The 2018 Brazilian presidential run-off: a complex network analysis approach using Twitter data

Juliano E. C. Cruz[1] and Marcos G. Quiles[2]

[1] Applied Computing Postgraduate Program, National Institute for Space Research (INPE), São José dos Campos, Brazil
[2] Institute of Science and Technology, Federal University of São Paulo (UNIFESP), São José dos Campos - SP, Brazil quiles@unifesp.br

Abstract. The political role of social networks has increased significantly in recent years. It has shifted from simple data analytics, employed to understand characteristics and desires of social groups, to a crucial digital platform for political campaigns and social influence. Here, we present an exploratory study of social data gathered from Twitter during the second round of the 2018 Brazilian Presidential Election. The study used complex network analysis to identify hidden communication patterns, important features, and key actors within Twitter data. The results were used to better understand how political polarization is embedded into Twitter during presidential campaigns. Moreover, the model also provided a straightforward manner to scrutinize the data and infer hypotheses from the obtained patterns. In summary, the number of communities of the network revealed an unbalanced division between the candidates and a lack of moderate behaviors.

Keywords: complex network · elections · social network analysis

1 Introduction

Social media has an important role in political discourse worldwide [4]. A considerable part of the global population uses Internet to read about general and political news. Many people who intend to have a more active role in politics, use social networks, blogs, or other online communities [10, 9]. A smaller group ends up organizing themselves in private communities, in which most members have homogeneous think alike and usually are aligned to the same party, political view, or ideological currents [10, 18]. Nowadays, one of the most popular social media for politicians is Twitter, which allows both direct and public communication with their supporters or ordinary citizens.

Brazil had become increasingly polarized starting with the 2014 presidential elections, oscillating during four years that followed and reaching a peak during the 2018 campaign. For the first time since the end of the military dictatorship, the far-right candidate won the 2018 election. After many years of centrist and left-wing tenures, the far-right coalition won by putting political "outsiders" on the ticket with a political agenda that appealed to the electorate.

Thus, this paper discusses Twitter data collected during the second round of the 2018 elections to verify the structure and patterns of communication and the shape and features of clusters and communities.

In this first section, background information on political polarization, the Brazilian political scenario up to the 2018 elections is presented. Related studies are analyzed in Section 2, followed by methodology. Next, centrality metrics and community detection algorithms employed in this paper are presented. The final section presents the resultant network topology, main nodes, clusters, and communities along with concluding remarks.

1.1 Political Polarization

Lee [16] defines polarization as the growth of extremists and the decline of moderates within an ideological spectrum. DiMaggio *et al.* [12] describe it as both a state and a process, indicating the opposition of opinions and its intensification over time. Polarization can be categorized into elite, mass, or pernicious types. Elite polarization occurs among political actors and institutions [8], while mass polarization involves the electorate's divided opinions and behaviors [1, 2]. Pernicious polarization combines elite and mass polarization into large opposing blocks, often around a binary division, and tends to last beyond the event that caused the polarization [8, 17].

The evolution of polarization is often fueled by the echo chamber effect, where beliefs are amplified within a closed system, leading to ideological segregation and political polarization [3, 22]. This effect is driven by selective exposure and confirmation bias, as individuals engage with information that aligns with their pre-existing beliefs, ignoring contrary viewpoints [3].

1.2 The Brazilian Political Scenario

The second round of the 2018 Brazilian presidential election had candidates from totally opposite political positions.

Jair Bolsonaro, a 63-year-old retired army officer and a seven-term federal congressman, represented the Social Liberal Party (PSL), advocating for a liberal economy and social conservatism. Known for his pro-gun stance and alignment with the United States and Israel, Bolsonaro's far-right views and "outsider" image underscored by his distance from national corruption scandals, made his campaign victorious.

Contrastingly, the Workers' Party (PT) nominated Fernando Haddad, a 55-year-old former São Paulo mayor and Minister of Education under PT administrations, after Luis Inácio Lula da Silva was barred from running due to legal convictions. Haddad's platform focused on expanding social welfare, opposing privatization, and enhancing Latin American and African ties.

Bolsonaro won with over 55% of the valid votes, securing victory in sixteen of the wealthiest states, while Haddad prevailed in eleven of the poorest. The election saw high rates of abstention (21.30%), blank (2.14%), and null (7.43%) votes, reflecting moderate voters' disillusionment with the polarized choice [13].

This polarization traced back to the 2014 elections and was exacerbated by corruption scandals involving PT, leading to widespread protests in 2013 and 2014. The impeachment of President Dilma Rousseff in 2016 and the subsequent presidency of Michel Temer marked a temporary dip in polarization. However, the announcement of the 2018 presidential candidates reignited polarization, reaching its zenith during the run-off.

2 Related work

Twitter was founded in 2006 and since then, it has been used for many purposes, like product advertisement, news, propaganda, and social networking for ordinary people. There are research articles that discuss the Twitter usage patterns.

Retweet is the main information dissemination mechanism on Twitter, but it was not known why certain information spread faster than others. Thus, Suh *et al.* [23] evaluated several features that may affect the propagation of tweets through retweets. The results showed that links and hashtags are strongly correlated to the retweet rate, as well as, followed user numbers, follower numbers, and account age.

There are also studies regarding exclusively political analysis on Twitter that can be cited. They analyze situations in Austria [15], Canada [14], Egypt [7], Germany [19], the United Kingdom [11], and the United States [10, 24].

