Journal of Physics: Complexity

RECEIVED
19 April 2023

REVISED
28 June 2023

ACCEPTED FOR PUBLICATION
4 September 2023

PUBLISHED
13 September 2023

PAPER

Group polarization, influence, and domination in online interaction networks: a case study of the 2022 Brazilian elections

Ruben Interian* and **Francisco A Rodrigues**

Institute of Mathematics and Computer Science, University of São Paulo, São Carlos, São Paulo, Brazil

* Author to whom any correspondence should be addressed.

E-mail: ruben@icmc.usp.br

Keywords: polarization, domination, interaction network, Twitter, modularity

Abstract

The erosion of social cohesion and polarization is one of the topmost societal risks. In this work, we investigated the evolution of polarization, influence, and domination in online interaction networks using a large Twitter dataset collected before and during the 2022 Brazilian elections. From a theoretical perspective, we develop a methodology called d-modularity that allows discovering the contribution of specific groups to network polarization using the well-known modularity measure. While the overall network modularity (somewhat unexpectedly) decreased, the proposed group-oriented approach reveals that the contribution of the right-leaning community to this modularity increased, remaining very high during the analyzed period. Our methodology is general enough to be used in any situation when the contribution of specific groups to overall network modularity and polarization is needed to investigate. Moreover, using the concept of partial domination, we are able to compare the reach of sets of influential profiles from different groups and their ability to accomplish coordinated communication inside their groups and across segments of the entire network. We show that in the whole network, the left-leaning high-influential information spreaders dominated, reaching a substantial fraction of users with fewer spreaders. However, when comparing domination inside the groups, the results are inverse. Right-leaning spreaders dominate their communities using few nodes, showing as the most capable of accomplishing coordinated communication. The results bring evidence of extreme isolation and the ease of accomplishing coordinated communication that characterized right-leaning communities during the 2022 Brazilian elections, which likely influenced the subsequent coup events in Brasilia.

1. Introduction

The Global Risks Report published by the World Economic Forum (WEF) is an annual study that explores some of the most severe risks may be faced over the next decade. The authors of the most recent report [20] state that the erosion of social cohesion and polarization ranked as the most significant societal risk among those pointed by WEF 2023. Moreover, societal polarization is listed as the fifth-most severe global risk in the short term (2 years) and the seventh-most significant one over the next 10 years. As an example that illustrates these asseverations, an analysis by the Pew Research Center discovered that, on average, Democrats and Republicans are farther apart ideologically today than at any time in the past 50 years in the USA [12].

Defined as the fracturing into sharply contrasting communities, polarization leads to declining social stability due to the insufficient communication between polarized, strongly-connected groups [22]. Extreme polarization, or radicalization, may lead to gridlocks or even violent conflicts [16]. Polarization may reduce the space for collective problem-solving, stimulating the adoption of short-term policy platforms to galvanize one side of the population [20]. There has been extensive research on the more general issue of competing interactions and the possibility of consensus formation [2, 35, 37].

Social networks and mass media are places where polarization manifests itself in a strong way [23]. They have attracted millions of individuals by allowing them to communicate, share their ideas, and discuss different topics, being an excellent tool for studying individual and group behavior. However, polarization and hostility are increasingly shifting from social media to the real world, as it was demonstrated by several political events, such as the protests of the Yellow Vest movement in France in 2018, the protests after George Floyd's death in 2020, the US Capitol attack in 2021, and the convoy protests in Canada in 2022 [21].

Previous studies analyzed partisan polarization in online participatory platforms, parliaments, and blogging networks in the USA [13, 17, 33], France [28], Italy [10], Israel [36], India and Pakistan [34], and a group of 16 European countries [26]. Markgraf and Schoch [27] reported a case study based on data from the German Federal Election of 2017 for illustrating their echo chambers research framework.

