classes 2 e 4 apresentam valores muito maiores indicando arestas com alto volume de disseminação, mostrando que as arestas mais relevantes, presentes no *backbone* tendem a também representar relações com alta co-interação. Por outro lado, as arestas das classes Classes 1 e 3 revelam um volume de disseminação bastante baixo, em que em todos os cenários aproximadamente pelo menos 95% das arestas possuem peso igual a 1. Isso revela a forte presença de co-interações esporádicas, porém rápidas. Em termos da velocidade com que a disseminação ocorre na plataforma,



**Figura 5: Função de distribuição acumulada do número de mensagens e tempo médio para cada classe nas redes do Telegram.**

as CDFs das Figuras 5(d), 5(e) e 5(f) mostram padrões de disseminação diferentes daqueles observados no Twitter/X. As arestas das Classes 3 e 4 apresentam, novamente, distribuições de tempo médio de compartilhamento de mensagens menores do que as demais, principalmente nas Figuras 5(d) e 5(e).

Entretanto, alguns valores da Classe 2 chamam a atenção, apresentando tempos médios tão pequenos quanto os das Classes 3 e 4. Relembre que o método de extração de *backbone Poly Urn* (Seção 3) é baseado em um modelo que assume que as arestas salientes de um nó são aquelas cujos pesos destoam significativamente das demais em relação às arestas incidentes ao próprio nó, isto é, uma análise local. Este resultado mostra que, na rede de disseminação do Telegram, algumas arestas da Classe 2 possuem pesos tão baixos quanto os das Classes 3 e 4, mas ainda assim, não são capturadas pelo *backbone* e, portanto, não são classificadas como esses perfis de disseminação. Isso ocorre devido ao fato de essas arestas pertencerem a pares de nós que possuem, em seus conjuntos exclusivos de arestas, distribuições de pesos de tempo médio de compartilhamento majoritariamente baixos em relação à mesma distribuição de toda a rede. Nestes casos, os valores ainda mais baixos dessas distribuições de pesos das arestas fazem com que algumas arestas sejam classificadas como sendo das Classes 3 ou 4, enquanto os demais, ligeiramente maiores em relação ao nó, mas bem menores para o padrão geral da rede, sejam classificados como sendo da Classe 2. Em resumo, esses pares de nós possuem conjuntos de arestas que apresentam grande potencial de disseminação em termos de velocidade se comparados aos demais nós na rede. Mesmo as arestas que não são capturadas pelas Classes 3 e 4 ainda têm tempos de disseminação menores comparados às demais. Dessa forma, nossos resultados revelam a existência de nós na rede com valores de disseminação predominantemente mais rápidos do que aqueles observados na rede como um todo. Por fim, ressalta-se que, durante os tumultos, a distribuição dos pesos baseados em

tempo médio (Figura 5(f)) apresenta, para todas as classes, tempos ainda menores de forma geral.

## 6 CONCLUSÕES E TRABALHOS E FUTUROS

Este estudo abordou a identificação e análise de padrões de disseminação de informações em redes sociais propondo um *framework* que combina volume e velocidade das interações para caracterizar diferentes perfis de disseminação de informações. Utilizando dois estudos de caso - Twitter/X e Telegram - analisamos períodos com evidências de coordenação e características topológicas distintas. Nossos principais resultados mostraram que o *framework* proporcionou uma visão mais detalhada dos padrões de coordenação entre contas de usuários. Identificamos diferentes classes de arestas com base no volume e na velocidade de disseminação, revelando padrões distintos de coordenação e comportamentos sincronizados nas plataformas analisadas. Mostramos também que, além do alto volume, arestas com outros perfis de disseminação formaram comunidades estruturadas, representativas de estratégias particulares de disseminação e coordenação.

Como trabalhos futuros, propomos estender a análise para outros cenários e eventos, focando na análise de conteúdo para entender como diferentes tipos favorecem cada perfil de disseminação. Investigaremos como narrativas, temas e formatos de mensagens contribuem para a velocidade e o volume de disseminação em cada classe de arestas.

## AGRADECIMENTOS

Este estudo recebeu suporte financeiro da Universidade Federal de Ouro Preto por meio do Programa de Bolsas de Iniciação Científica e Tecnológica, e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).