Cram *et al.* [11] analyzes the data from one month before the 2017 British general election. It used over 34 million posts, where 9.6 million are original tweets and 25 million are retweets. The study employed time series analysis regarding relevant news, most cited topics, and most active and popular users. It detected the overwhelming dominance of pro-Labour posts and a disproportionate presence of the Scottish National Party, even though only Scottish voters could elect this party. The study found that, in this case, social media was just an extension of traditional media. It also claimed that, even though Twitter cannot be used to predict elections due to lack of representativeness, it was a useful tool to access the mood of a particular population niche.

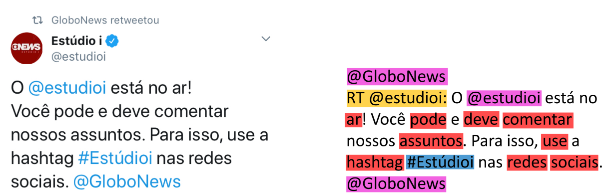
As for the 2017 German federal elections, Morstatter *et al.* [19] measured the election dynamics using Twitter data, from a dataset of more than 39 million tweets, with a little more than 130 thousand users. The study analyzed how the party Alternative for Germany (AfD), was able to take control of a large number of parliament seats. Initially, the study performed community detection to then identify how each cluster interacted with others and also to determine the most relevant themes for each cluster.

Yaqub *et al.* [24] reports on the 2016 American presidential election, where the main candidates were Hillary Clinton and Donald Trump. The dataset had almost 2.9 million tweets, which were acquired during the elections. One of the objectives was to evaluate how accurately tweets represented the public opinion. It found that sentiment and topics expressed on Twitter could be a good proxy for public opinion. Another finding was that little original content was created by users. They normally retweeted opinions and the user-to-user communication rate was quite low. Finally, the last relevant finding was that sentiments generated by Donald Trump discourses were more optimistic and positive than those generated Hillary's, directly reflecting the comment sentiments of Twitter users at the time.

3 Methodology

3.1 The Abstraction Model

The proposed abstraction model is able to keep all the information extracted from tweets by merging different nodes and relationship types into a unique network. In addition, it also allows node and relationship types to be filtered out in particular queries. The node types are: users, hashtags, retweets and words, called stems. The relationships are: copresence and authorship. Thus, in a scenario in which it is necessary to filter out only the users with a copresence relationship, it can be easily done with this approach. Figure 1 shows what a tweet looks like initially (a) and, in (b), the colored particles are the ones which are selected to become nodes in the preprocessing phase, as detailed in Section 3.2. Some words are not tagged due to stop word removal procedure.

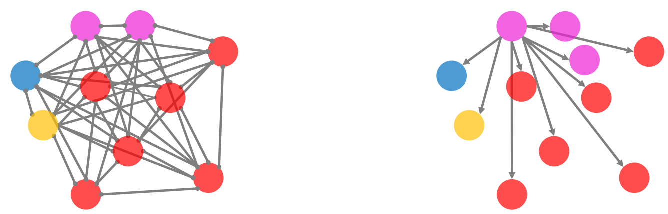


(a) (b)

Fig. 1: (a) Original tweet and (b) extracted text with color tag representing the node types, where pink highlights users, yellow-retweets, red-words, and blue-hashtags.

The example shown in Figure 1(a) is a retweet of a TV news account. It was chosen for having all of the node types in just one tweet. The translation is: "@estudioi is on air! You can and should comment our stories. Therefore, use the hashtag #Estúdioi on social networks. @GloboNews"

Networks would not exist with only nodes. Therefore, connections are also a really important feature. The copresence relationship in Figure 2(a) refers to the relationship created when the terms appear in the same tweet body. It is not directional because the terms are together and there is no action between them. The authorship relationship in Figure 2(b) refers to the act of a user writing a tweet. In this case, the relationship is directional and goes from the user to tweet body terms. Both edge types are weighted, where it means the number of times that particular relationship happened.



(a) Copresence relationships (b) Authorship relationships

Fig. 2: Node relationship types: (a) represents terms that appear in the same tweet and (b) represents a user that wrote a tweet with certain terms. This example refers to the Figure 1(b).

3.2 Dataset and Preprocessing

The data used in this study was acquired during the second round of the 2018 Brazilian presidential election, from October 8th to 27th. It has 1 million public tweets with user name and tweet body. For a faster and more dynamic analysis, an abstraction model, described in Section 3.1, was employed. Several processing approaches were employed to design the model some of which are related to Natural Language Processing (NLP).

The following steps were performed in the preprocessing phase: (A) relational database read; (B) language detection; (C) translation to Portuguese when applicable; (D) tweet segmentation in hashtags, user mentions, author users, retweets and words (regular text); (E) stop word removal (colloquial writing was not considered) and word stemming; (F) insertion in document-oriented database and data transformation; and (G) insertion in graph database.

There was a considerable amount of tweets in other languages. So, language detection was needed for each tweet, which was done in step B. About 91.7% (917,311) of the tweets were in Portuguese, 5.5% (55,169) in English, 1.7% (17,293) in Spanish and 0.4% (4,015) in French. In step C, non-Portuguese tweets were automatically translate to Portuguese to avoid language clusters and create a network that focused on meaning.

