

# Pipeline para monitoramento de discussões políticas no Twitter: estudo de caso com o evento político de 8 de janeiro de 2023

Luciano José Silva<sup>1</sup>, Renata Mendes de Araujo<sup>1,2,3</sup>, Ana Grasielle Dionisio Correa<sup>1</sup>

<sup>1</sup>Programa de Pós-Graduação em Informática – Universidade Presbiteriana Mackenzie  
Rua da Consolação, 930 – 01302-907 – São Paulo – SP – Brasil

<sup>2</sup>Programa de Pós-Graduação em Sistemas de Informação – EACH/USP  
São Paulo - SP - Brasil.

<sup>3</sup>Doutorado Profissional em Políticas Públicas - Escola Nacional de Administração  
Pública – Brasília, DF – Brasil

luciano.jose-silva@outlook.com, {renata.araujo, ana.correa}@mackenzie.br

**Abstract.** *This research aimed to understand the dynamics of social networks during the Brazilian presidential elections in 2022. The objective was to identify the main groups of profiles that expressed themselves on the network, their positions, interactions, and speech during this electoral period. To this end, a Python pipeline associated with open source tools was developed, from which it was possible to extract, process, analyze and visualize Twitter data during relevant political events in 2022 and 2023. A case study is presented, based on data collected during the events of 8 January 2023 in Brasília, validating the application of the pipeline. This study presents a view of the network formed over time, the clusters found, and the main gatekeepers.*

**Resumo.** *Essa pesquisa buscou compreender a dinâmica das redes sociais durante as eleições presidenciais brasileiras em 2022. O objetivo foi identificar os principais grupos de perfis que se manifestaram na rede, seus posicionamentos, interações e discurso durante este período eleitoral. Para isso, foi desenvolvido um pipeline em Python associado a ferramentas de código aberto, a partir do qual foi possível extrair, processar, analisar e visualizar dados do Twitter durante eventos políticos relevantes em 2022 e 2023. Um estudo de caso é apresentado, baseado nos dados coletados durante os eventos de 8 de janeiro de 2023 em Brasília, validando a aplicação do pipeline. Esse estudo apresenta uma visão da rede formada ao longo do tempo, os clusters encontrados e os principais gatekeepers.*

## 1. Introdução

No dia 01 de janeiro de 2023, marcou-se a posse do Presidente Luiz Inácio Lula da Silva, dando fim a um dos processos eleitorais mais polarizados da política brasileira desde o final da ditadura militar nos anos 80. Ao longo do ano de 2022, os dois principais candidatos (o ex-presidente Luiz Inácio Lula da Silva e o presidente em exercício Jair Messias Bolsonaro) travaram uma “batalha” pelos votos dos eleitores brasileiros, como previsto pelo cientista político Antônio Lavareda [Lavareda 2022].

Como ocorrido em eleições nos Estados Unidos, com destaque as eleições de Barak Obama em 2008 [Cogburn e Vasquez 2009] e 2012 [Hong e Nadler 2012] e de Donald Trump em 2016 [Enli 2017], as redes sociais tiveram um impacto relevante no

processo eleitoral. De acordo com Braga e Carlomagno (2018), no Brasil, apesar das eleições de 2006 e 2008 possuírem estudos acerca do uso dos meios digitais no processo eleitoral, foi durante as eleições de 2010 que as redes sociais passaram a ter maior relevância, a partir da flexibilização do uso da Internet pelos candidatos.

Esse cenário evidencia a necessidade cada vez maior de entender como as discussões políticas se manifestam nas redes sociais. Para isso, propomos um *pipeline* cujo propósito é extrair dados do *Twitter* a partir de um evento desejado pelo usuário e analisar o comportamento da rede em relação a este evento. O principal objetivo do *pipeline* é apoiar no processo de evidenciar a dinâmica de comportamento das redes sociais em relação a eventos politicamente relevantes: seus principais participantes, configuração de grupos de perfis em seu posicionamento político e assuntos debatidos nas postagens. Como exemplo de uso do *pipeline*, apresentamos um estudo de caso baseado nos atos de 8 de janeiro de 2023, ocorrido em Brasília, desencadeado pelo resultado das eleições de 2022, e como eles se refletiram nas interações dos usuários do *Twitter*.

O artigo segue a seguinte estrutura: na Seção 2 são apresentadas os principais conceitos e métricas de redes sociais utilizados no trabalho; a Seção 3 discute trabalhos relacionados; a Seção 4 apresenta o *pipeline* proposto, e um exemplo de sua execução é ilustrado no estudo de caso descrito na Seção 5. A Seção 6 conclui o artigo e apresenta possibilidades de trabalhos futuros.