## REFERENCES

- [1] Marcelo MR Araujo, Carlos HG Ferreira, Julio CS Reis, Ana PC Silva, and Jussara M Almeida. 2023. Identificação e Caracterização de Campanhas de Propagandas Eleitorais Antecipadas Brasileiras no Twitter. In *Anais do XII Brazilian Workshop on Social Network Analysis and Mining*. SBC, 67–78.
- [2] Albert-László Barabási. 2013. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371, 1987 (2013), 20120375.
- [3] Carlos Barbosa, Lucas Félix, Vinícius Vieira, and Carolina Xavier. 2019. Sara - A Semi-Automatic Framework for Social Network Analysis. In *Anais Estendidos do XXV Simpósio Brasileiro de Sistemas Multimídia e Web*.
- [4] Daniele Bellutta and Kathleen M Carley. 2023. Investigating coordinated account creation using burst detection and network analysis. *Journal of big Data* 10, 1 (2023), 20.
- [5] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefevre. 2008. Fast unfolding of comm. in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, 10 (oct 2008), P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- [6] Simone Castagna, Giulia Porrino, and Federico Borgonovo. 2023. The Italian pro-Russia digital ecosystem on Telegram. *Cybersecurity and Law* 10, 2 (2023), 299–317.
- [7] Matteo Cinelli, Stefano Cresci, Walter Quattrociocchi, Maurizio Tesconi, and Paola Zola. 2022. Coordinated inauthentic behavior and information spreading on twitter. *Decision Support Systems* 160 (2022), 113819.
- [8] Michele Coscia and Frank MH Neffke. 2017. Network backboning with noisy data. In *2017 IEEE 33rd international conference on data engineering (ICDE)*. IEEE, 425–436.
- [9] Jose Martins da Rosa, Renan Saldanha Linhares, Carlos Henrique Gomes Ferreira, Gabriel P. Nobre, Fabricio Murai, and Jussara M. Almeida. 2022. Uncovering Discussion Groups on Claims of Election Fraud from Twitter. In *Proc. of Social Informatics: 13th International Conference*. [https://doi.org/10.1007/978-3-031-19097-1\\_20](https://doi.org/10.1007/978-3-031-19097-1_20)
- [10] Arash Dargahi Nobari, Negar Reshadatmand, and Mahmood Neshati. 2017. Analysis of Telegram, An Instant Messaging Service. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17)*.
- [11] Tuğrulcan Elmas, Rebekah Overdorf, Ahmed Furkan Özkalay, and Karl Aberer. 2021. Ephemeral astroturfing attacks: The case of fake twitter trends. In *2021 IEEE European symposium on security and privacy (EuroS&P)*. IEEE, 403–422.
- [12] Carlos Ferreira, Fabricio Murai, Ana Silva, Jussara Almeida, Martino Trevisan, Luca Vassio, Marco Mellia, and Idilio Drago. 2021. On the Dynamics of Political Discussions on Instagram: A Network Perspective. *Online Social Networks and Media* (2021).
- [13] Carlos H.G. Ferreira, Fabricio Murai, Breno Matos, and Jussara M. Almeida. 2019. Modeling Dynamic Ideological Behavior in Political Networks. *Web Science Journal* 7 (2019).
- [14] Carlos Henrique Gomes Ferreira, Fabricio Murai, Ana Paula Couto da Silva, Jussara Marques de Almeida, Martino Trevisan, Luca Vassio, Idilio Drago, and Marco Mellia. 2020. Unveiling Community Dynamics on Instagram Political Network. In *12th ACM Conference on Web Science*.
- [15] Luis Guilherme G. Da Fonseca, Carlos Henrique Gomes Ferreira, and Julio Cesar Soares Dos Reis. 2024. The Role of News Source Certification in Shaping Tweet Content: Textual and Dissemination Patterns in Brazil's 2022 Elections. In *Proceedings of the 20th Brazilian Symposium on Information Systems*. 1–10.
- [16] Fabio Giglietto, Nicola Righetti, Luca Rossi, and Giada Marino. 2020. It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 Italian elections. *Information, Communication & Society* 23, 6 (2020), 867–891.
- [17] Carlos Henrique Gomes Ferreira, Fabricio Murai, Ana P. C. Silva, Martino Trevisan, Luca Vassio, Idilio Drago, Marco Mellia, and Jussara M. Almeida. 2022. On network backbone extraction for modeling online collective behavior. *PLOS ONE* 17, 9 (09 2022), 1–36. <https://doi.org/10.1371/journal.pone.0274218>
- [18] Timothy Graham, Sam Hames, and Elizabeth Alpert. 2024. The coordination network toolkit: a framework for detecting and analysing coordinated behaviour on social media. *Journal of Computational Social Science* (2024), 1–22.
- [19] Franziska B Keller, David Schoch, Sebastian Stier, and JungHwan Yang. 2020. Political astroturfing on twitter: How to coordinate a disinformation campaign. *Political communication* 37, 2 (2020), 256–280.
- [20] Candice Lanius, Ryan Weber, and William I MacKenzie Jr. 2021. Use of bot and content flags to limit the spread of misinformation among social networks: a behavior and attitude survey. *Social network analysis and mining* 11, 1 (2021), 32.
- [21] Renan S. Linhares, José M. Rosa, Carlos H. G. Ferreira, Fabricio Murai, Gabriel Nobre, and Jussara Almeida. 2022. Uncovering Coordinated Communities on Twitter During the 2020 U.S. Election. In *Proc. of ASONAM*.
- [22] Riccardo Marcaccioli and Giacomo Livan. 2019. A pólya urn approach to information filtering in complex networks. *Nature communications* 10, 1 (2019), 745.