Step D performed the tweet segmentation, where tweet body words were separated into five classes or, what we call, tweet particles: hashtag, author user, user mention, retweet, and text. This segmentation procedure is very important for Step G, which needs to know what type of tweet particle it is handling, so it can than create the right type of node and relationship. Due to the fact that each term type has a well-defined pattern, simple regular expressions were used to perform such segmentation. Step E seeks to avoid words that do not have intrinsic meaning and to avoid meaning duplicity by performing stop word removal and stemming, respectively. In step F, the tweets were stored in the document-oriented database, where the document properties are the five tweet particles obtained in segmentation performed by step D. The objective of this task was to transform how data were organized. In step G, all nodes and relationships among them were inserted in graph-oriented database.

4 Methods

The methods employed for further network analysis are explained in this section.

4.1 Centrality Metrics

Centrality metrics measure how important a node is to the network. There are several approaches, but the ones used in this paper are approaches that consider the influence beyond the first connection layer. The methods are eigenvector centrality and PageRank.

Eigenvector centrality Eigenvector centrality was proposed in 1986 [6]. It is the first centrality metric to consider the transitive importance of a node in a graph rather than just its direct connections. Thus, relationships with high scoring nodes can be said to contribute more to the scoring of a particular node than connections to low scoring nodes.

PageRank PageRank was initially created to rank websites in Google search [20]. The score is based on the inbound link quantity and quality. It also relies on the assumption that an Internet user can get bored after several clicks, going then to a random page. It can be understood as a Markov chain, where states are pages and transitions are links, all of which have equally transition likelihood. Thus, if the method reaches a page with no outbound link, it will randomly choose a page to continue the process. PageRank pragmatically considers that pages without outbound links are connected to all pages in the network. Therefore, scores found for this particular page are divided equally among all other pages. This residual transition probability is typically set to 0.85. The value is estimated by averaging how often users use their bookmark list to go to a new page.

4.2 Community Detection

The objective of community detection methods is to find clusters in the network. Contrary to clustering methods that group samples in terms of their features, community detection only uses nodes and theirs relationships. This study employs label propagation and Louvain methods.

Label Propagation The Label Propagation algorithm is a fast algorithm for finding communities in graphs, which was proposed by Raghavan *et al.* [21]. It can detect communities using only the network structure but it has a feature allows to use initial labels to narrow down the final solution.

The algorithm assumes that a single label can easily become dominant in a densely connected group of nodes, which is unlikely to happen in a poorly connected region. At the end, nodes with the same label belong to the same community. The algorithm's name comes from the fact that some labels propagate through the network during the iterative process of label updating. Densely connected groups of nodes quickly reach consensus on a single label during the iterations. So, only a few labels will remain at the end.

The Louvain Algorithm Proposed by Blondel *et al.* [5] the Louvain algorithm performs hierarchical community detection. It is based on modularity and is one of the fastest algorithms, also performing well on very large graphs [5]. The basic idea consists in optimizing the communities' modularity and then, aggregating the community nodes. The modularity score quantifies the assignment quality of a community to a node by comparing how densely connected that community has become compared to a scenario in which it is a random network.

Louvain is a hierarchical method. Thus, it tries to go a level further every iteration, merging communities whenever possible. The overall stop criterion is met when an iteration does not result in any reassignments.

5 Results

5.1 Network Basic Features

The network under study comprises 468,643 nodes and 3,854,159 edges, featuring four distinct node types: 338,899 unique users, 66,831 unique stems, 40,790 unique retweets, and 22,123 unique hashtags. The edges are divided into 2,178,039 copresence relationships and 1,676,120 authorship relationships. Notably, 52% (176,079) of user nodes were active in writing tweets, while 48% (162,820) were mentioned without actively participating. A significant portion of nodes (32%, 149,095) lacked connections, often due to tweets containing only links or images without text.

Table 1 compares network segments by average degree, connection weight, and clustering coefficient. The network's average degree stood at 2842.0, with an edge weight of 3.65 per node, indicating the diversity of tweet particles and the frequency of node usage in tweets. The clustering coefficient, averaging over 0.16, reflects the degree of neighbor connectivity, with 1 indicating full connection and 0 none. The authorship-only network segment exhibited a markedly lower clustering coefficient (0.0007), suggesting less interconnectedness compared to the broader network. In contrast, retweets had a slightly higher average clustering coefficient (0.2269), hinting at more dense connections among these nodes. This

Table 1: Average degree, average weight per node, and average clustering coefficient for the entire network and for different node types.

Node Type	Population	Avg. Degree	Avg. Weight	Avg. Clustering Coef.
all	468,643	2842.0	3.65	0.1618
all (authorship)	236,563	2345.3	1.00	0.0007
all (copresence)	139,055	1746.9	6.09	0.0455
hashtags	22,123	6145.0	8.59	0.1303
stems	66,831	1823.4	3.94	0.1682
retweets	40,790	789.0	8.10	0.2269
users	338,899	259.1	1.55	0.1548

structural analysis reveals the network's complexity and the varying interaction patterns among different node types.

In analyzing the network by node type, hashtags lead with an average degree of 6145.0, followed by stems (1823.4), retweets (789.0), and users (259.1) due to their frequency and repetition in tweets. Hashtags, indexing keywords or topics, appear more repetitively (8.59 times on average) and are crucial for topic identification. Stems, representing morphological roots, are often neutral and used across political discussions, explaining their high repetition but lower average weight compared to hashtags and retweets. Retweets, reflecting endorsements of views, show high repetition due to the verbatim sharing of messages, enhancing their relationship weight.