## 2. Conceitos Fundamentais

Para entender a dinâmica das redes sociais, como os diferentes perfis se conectam e se agrupam, este estudo se baseou em alguns conceitos e métricas de análise de redes sociais, apresentados a seguir.

**Gatekeeping:** O conceito de gatekeeping na comunicação refere-se à filtragem da enorme quantidade de informações disponíveis para torná-las acessíveis a uma pessoa em particular. Ele também é descrito como um processo mais amplo de construção da realidade social, através da decisão do que é ou não transmitido pela mídia [DeJuliis, 2015].

**Medidas de Centralidade:** Um dos principais conceitos de análise de rede sociais é o de métricas típicas dos nós da rede, e se referem principalmente à sua centralidade [Recuero, Bastos e Zago 2020] ou seja, ao quão central é um nó em uma rede. Para esse estudo, foram usados quatro tipos de centralidade para evidenciar os principais influenciadores: (i) **Autovetor (*Eigenvector*)** [Bonacich 1987]: afere a influência de um nó na rede e serve de base para outras centralidades. A medida do autovetor serve para determinar a importância de um nó em função da importância de seus vizinhos. Isto significa que, mesmo que um nó se conecte a poucos nós na rede, seus vizinhos podem ser importantes e, conseqüentemente, o nó também é importante. (ii) **Pagerank:** Introduzido em 1998, essa medida de centralidade mede o grau de influência de um nó em uma rede e se baseia no algoritmo que foi usado como base para ordenar buscas do Google [Page et al. 1999]. (iii) **HITS (*Hyperlink Induced Topic Search*)** [Kleinberg 1998]: assim como **Pagerank** e **Autovetor**, essa medida de centralidade mede o grau de influência de um nó em uma rede, sendo que para isso ele faz uso dos graus de entrada e saída para classificar os nós como **Authority** (se ele for conectado por muitos nós de qualidade) e **Hub** (se ele se conectar a muitos nós de

qualidade). (iv) **Grau:** representa o número de conexões que um determinado nó possui, sendo que, em uma rede direcionada, pode ser representado por dois tipos principais: **Grau de Saída**, que representa a quantidade de conexões que um nó faz, e **Grau de Entrada**, que representa a quantidade de conexões que um nó recebe. Quanto maior o grau de entrada, maior a importância deste nó como influência na rede [Recuero et. al. 2019].

**Modularidade (*Clusters*):** É definida como uma métrica de análise de redes que identifica os grupos adjacentes à estrutura de rede, sendo calculada tendo como base a separação dos nós da rede [Recuero et. al. 2015]. Essas comunidades são definidas como agrupamentos de nós altamente interligados entre si e pouco conectadas com o restante da rede. Com isso, redes com alto grau de modularidade tendem a ter conexões densas entre os nós de cada comunidade e poucas conexões com nós de outras comunidades. Fatores como homofilia e câmara de eco tem impacto direto na formação dos *clusters*. Estudos apontam que interações políticas no *Twitter* apresentam fortes padrões polarizados de *retweets*, demonstrando um alto grau de *clustering* pela tendência ideológica dos usuários [Conover et. al. 2011].

Além das métricas descritas acima, este trabalho também utilizou diferentes *layouts* de visualização de redes: o **Circle Pack Layout**, baseado em um algoritmo desenvolvido por Mike Bostock [Bostock 2018], sendo utilizado no trabalho para criar uma visão sintetizada da rede; o **OpenOrd**, criado em 2010, baseado originalmente no algoritmo Fruchterman-Reingold [Fruchterman e Reingold 1991]; e o *layout* **Yifan Hu**, criado em 2005 [Hu 2005], combinados para criar uma visão detalhada da rede.

### 3. Trabalhos Relacionados

No âmbito da análise de redes sociais em contextos políticos, destaca-se o estudo de Stieglitz e Dang-Xuan [2013], que propõem uma estrutura metodológica para esse fim, usando diferentes plataformas (*Twitter*, *Facebook* e *Microblogs*). Essa metodologia é dividida em duas partes: a primeira voltada para busca de conteúdo politicamente relevante em redes sociais, e a segunda na sugestão de estratégias para análise dos dados, como, por exemplo, nuvem de palavras, para auxiliar no entendimento dos resultados encontrados. Em outro estudo, Stieglitz et al. [2018] apresentam os desafios encontrados em cada etapa da estrutura metodológica apresentada no trabalho anterior [Stieglitz e Dang-Xuan 2013], que estão ligados à natureza interdisciplinar dos temas pesquisados, a detecção correta dos tópicos e eventos, a arquitetura do software de coleta de e processamento das informações, a qualidade dos dados e a visualização das informações.