It is known that in the last decade, Brazilian society has been highly polarized. Cota *et al* [9] analyzed the effects of echo chambers in information spreading in a Twitter interaction network related to the impeachment of the former Brazilian President Dilma Rousseff. They showed that, on average, users with pro-impeachment leanings could transmit information to a larger audience than users expressing anti-impeachment inclinations. Online polarization in Brazil was also analyzed in [4], showing how news outlets influenced political discussions during the 2018 Brazilian presidential campaign.

In this study, we investigated the evolution of polarization, influence, and domination during the 2022 presidential election in Brazil. In 2022, president Lula da Silva won the Brazilian presidential election of 2022 by 1.8 points—the slimmest margin recorded since redemocratization, which happened between 1974 and 1985. To the best of our knowledge, this is the first work that analyzes this period. Among other implications, this research could help us to identify the reasons that may lead to the coup events that occurred on 8 January 2023 in Brasília, when protesters attempted to forcefully depose Brazil's democratically elected president by breaking into and vandalizing the Supreme Federal Court, the National Congress building, and the Planalto Presidential Mansion in the Three Powers Plaza. We are also interested in information spreading and domination before and during the elections, both in the interaction network among individuals and in each community individually.

In our investigation, we used Twitter social network data. Many public figures—politicians, personalities, and researchers—use Twitter frequently to communicate and expose their viewpoints, often using this social network as their official public profile [1]. We built the interaction network using over 15 million tweets published by almost two million users.

In our case study, we are interested in answering these specific research questions:

1. How did the interaction network polarization evolved before and during the Brazilian election, and how did specific groups contribute to overall network polarization?
2. How was the information spread before and during the elections, both in the interaction network among all users and in each group individually, and what were the differences in the difficulty of achieving coordinated communication between the groups?

To answer the first research question, we adapt the widely used modularity measure based on the principle that the density of internal links is higher than that of external links [32]. We develop a methodology called d-modularity that allows us to evaluate the relative contribution of one specific community to the network's modularity. The proposed d-modularity satisfies a fundamental property for such a measure: the sum of the contributions of all communities produces the overall modularity of the network. Our methodology is general enough to be used in any network with several communities, where the contribution of specific groups to overall network polarization is needed to investigate.

We found that, in our case study, while the overall modularity (somewhat unexpectedly) decreased, the contribution of the right-leaning group to this modularity stayed very high during the analyzed period. The contribution of other groups was less significant and more unsystematic. We present several pieces of evidence to explain this behavior.

For the second research question, the concept of partial domination is applied. Domination is used for studying coordinated communication in social networks [5, 19]. Partial domination reveals the ability of spreaders to influence their communities. We show that right-leaning and left-leaning high-influential users have differences in the way they dominate their groups and the whole interaction network. We also compared the domination profiles of these groups to those of a different, more neutral community.

This work is organized as follows. In the next section, we describe the data capturing process and the collected dataset. Section 3 introduces the theoretical background and our methodology, including the approach to assess the contribution of specific groups to network polarization, and the partial domination concept. The results are presented in section 4. Conclusions and implications are drawn in the last section.