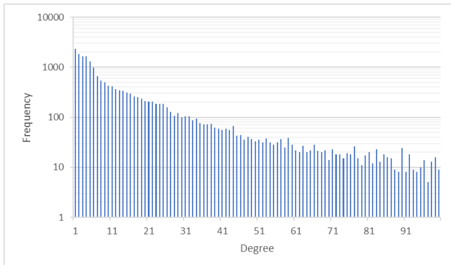
User nodes exhibit the least diversity (259.1) and repetition (1.55), impacted by the predominance of single-use authorship relationships, with over 72% of user nodes not repeating tweet particles. This uniqueness of user nodes stems from their mandatory role in authorship relationships, directly affecting their average weight and degree distribution.

Degree distribution, illustrated in Figure 3, reveals distinct patterns for each node type. Users and stems start with higher magnitudes (10000), while hashtags and retweets start at 1000. Notably, user nodes display a significant plateau, ending at degree 17, whereas retweets show a smaller plateau, concluding at degree 7, indicating varied engagement levels across node types.

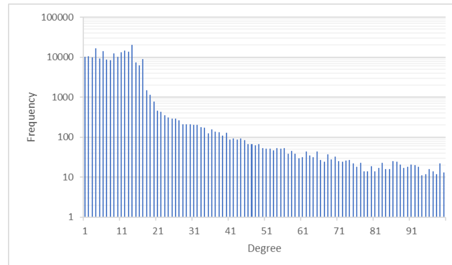
Figure 4 segments the network's degree distribution by connection type, revealing distinct patterns. The authorship relationship, depicted in Figure 4(a), mirrors the user node distribution from Figure 3(b), highlighting the significant role of user nodes in authorship connections. Conversely, the copresence relationship, shown in Figure 4(b), aligns more closely with the distribution patterns of hashtags, retweets, and stems, indicating their prevalent use in tweets. The combined network view in Figure 4(c) predominantly reflects the copresence relationship's distribution but also features a secondary peak from degrees 12 to 17, attributable to authorship connections.

The network's average weight per node is 3.65, with 87% (296,354) of nodes having an average weight between 1 and 2, as shown in Figure 5. This metric indicates the frequency of interactions between nodes, underscoring the network's dynamic nature.

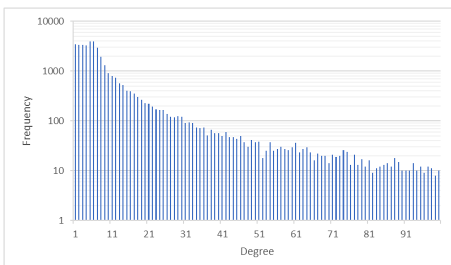
The highest degree nodes are predominantly hashtags, with five stems and one user also making the list. Degrees range from approximately 17,000 to over



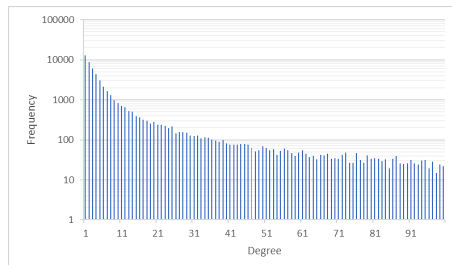
(a) hashtags



(b) users



(c) retweets



(d) stems

Fig. 3: Degree distribution for different node types.

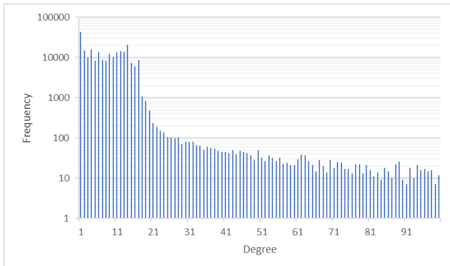
42,000, with most nodes having an average weight below 10, highlighting the significant role of hashtags in the network. Notably, three candidate-related hashtags (#haddad13, #bolsonaropresidente, and #bolsonarosim) and one candidate account (@jairbolsonaro) have higher average weights, reflecting their central role in the network's discourse.

Focusing on hashtags, all of which are associated with the two main candidates, indicating a lack of neutral hashtags among the top ranks. The distribution of average weights suggests varying degrees of engagement and repetition among these hashtags, with no direct correlation between degree and average weight. This pattern underscores the hashtags' importance in organizing and amplifying political discourse on Twitter.

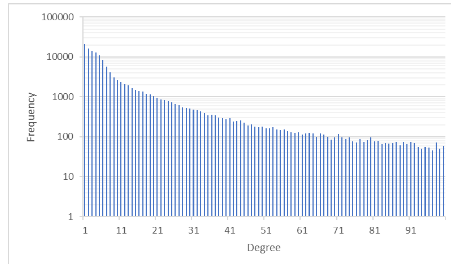
Stems, ranking second in average degree, exhibit lower average weights compared to hashtags. Their neutrality and widespread use across political discussions contribute to their high degree but lower repetition rate.

Retweets and user nodes, despite their lower ranking in degree, show significant engagement levels, with retweets particularly standing out for their higher average weights, indicative of their role in disseminating messages verbatim across the network.

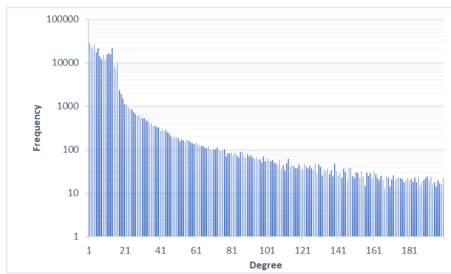
The cumulative distribution function (CDF) of the network and its subnetworks, presented in Figure 6, contrasts with example power-law and Poisson distributions. The overall network and authorship subnetwork distributions deviate



(a) only authorship relationship



(b) only copresence relationship



(c) network

Fig. 4: Degree distribution for each relationships and for the entire network.