Outras pesquisas no mesmo campo incluem o estudo de Bastos et al. (2013) que exploraram a estrutura de *gatekeeping* do *Twitter* por meio da análise de *hashtags* políticas como “#FreeIran”, “#FreeVenezuela” e “#Jan25”. Conover et al. (2011), por sua vez, examinaram duas redes de comunicação política no *Twitter*, compostas por mais de 250.000 *tweets* das seis semanas que antecederam as eleições de meio de mandato do Congresso dos EUA em 2010, com o intuito de entender como as mídias sociais moldam a esfera pública em rede e facilitam a comunicação entre comunidades com diferentes orientações políticas.

No Brasil, as manifestações acontecidas em 2013, e seus reflexos nas discussões dentro do *Twitter*, puderam ser observadas em [França e Oliveira 2014] e [Theodoro et.

al. 2013]. Também Pinto, Theodoro e Oliveira [2017] mapearam o comportamento das *hashtags* ligadas aos diferentes debates acontecidos durante o pleito e quais foram os principais candidatos a se destacarem nas eleições brasileiras de 2014.

#### 4. Pipeline Proposto

Construído para atender pesquisadores interessados em entender e visualizar a dinâmica das interações das redes sociais durante a disputa eleitoral brasileira de 2022, o presente *pipeline*, inspirado no trabalho de Stieglitz et al. (2018), foi construído para facilitar a identificação de temas de interesse, permitindo a extração dos dados, processamento e visualização das informações. A rede social escolhida para este estudo foi o *Twitter*, por sua representatividade global e também no Brasil e por permitir (à época que esta pesquisa foi iniciada) acesso via APIs para extração de dados.

O pipeline apresentado na Figura 1 acrescenta uma camada no trabalho de Bastos et. al. (2013) com a visão da rede, a identificação dos principais *hubs* de replicação e suas redes afiliadas, bem como os tópicos em discussão dentro dessas redes. Quando se confronta as linhas do tempo desses eventos, é possível observar a ação dos *network gatekeepers*, atuando como direcionadores dos temas nos diferentes *clusters*. A proposta era chegar a uma visualização de como as redes se formam a partir de eventos políticos importantes, identificar seus principais participantes e grupos, além de categorizar as posições de voto e os termos-chave discutidos neste grupo.

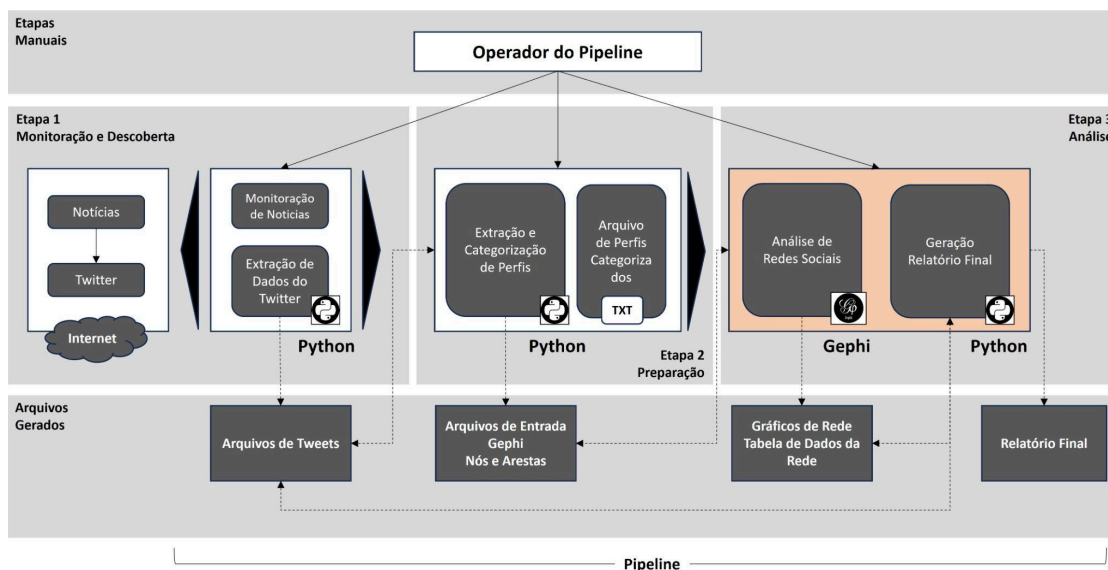


Figura 1 – Etapas do *pipeline*

O *pipeline* é composto pelas seguintes etapas:

**Etapa 1 - Descoberta e Monitoração de eventos:** A partir da seleção do evento de interesse da pesquisa, o usuário do pipeline identifica como esse evento é discutido dentro da *timeline* do *Twitter*, e seleciona as palavras-chave e/ou *hashtags* ligadas a ele. O próximo passo é a configuração da *query* de extração. Nessa etapa, o usuário informa a palavra chave para pesquisa, a quantidade máxima de *tweets* a serem extraídos, e a data de início da pesquisa. A execução da *query* gera arquivos com os *tweets* que atendam aos requisitos informados. Por exemplo, em 08 de janeiro de 2023, muitos sites de notícias destacaram as manifestações que ocorreriam em Brasília. Isso levou o termo

"Brasília" a ser uma tendência no Twitter desde as 6 da manhã daquele dia. Com base nesses eventos, uma palavra-chave que poderia ser usada para extrair *tweets* para a execução do *pipeline* seria "Brasília".

**Etapa 2 - Preparação:** Nessa fase, após dados coletados, é feita a criação dos arquivos que contêm as informações detalhadas sobre cada um dos perfis (nós) e o relacionamento entre eles (arestas). Essa etapa é importante porque ilustra as conexões entre os usuários envolvidos no debate eleitoral do *Twitter*. Nessa etapa também é realizada a categorização dos perfis do *Twitter* de acordo com seu posicionamento nas eleições. Para tanto, foi criado um arquivo contendo uma tabela com 3.850 contas abrangendo artistas, políticos, jornais, revistas, influenciadores digitais, extraídos a partir de listas disponibilizadas pelo *Twitter*. Para determinar o posicionamento político desses perfis, verificamos manualmente no *Twitter* se havia declarações explícitas de apoio a algum candidato em suas postagens. Dessa forma, os perfis foram classificados como "Pro\_Bolsonaro", "Pro\_Lula" ou "Neutro\Indefinido". Após a execução desses passos, são gerados dois arquivos de entrada para a ferramenta *Gephi*<sup>1</sup>, sendo o primeiro arquivo com os nós (autores dos *tweets* e perfis mencionados) identificados de acordo com seu posicionamento político e categoria (por exemplo, jornalista, políticos, etc) e o segundo o arquivo com as relações de menções entre os perfis (arestas).

**Etapa 3 - Análise de Dados:** Nessa fase é executado o cálculo das métricas de rede (análise quantitativa), de maneira a conhecer melhor a composição, densidade, centralidade e agrupamento dos elementos que compõem a rede. Em seguida, são gerados gráficos de visualização da rede, usando o *layout Circle Pack* para criar uma visão geral da rede (simplificada) e a combinação dos dois modelos de *layouts* (*Openord* e *Yifan Hu*) de maneira a criar uma visão detalhada da rede, ressaltando as principais modularidades, bem como os nós com maior grau de centralidade, identificando os *gatekeepers* da rede. Por meio de um *script*, os dados calculados e exportados pela ferramenta *Gephi* (em formato CSV) e os arquivos extraídos na primeira fase, são usados para se identificar os principais *gatekeepers* da rede (centralidade) e seus respectivos alinhamentos políticos, agrupamentos de interesse (*clusters*), tópicos que estavam sendo discutidos dentro desse grupo e sua convergência com os fatos identificados na rede.

## 5. Estudo de Caso

Para demonstrar os resultados obtidos a partir da execução do *pipeline*, será apresentado um estudo de caso focado em um tema relevante no contexto eleitoral brasileiro. O estudo de caso teve como objetivo demonstrar a viabilidade de execução das etapas previstas para o *pipeline* e a capacidade dos resultados obtidos com sua execução em apoiar o entendimento da dinâmica da rede social *Twitter* em relação a um evento específico de interesse do usuário do *pipeline*. As etapas de execução do *pipeline* são demonstradas, conforme apresentado a seguir.

**Descoberta e Monitoração de eventos:** Uma manifestação programada para o dia 8 de janeiro de 2023 em Brasília, por apoiadores do ex-presidente Jair Messias Bolsonaro, que contestavam os resultados das eleições presidenciais de 31 de outubro de 2022, saiu do controle e se tornou um ataque aos prédios do Congresso, Senado, Superior Tribunal de Justiça e monumentos da Esplanada dos Ministérios. O *pipeline*

---

<sup>1</sup> <https://gephi.org/>

foi acionado no dia 09 de janeiro de forma a coletar os dados referentes ao dia anterior. Para identificar o termo-chave para pesquisa, foi analisado o *Trend Topics* Brasil no *Twitter*, onde se destacava o termo "Brasília". Em seguida, foi configurada a *query* de extração usando esse termo. Realizou-se a extração das mensagens em português no *Twitter* ao longo de 24 horas (das 6:00 do dia 08 de janeiro às 6:00 do dia 09 de janeiro). A repercussão dos acontecimentos ocasionou um significativo volume de acesso aos servidores do *Twitter*, ocasionando tempo de resposta mais lento e impactando a capacidade de coleta de dados. Para contornar essa situação e viabilizar o estudo, a execução do pipeline teve que ser particionada em diferentes intervalos de tempo, totalizando sete intervalos (Figura 2).