- [23] Michele Mazza, Guglielmo Cola, and Maurizio Tesconi. 2022. Ready-to-(ab) use: From fake account trafficking to coordinated inauthentic behavior on Twitter. *Online Social Networks and Media* 31 (2022), 100224.
- [24] Zachary P Neal, Rachel Domagalski, and Bruce Sagan. 2021. Comparing alternatives to the fixed degree sequence model for extracting the backbone of bipartite projections. *Scientific reports* 11, 1 (2021), 23929.
- [25] Mark EJ Newman and Michelle Girvan. 2004. Finding and evaluating community structure in networks. *Physical review E* 69 (2004), 026113.
- [26] Lynnette Hui Xian Ng, Ian Kloo, Samantha Clark, and Kathleen M Carley. 2024. An exploratory analysis of COVID bot vs human disinformation dissemination stemming from the Disinformation Dozen on Telegram. *Journal of Computational Social Science* (2024), 1–26.
- [27] Leonardo Nizzoli, Serena Tardelli, Marco Avvenuti, Stefano Cresci, and Maurizio Tesconi. 2021. Coordinated behavior on social media in 2019 UK general election. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 443–454.
- [28] Gabriel Nobre, Carlos Ferreira, and Jussara Almeida. 2020. Beyond Groups: Uncovering Dynamic Communities on the WhatsApp Network of Information Dissemination. In *SocInfo' 2020*.
- [29] Gabriel Peres Nobre, Carlos H.G. Ferreira, and Jussara M. Almeida. 2022. A Hierarchical Network-Oriented Analysis of User Participation in Misinformation Spread on WhatsApp. *Information Processing and Management* 59, 1 (jan 2022), 21 pages. <https://doi.org/10.1016/j.ipm.2021.102757>
- [30] Diogo Pacheco, Alessandro Flammini, and Filippo Menczer. 2020. Unveiling Coordinated Groups Behind White Helmets Disinformation. In *The Web Conference*.
- [31] Diogo Pacheco, Pik-Mai Hui, Christopher Torres-Lugo, Bao Tran Truong, Alessandro Flammini, and Filippo Menczer. 2021. Uncovering Coordinated Networks on Social Media: Methods and Case Studies. In *International Conference on Web and Social Media*.
- [32] Janina Susanne Pohl, Simon Markmann, Dennis Assenmacher, and Christian Grimme. 2023. Invasion@ Ukraine: providing and describing a twitter streaming dataset that captures the outbreak of war between Russia and Ukraine in 2022. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 17. 1093–1101.
- [33] Naman Saxena, Adwitiya Sinha, Tanishk Bansal, and Ankita Wadhwa. 2023. A statistical approach for reducing misinformation propagation on twitter social media. *Information Processing & Management* 60, 4 (2023), 103360.
- [34] M Ángeles Serrano, Marián Boguná, and Alessandro Vespignani. 2009. Extracting the multiscale backbone of complex weighted networks. *Proceedings of the national academy of sciences* 106, 16 (2009), 6483–6488.
- [35] Karishma Sharma, Yizhou Zhang, Emilio Ferrara, and Yan Liu. 2021. Identifying coordinated accounts on social media through hidden influence and group behaviours. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 1441–1451.
- [36] Stavros Souravlas, Angelo Sifaleras, Maria Tsintogianni, and Stefanos Katsavounis. 2021. A classification of community detection methods in social networks: a survey. *International Journal of General Systems* 50, 1 (2021), 63–91.
- [37] Serena Tardelli, Leonardo Nizzoli, Maurizio Tesconi, Mauro Conti, Preslav Nakov, Giovanni Da San Martino, and Stefano Cresci. 2024. Temporal dynamics of coordinated online behavior: Stability, archetypes, and influence. *Proceedings of the National Academy of Sciences* 121, 20 (2024), e2307038121.
- [38] MS Ulizko, AA Artamonov, RR Tukumbetova, EV Antonov, and MI Vasilev. 2022. Critical Paths of Information Dissemination in Networks. *Scientific Visualization* 14, 2 (2022), 98–107.
- [39] Aleksandra Urman and Stefan Katz. 2020. What they do in the shadows: examining the far-right networks on Telegram. *Information, Communication & Society* (2020), 1–20.
- [40] Luis Vargas, Patrick Emami, and Patrick Traynor. 2020. On the Detection of Disinformation Campaign Activity with Network Analysis. In *ACM Cloud Computing Security Workshop*. 133–146.
- [41] Deepika Varshney and Dinesh Kumar Vishwakarma. 2021. A review on rumour prediction and veracity assessment in online social network. *Expert Systems with Applications* 168 (2021), 114208.
- [42] Otavio R Venâncio, Carlos HG Ferreira, Jussara M Almeida, and Ana Paula C da Silva. 2024. Unraveling User Coordination on Telegram: A Comprehensive Analysis of Political Mobilization during the 2022 Brazilian Presidential Election. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 18. 1545–1556.
- [43] Derek Weber and Frank Neumann. 2020. Who's in the gang? Revealing coordinating communities in social media. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 89–93.
- [44] Derek Weber and Frank Neumann. 2021. Amplifying influence through coordinated behaviour in social networks. *Social Network Analysis and Mining* 11, 1 (2021), 111.