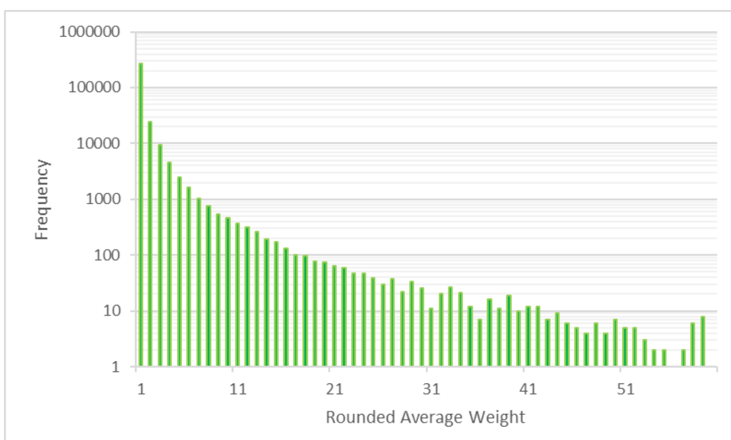


Fig. 5: The distribution of average connection weight per node.

from these models, suggesting a unique network structure not fully captured by traditional distribution models. However, the copresence subnetwork's alignment with a power-law distribution hints at the scale-free nature of these interactions.

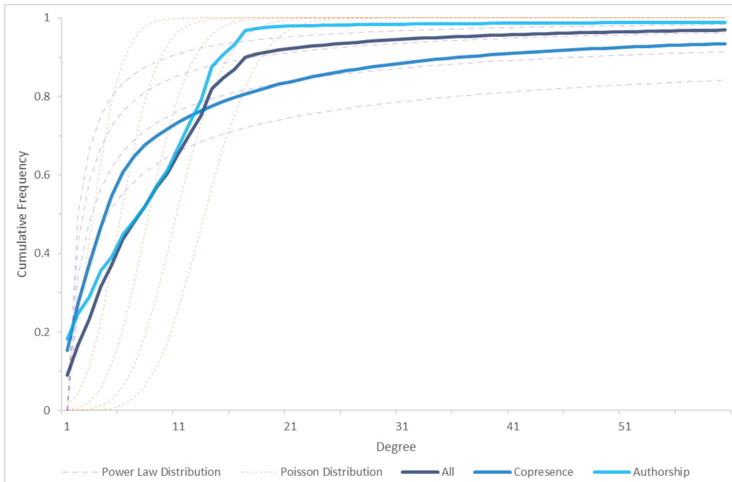


Fig. 6: Cumulative distribution functions of the network and its subnetworks (solid lines), example of power law distributions (dash-dot lines), and example of Poisson distributions (dashed lines).

This analysis elucidates the complex interplay of node types within the network, highlighting the pivotal role of hashtags in structuring political discourse and the distinct patterns of engagement across different types of connections. Detailed information about each node type or other analyses done can be found in Supplementary Material.

5.2 Authorship Relations

Analyzing authorship relationships between users and retweets reveals the network's main influencers. Popularity is gauged not just by the number of retweets or mentions but by the diversity of users engaging. This approach values breadth over depth of interaction, focusing on user-user and user-retweet dynamics.

Figure 7 identifies the twenty most cited users, with @jairbolsonaro and @haddad_fernando emerging as principal hubs within their respective clusters. Other notable mentions include @manueladavila, @rogerwaters, and @lulaoficial supporting Haddad, and @lobaoeletrico, @carlosbolsonaro supporting Bolsonaro, alongside neutral entities like @youtube and @tsejusbr. The peculiar case of @jairbolsonarohttps highlights issues in data parsing.

Distinct clusters around these hubs suggest a polarized network structure, with a central cluster engaging with both sides. The comparison between the top 15 users regarding degree and top 20 most mentioned usersshows a shift in ranking when focusing solely on inbound authorship connections, underscoring the different interaction patterns in authorship versus copresence relationships.

Retweet analysis regarding the authors with higher degreefurther illustrates the network's segmentation, with no cross-cluster retweeting among the top 20

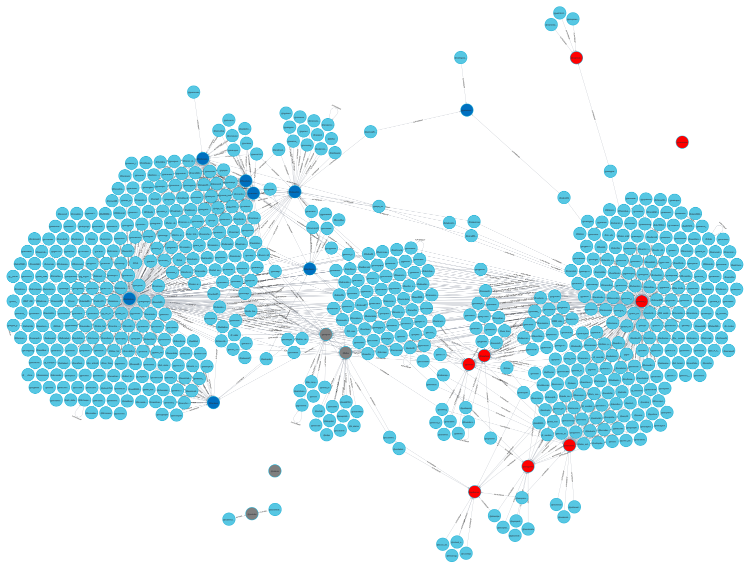


Fig. 7: The twenty most mentioned users and the main users who mentioned them (light blue). The color of mentioned users represent which side they were related to. Blue is for nodes supporting Bolsonaro, red Haddad and gray neutral.

retweets. This isolation reflects echo chamber dynamics, as discussed by Conover *et al.* [10]. The majority of retweets support Haddad, with a smaller fraction backing Bolsonaro or remaining neutral. Detailed information about the analyses done can be found in Supplementary Material.