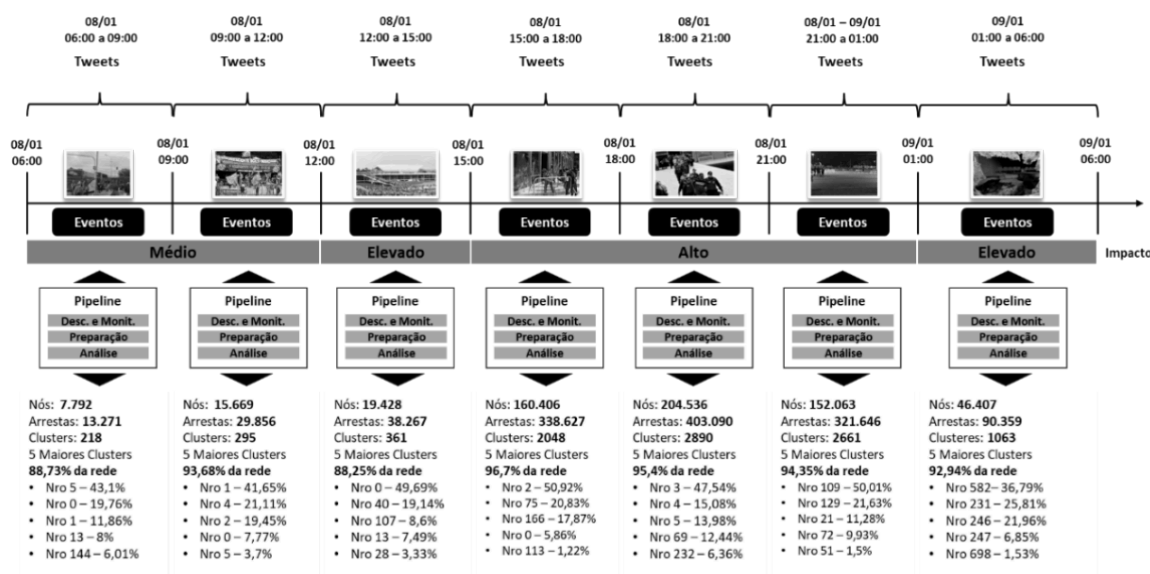


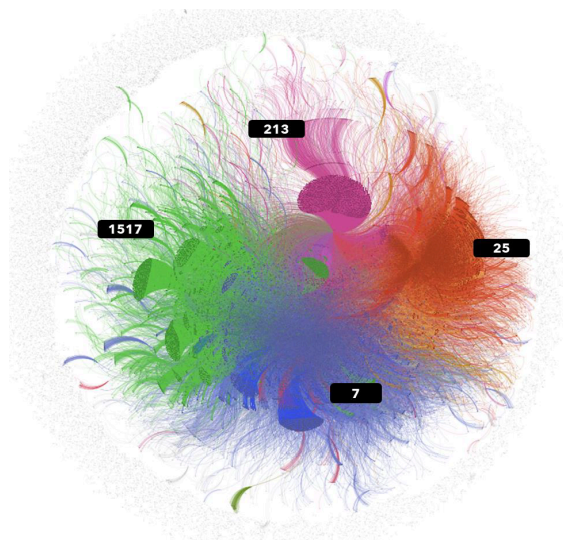
Figura 2 – Dados analisados por período

**Preparação:** Em cada período de coleta, foram criados arquivos contendo os *tweets* extraídos. Nessa etapa, os principais *gatekeepers* são classificados cruzando esses dados com a categorização de cada perfil e seu posicionamento político, gerando arquivos com essas informações. Na Tabela 1, há um exemplo desse arquivo categorizado, com IDs fictícios.

Tabela 1 – Exemplo do Arquivo Categorizado

Id	Categoria	Posição
00000001	Governadores do Brasil	Pro_Lula
00000002	Jornalismo brasileiro	Pro_Bolsonaro
00000003	Influenciador	Neutro/Indefinido
00000005	Políticos do Brasil	Pro_Lula

**Análise de Dados:** A rede formada para todo o período de vinte quatro horas compreendeu um total de 436.640 nós (perfis) e 1.164.952 arestas (*tweets*), 4.847 *clusters* e um grau médio de cada nó de 3,258. Os *clusters* calculados foram numerados (Figura 3) e os cinco principais *clusters* formavam 96,92% de toda a rede (*cluster* nro. 7: 41,58%, *cluster* nro 1517: 21,1%, *cluster* nro 25: 18,25% , *cluster* nro 213: 15,24% e *cluster* nro 433: 0,75%).