5.3 Top 50

Centrality analysis using eigenvector and PageRank metrics reveals the network's key nodes. The eigenvector centrality, shown in Figure 8, highlights a network with 771 connections, including fifteen users, twelve hashtags, and twenty-three stems, but no retweet nodes. This mix of nodes indicates diverse political affiliations and topics of discussion, from election-related hashtags to neutral stems and media outlets.

PageRank centrality, detailed in Figure 9, presents a slightly different composition with nineteen hashtags, nineteen stems, nine retweets, and three users. This variation underscores the different focuses of the two centrality measures, with PageRank incorporating a broader array of retweets and hashtags into the top ranks.

Both centrality measures capture the network's critical nodes but from different perspectives. Eigenvector centrality emphasizes users and stems, while PageRank gives more weight to retweets and hashtags, reflecting their role in disseminating and amplifying content.

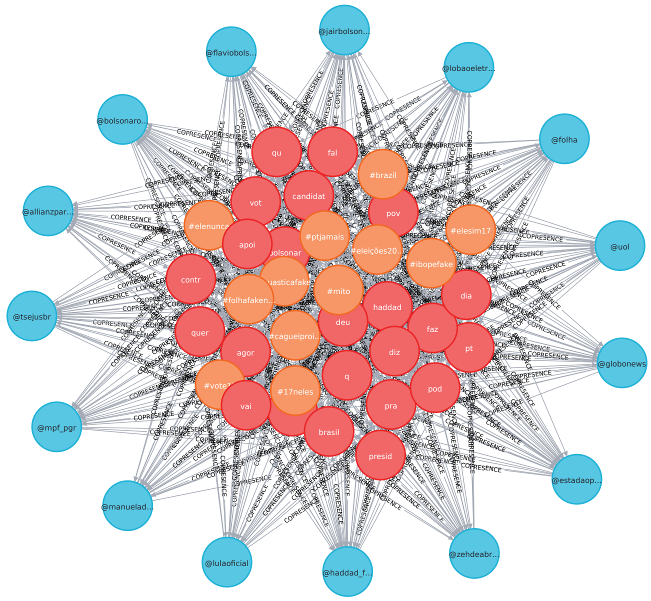


Fig. 8: Top 50 using eigenvector centrality.

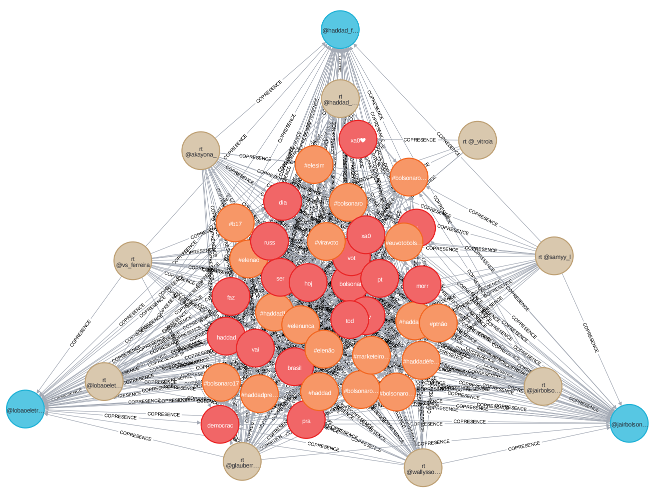


Fig. 9: Top 50 using PageRank.

5.4 Communities

Community detection using the Louvain algorithm identified 150,574 communities, with the 23 largest encompassing 67.7% of all nodes. These communities range from large, politically aligned clusters to smaller, topic-specific groups. The largest community had 43,415 nodes, while the smallest in the top 23 had

84 nodes. Communities with only one node accounted for over 32% of the total, highlighting the network's fragmentation.

Figure 10 shows the distribution of these communities by size and political affiliation, with Bolsonaro-related communities numbering nine and Haddad-related six. Neutral communities, often hosting debates between political sides, make up eight of the total. Notably, some communities focus less on political debate and more on specific themes or events, such as concerts or personal discussions.

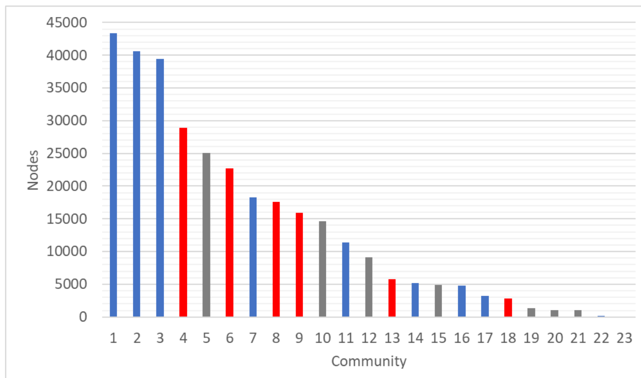


Fig. 10: Largest communities obtained with Louvain algorithm. Blue are Bolsonaro-related, red-Haddad-related, and gray-neutrals.