**Figura 3 – Principais *clusters* de toda a rede analisada**

Para cada período do dia, os dados foram analisados para a construção de visualização das redes, relacionando o comportamento da rede com os acontecimentos relatados pela mídia ao longo do dia.

Como exemplo, passamos a apresentar as análises para o período entre 09h e 12h do dia 08/01/2023. Neste período, alguns dos eventos relatados pela mídia [OGLOBO 2023] [PODER360 2023], incluíam:

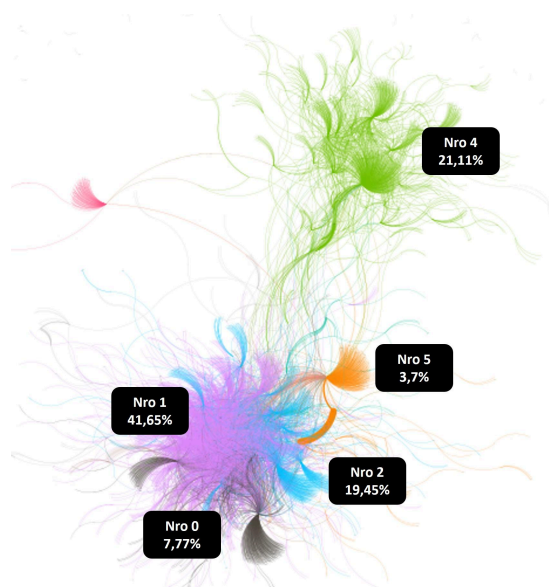
*7h36: Dino publicou no Twitter que esperava não haver atos violentos e que não fosse necessário a polícia atuar. O acampamento em frente ao QG do Exército contava com mais pessoas. Já se sabia, pela manhã, que os manifestantes planejavam caminhar até o Palácio do Planalto. Manifestantes convocavam para o ato em frente ao Congresso.*

*10:00: O dia começa com a preparação para a invasão do Congresso. Nas redes sociais, são compartilhados vários vídeos em que bolsonaristas radicais relatam supostas conversas com oficiais da Polícia Militar.*

*11:00: A Agência Brasileira de Inteligência (Abin) alertou sobre risco de invasão a prédios públicos horas antes dos atos de terrorismo. O órgão encaminhou pela manhã um relatório para as forças de segurança de Brasília, incluindo o governo do Distrito Federal, informando que os acampados do QG do Exército ameaçavam invadir a Praça dos Três Poderes.*

A análise dos dados neste período mostra uma rede com 15.669 nós (perfis), 29.856 arestas (*tweets*) e 295 *clusters*. Os 5 maiores *clusters* perfazem 93,68% da rede (Nro 1 – 41,65% - Pro\_Bolsonaro / Nro 4 – 21,11% - Pro\_Lula / Nro 2 – 19,45%

Pro\_Bolsonaro / Nro 0 – 7,77% - Pro\_Bolsonaro / Nro 5 – 3,7% - Pro\_Bolsonaro), conforme apresentado na Figura 4.



**Figura 4 – Mapa de rede com os principais clusters entre 9h e 12h de 08/01/23**

De um modo geral, os maiores *gatekeepers* identificados nesse período possuíam o posicionamento Pro\_bolsonaro (Tabela 2) e a nuvem de palavras desse período destaca além de “Brasília”, outros termos como: bolsonaristas, patriotas, força nacional, entre outros (Figura 5). Nas três centralidades apresentadas quanto mais próximos os valores estiverem de 1.0, mais relevantes os Gatekeepers serão dentro do Cluster.

**Tabela 2 – 10 maiores gatekeepers no período de 09h às 12h**

Gatekeeper	Ranking	eigencentrality	pageranks	Authority	Tweets Citados	Posicao
Gatekeeper_0	0	1.0	0.034426	0.788517	2734	Pro_Bolsonaro
Gatekeeper_1	1	0.404086	0.009283	0.224202	1024	Pro_Bolsonaro
Gatekeeper_2	2	0.390135	0.009828	0.22672	976	Pro_Bolsonaro
Gatekeeper_3	3	0.322372	0.009221	0.162346	648	Pro_Bolsonaro
Gatekeeper_4	4	0.243647	0.005032	0.152361	497	Pro_Bolsonaro
Gatekeeper_5	5	0.21425	0.004704	0.130471	475	Pro_Bolsonaro
Gatekeeper_6	6	0.187344	0.00431	0.086214	385	Pro_Bolsonaro
Gatekeeper_7	7	0.13702	0.003441	0.072082	361	Pro_Bolsonaro
Gatekeeper_8	8	0.151968	0.003819	0.081046	305	Pro_Bolsonaro
Gatekeeper_9	9	0.130045	0.003379	0.06636	265	Pro_Bolsonaro





Figura 5 – Nuvem de palavras no período de 09h às 12h

Ao extrair as nuvens de palavras dos 5 maiores clusters, notamos que os termos usados nos clusters pró-Bolsonaro têm similaridades, enquanto no cluster classificado como pró-Lula, os termos são distintos (Figura 6).



Figura 6 – Nuvem de Palavras nos 5 maiores cluster no período das 09h às 12h

As análises dos dados coletados mostraram como os eventos afetaram a dinâmica da rede, incluindo como o posicionamento político dos *gatekeepers* influenciou as discussões em suas "bolhas" e o crescimento da rede ao longo do dia. Também foi possível comparar diferentes momentos do dia, observando a formação dos principais grupos, suas orientações políticas e o volume de *tweets* trocados nos diferentes momentos.

## 6. Conclusão

Esse trabalho apresentou um *pipeline* que possibilita a extração de dados do *Twitter* ligados a eventos políticos de interesse, e o posterior tratamento, categorização e análise das informações processadas para a identificação de principais atores, grupos e seus perfis políticos.

A aplicação do *pipeline* permitiu que, ao longo de 2022 e início de 2023, fossem extraídos cerca de 28Gb de dados de 81 eventos políticos ocorridos no cenário nacional relacionados às eleições presidenciais, incluindo a divulgação de pesquisas eleitorais, debates para presidência da república e governadores estaduais, e acontecimentos de

impacto para campanha eleitoral. Em particular, foi possível estudar o comportamento da rede durante um episódio considerado como histórico, a invasão dos três poderes em 08 de janeiro de 2023.

As análises possibilitadas pelo *pipeline* geraram informações capazes de apoiar o entendimento da topologia da rede em relação aos principais atores, a formação de grupos e o teor das falas de cada grupo em diferentes momentos do dia 08 de janeiro. As informações geradas subsidiaram a construção de um artefato - Linha do Tempo <omitida> voltado a auxiliar cidadãos em geral a acompanhar o movimento das redes passo a passo ao longo do dia.

As principais limitações do *pipeline* incluem, por exemplo, a identificação e escolha manual de eventos para análise pelo pesquisador. O processo de extração de dados do *Twitter* depende diretamente do desempenho da API, que durante eventos que gerassem alto engajamento tornaram a captura de dados lenta e sujeita a *timeouts*. O *Twitter* possuía uma cota básica de 500 mil *tweets* por mês, mas para essa pesquisa ainda foi possível, à época, obter um acesso de pesquisador acadêmico (10 milhões de *tweets* por mês). Infelizmente, esse modelo de acesso para pesquisa foi descontinuado pela plataforma. O processo de classificação do posicionamento político dos perfis relevantes nas análises foi baseado na declaração de voto explícita de cada um deles em seus *tweets* publicados na plataforma. Essa condição pode ser afetada por diversos fatores como a mudança de posicionamento político, fatos relevantes ligados aos candidatos etc. Por fim, a polarização do cenário político brasileiro de 2022, fez com que os votos ficassem concentrados em apenas dois candidatos. Em cenários menos concentrados, o processo de classificação de posicionamento pode ficar mais disperso, impactando nos resultados das análises.

A pesquisa possibilitou a observação dos impactos de eventos políticos no *Twitter* durante um dos períodos mais polarizados da política brasileira nos últimos anos, proporcionando uma compreensão de como esses eventos influenciam a formação das redes. Os dados coletados possibilitarão futuras investigações, incluindo análises das redes sociais (como o impacto durante os debates presidenciais) e a criação de uma linha do tempo detalhada desses eventos.

Em trabalhos futuros exploraremos a correlação entre períodos avaliados, a associação entre notícias e interações nas redes sociais, a categorização política dos clusters gerais e a automatização de etapas do pipeline. Também serão explorados dados coletados em outras ocasiões. Já é possível visualizar uma dessas iniciativas no site do Ciberdem [Ciberdem 2023] onde está disponível a Linha de Tempo dos eventos de 8 de janeiro de 2023.

## **Agradecimentos**

Renata Araujo agradece o apoio financeiro do CNPq (#305645/2022-6) e da FAPESP (#2021/14772-1).

## **Referências**

Bastos, M. T., Raimundo, R. L. G. e Travitzki, R. (2013). “Gatekeeping Twitter: message diffusion in political hashtags”. *Media, Culture & Society*, 35(2), p. 260-270.