The internal organization of these communities, as depicted in Figure 11, reveals a predominance of user nodes, with significant variation in the presence of other node types like retweets, stems, and hashtags. This diversity indicates the range of discussion and interaction patterns within each community.

PageRank scores within communities, shown in Figure 12 and Figure 13, suggest a wide disparity in influence, with most nodes holding low scores. This distribution points to a few highly influential nodes within communities, surrounded by a large number of less influential ones.

PageRank was chosen because it generated a top 50 composed of all node types, which did not happen with eigenvector centrality, see Section 5.3. Having a more diverse node population, theoretically, helps in order to have a better inference about community theme. Thus, from visually analyzing the top 30 of every community it is possible to note that only two communities (19 and 11) do not have all node types present. Also, most of the top 30 have stems as the majority nodes, except for Community 1. Stems do not have the highest PageRank scores on average (Figure 14), but at the top 30 they are the majority in most communities. Another notable trend is that, the smaller the community is with the top 30, the more likely it is to be less connected.

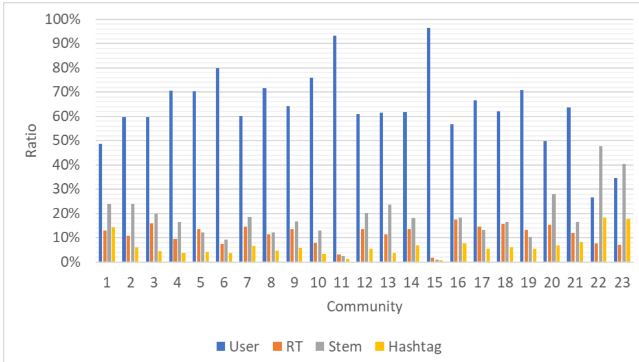


Fig. 11: Node-type ratio for each community.

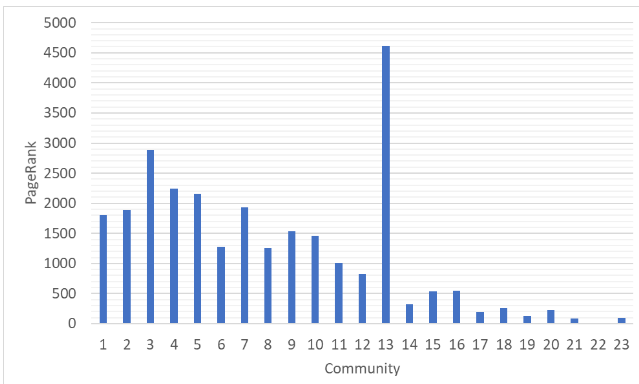


Fig. 12: Highest PageRank scores for each community.

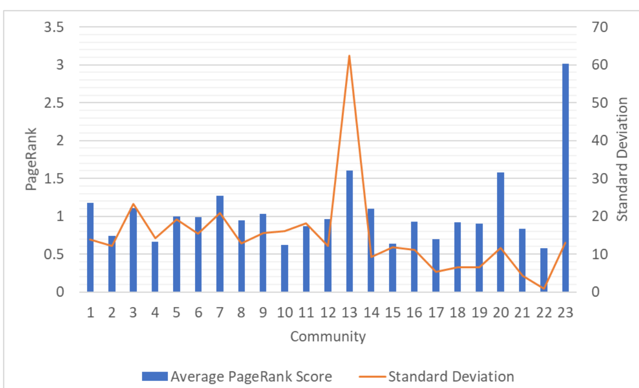


Fig. 13: PageRank score average and standard deviation for each community.

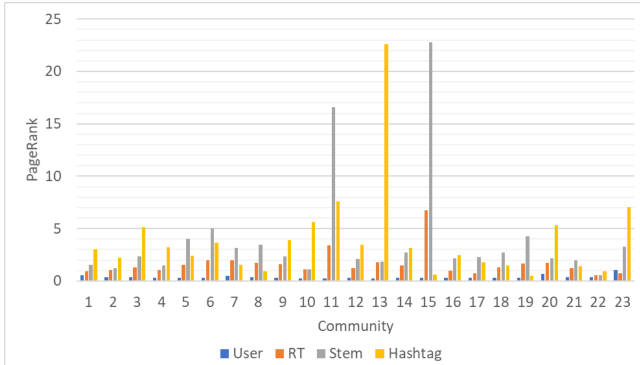


Fig. 14: PageRank average per node type in each community.

The analysis of community themes, based on the top 30 nodes by PageRank, reveals a rich tapestry of political affiliations, key issues, and public figures dominating the discourse. While political nodes are prevalent, the presence of neutral and thematic nodes underscores the network's complexity and the multifaceted nature of online political discourse.

In summary, the network's structure, from its influencers to its communities, paints a detailed picture of the online political landscape during the election period. The interplay of user interactions, centrality measures, and community dynamics offers insights into the digital public square's polarized yet interconnected nature. Detailed information about all the analyses done can be found in Supplementary Material.

6 Conclusions

The objective of this paper was to discuss Twitter data from the Brazilian 2018 presidential run-off election. The study employed tools and methods to facilitate storage, processing, and analysis of unstructured data.

The employed tweet abstraction model made the network analysis easier and much more dynamic. Just one network was necessary. Analysis was performed by queries that allowed selecting multiple types of nodes and connections. The copresence subnetwork seems to follow a power-law distribution, while the entire network and authorship subnetwork do not seem to follow either power-law or Poisson distribution. When the most mentioned users were filtered, the resultant hubs coincide with the most prominent personalities and organizations during the period of the political campaign. The main users are both presidential candidates and secondarily there are users orbiting them. The fact that there is just a medium-size group of users connected to both candidates may suggest that the majority of users might get focused exclusively on attacking or supporting one of the sides, with regard to user mention.

The claim that users who support opposite parties or politicians hardly interact through retweets [10] was confirmed by the analysis done. In addition, no user retweeted more than one node from the top 20.

The results from Label Propagation algorithm were not satisfactory. However, the Louvain algorithm produced good results. The community sizes have logarithmic-alike distribution, then only the main ones where analyzed, the 23 largest communities. Most of the top 30 nodes of the communities had stems as majority. Their main node was hashtag (9 communities), followed by stems (6 communities).

Frequently research creates more questions than answers. Further studies should examine the communities in greater depth, to study main user roles and influences, hashtag and stem themes. They also should probe the relationship among communities in order to understand the relationship structure and to examine whether neutral communities orbit polarized groups or form an independent major group.

Finally, polarization normally means expanding extremists and diminishing moderates. Moderate numbers diminish, but do not disappear completely. Therefore, except for Community 15, one could not find explicitly moderate users or moderate behavior in the data analyzed. It is impossible to say if moderates ignored the second round election completely or supported one of the sides; there is no data to establish by what ratio either happened. Nor can it be ascertained what role moderates had in communities, or even if they were present in the major ones.

References

1. Abramowitz, A.: The disappearing center: engaged citizens, polarization, and American democracy. Yale University Press (2011)
2. Abramowitz, A.A., Saunders, K.L.: Is polarization a myth? The Journal of Politics **70**(2), 542–555 (2018)
3. Barberá, P., Jost, J., Nagler, J., Tucker, J., Bonneau, R.: Tweeting from left to right: Is online political communication more than an echo chamber? Psychological Science **26**(10), 1531–1542 (2015)
4. Barberá, P., Gohdes, A.R., Iakhnis, E., Zeitzoff, T.: Distract and divert: How world leaders use social media during contentious politics. The International Journal of Press/Politics **29**(1), 47–73 (2024)
5. Blondel, V.D., Guillaume, J., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment **2008**(10), P10008 (2008)
6. Bonacich, P.: Power and centrality: a family of measures. American journal of sociology **92**(5), 1170–1182 (1987)
7. Borge-Holthoefer, J., Magdy, W., Darwish, K., Weber, I.: Content and network dynamics behind egyptian political polarization on twitter. In: Proceedings of ACM Conference on Computer Supported Cooperative Work & Social Computing. pp. 700–711 (2015)
8. Carothers, T., O'Donohue, A.: Democracies Divided: The Global Challenge of Political Polarization. Brookings Institution Press (2019)

9. Chan, M., Yi, J.: Social media use and political engagement in polarized times. examining the contextual roles of issue and affective polarization in developed democracies. Political Communication **0**(0), 1–20 (2024)
10. Conover, M.D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., Flammini, A.: Political polarization on twitter. In: AAAI Conference on Weblogs and Social Media (2011)
11. Cram, L., Llewellyn, C., Hill, R., Magdy, W.: Uk general election 2017: a twitter analysis. arXiv:1706.02271 (2017)
12. DiMaggio, P., Evans, J., Bryson, B.: Have american's social attitudes become more polarized? American Journal of Sociology **102**(3), 690–755 (1996)
13. Federal, S.T.: Brazilian electoral data repository. http://www.tse.jus.br/hotsites/pesquisas-eleitorais/index.html, accessed: 2020-10-01
14. Gruzd, A., Roy, J.: Investigating political polarization on twitter: A canadian perspective. Policy & Internet **6**(1), 28–45 (2014)
15. Kušen, E., Strembeck, M.: Politics, sentiments, and misinformation: An analysis of the twitter discussion on the 2016 austrian presidential elections. Online Social Networks and Media **5**, 37–50 (2018)
16. Lee, J.M.: The political consequences of elite and mass polarization. Ph.D. thesis, University of Iowa (2012)
17. McCoy, J., Somer, M.: Toward a theory of pernicious polarization and how it harms democracies: Comparative evidence and possible remedies. Annals of the American Academy of Political and Social Science **681**
18. Moe, H., Lindtner, S., Ytre-Arne, B.: Polarisation and echo chambers? making sense of the climate issue with social media in everyday life. Nordicom Review **44**(1), 23–43 (2023)
19. Morstatter, F., Shao, Y., Galstyan, A., Karunasekera, S.: From alt-right to alt-rechts: Twitter analysis of the 2017 german federal election. In: Proceedings of the The Web Conference. pp. 621–628 (2018)
20. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Stanford InfoLab (1999)
21. Raghavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. Physical Review E **76**(3), 036106 (2007)
22. Rum, S.N.M., Mohamed, R., Asfarian, A.: Identifying political polarization in social media: A literature review. Journal of Advanced Research in Applied Sciences and Engineering Technology **34**(1), 80–89 (2023)
23. Suh, B., Hong, L., Pirolli, P., Chi, E.H.: Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In: IEEE International Conference on Social Computing. pp. 177–184 (2010)
24. Yaqub, U., Chun, S.A., Atluri, V., Vaidya, J.: Analysis of political discourse on twitter in the context of the 2016 us presidential elections. Government Information Quarterly **34**(4), 613–626 (2017)