- Braga, S. e Carlomagno, M. (2018). “Eleições como de costume? Uma análise longitudinal das mudanças provocadas nas campanhas eleitorais brasileiras pelas tecnologias digitais (1998-2016)”. *Revista Brasileira de Ciência Política*, p. 07-62.
- Bonacich, P. (1987). “Power and centrality: A family of measures”. *American journal of Sociology*, 92(5), p. 1170-1182.
- Bostock, M. (2018). “Zoomable circle packing”. Github. <https://gist.github.com/mbostock/7607535>.
- Ciberdem. (2023, 9 de janeiro). “Linha do Tempo: Eleições Presidenciais de 2022 - Eventos de 08 de Janeiro”. Disponível em: <https://ciberdem.mack.com.br/index.php/linha-do-tempo-eleicoes-presidenciais-2022/brasil-janeiro-de-2023/>
- Cogburn, D. L. e Vasquez, F. K. (2009). “Examining the impact of web 2.0 and social media on political participation and civic engagement in the 2008 Obama campaign”. Em: *American Political Science Association Annual Meeting and Exhibition*. Toronto.
- Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F. e Flammini, A. (2011). “Political polarization on twitter” Em: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 5, No. 1. pp. 89-96.
- DeJuliis, D. (2015). “Gatekeeping theory from social fields to social networks”. *Communication Research Trends*. 34(1). p. 4-23.
- França, T. C., e Oliveira, J. (2014). "Análise de sentimento de tweets relacionados aos protestos que ocorreram no Brasil entre junho e agosto de 2013". Em: *Brazilian Workshop on Social Network Analysis and Mining*. Sociedade Brasileira de Computação. pp. 128-139.
- Enli, G. (2017). “Twitter as arena for the authentic outsider: exploring the social media campaigns of Trump and Clinton in the 2016 US presidential election”. *European Journal of Communication*. 32(1). p. 50-61.
- Fruchterman, T. M. e Reingold, E. M. (1991). “Graph drawing by force-directed placement”. *Software: Practice and Experience*. 21(11). p. 1129-1164.
- Hong, S. e Nadler, D. (2012). “Which candidates do the public discuss online in an election campaign?: The use of social media by 2012 presidential candidates and its impact on candidate salience”. *Government Information Quarterly*. 29(4), p. 455-461.
- Hu, Y. (2005). “Efficient, high-quality force-directed graph drawing”. *Mathematica Journal*. 10(1). p. 37-71.
- Kleinberg, J. M. (1999). “Authoritative sources in a hyperlinked environment”. *Journal of the Association for Computing Machinery*. 46(5). p. 604-632.
- Lavareda, A (2022, 2 de maio) [Entrevista ao Site Metro1]. Disponível em <https://www.metro1.com.br/noticias/politica/122677,eleicao-de-2006-foi-a-mais-polarizada-da-nova-republica-mas-2022-pode-superar-diz-lavareda>
- OGlobo. (2023, 9 de janeiro). “Cronologia do Terror: Como Golpistas Promoveram um Ataque Histórico à República”. Disponível em

<https://infograficos.oglobo.globo.com/politica/cronologia-golpistas-atacam-congresso-planalto-stf-brasil.html>

Page, L. (1998). "The pagerank citation ranking: Bringing order to the web". Technical report. Stanford Digital Library Technologies Project.

Pinto, P., Theodoro, I., e Oliveira, J. (2017). "Comportamento das hashtags durante grandes eventos". Em: *Anais do VII Workshop sobre Aspectos da Interação Humano-Computador para a Web Social*. Sociedade Brasileira de Computação. pp. 35-42.

Poder360. (2023, 9 de janeiro). "Leia a cronologia dos fatos que levaram às invasões". Disponível em: <https://www.poder360.com.br/brasil/leia-a-cronologia-dos-fatos-que-levaram-as-invasoes/>

Recuero, R., Bastos, M. e Zago, G. (2015). *Análise de redes para mídia social*. Editora Sulina.

Recuero, R., Zago, G. e Soares, F. (2019). "Using social network analysis and social capital to identify user roles on polarized political conversations on Twitter". *Social media and society*. 5(2).

Stieglitz, S., Mirbabaie, M., Ross, B. e Neuberger, C. (2018). "Social media analytics: Challenges in topic discovery, data collection, and data preparation". *International Journal of Information Management*. 39. p. 156-168.

Stieglitz, S. e Dang-Xuan, L. (2013). "Social media and political communication: a social media analytics framework". *Social Network Analysis and Mining*. (3). p. 1277-1291.

Theodoro, I., Stearns, B., Rangel, F., França, T. C., & Oliveira, J. Padrões de Interação no Twitter durante os Protestos de 2013.