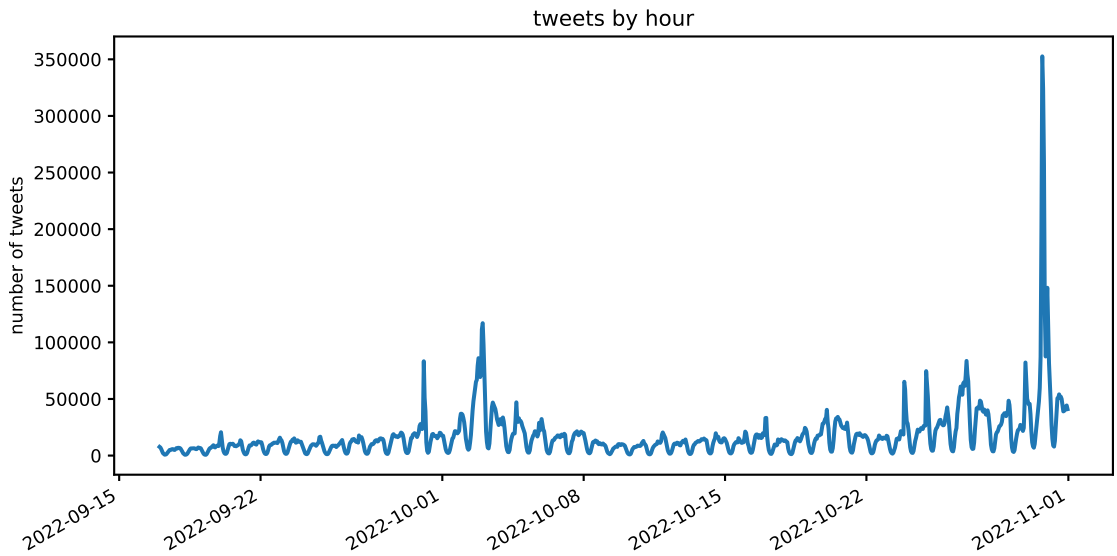

Figure 1. Number of tweets published by hour during the period from 17 September to 30 October 2022.

Table 1. Main descriptive statistics of the collected Twitter dataset on the 2022 Brazilian election.

Statistic	Value
Number of tweet authors	1 890 956
Number of tweets	15 208 839
Number of retweets	13 449 880
Number of replies	1 354 211
Number of quotes	237 601

2. Data collection

We collected a large dataset [24] related to the 2022 Brazilian presidential election from the Twitter social network. There is no single hashtag that identifies all tweets related to the election, since the most used hashtags covered only a small proportion of election-related tweets. Therefore, for constructing our dataset, we used a query composed of a list of election-related words in Portuguese: *eleição OR eleições OR eleitoral OR eleitorais*. The period for data collection was from 17 September to 30 October 2022, resulting in more than 15 million tweets.

The collected data can be roughly divided in two periods: the period of 15 days before the first round of voting, including the day of the elections first round that took place on 2 October 2022; and the period between the two rounds of elections. The second round happened on 30 October 2022. Figure 1 shows the users' hourly tweeting pattern during the considered period. The curve reaches two maximums, corresponding to the days of two rounds of the elections.

Table 1 shows the main characteristics of the collected data, namely the number of collected tweets, unique users, and the number of tweets by its type: retweet, reply, quote. This classification is not a partition, since a tweet can be at the same time a quote and a reply. A tweet may also be unclassified if it is not replying, quoting, or reposting another message. As can be seen, most tweets are retweets, generally representing endorsements of others' opinions.

3. Methods

We have shown that most tweets (specifically, 88.4%) are retweets, representing some kind of support for the opinions of others. These links generally reflect positive relations to the idea or message shared by others. For this reason, we used a network approach to model the interactions during the analyzed time window.

We model the retweet network as a directed graph $G(V,A)$, being V a set of n vertices that represents the users, and A a set of m arcs representing the retweets. An arc $a=(x,y)$ indicates that the user y retweeted one or more tweets published by the user x, and indicate the direction of information propagation, from x to y. We also created networks that represent the interactions restricted to each day.

Table 2. Largest communities identified by the multilevel algorithm by Blondel *et al* [3].

Community	# of users	Some representative members
Left-leaning	552 578	@LulaOficial, @siteptbr, @HaddadDebochado
News media, press	244 296	@JornalOGlobo, @g1, @UOLNoticias, @folha
Right-leaning	241 641	@JovemPanNews, @revistaoeste, @jairbolsonaro

3.1. Community detection

Conover *et al* showed that, since retweets generally represent endorsements of others' opinions, the community structure of retweet networks can be used to predict the political alignment of users with at least 95% accuracy [8]. The graph of retweets analyzed in this work is sparse, and very large, containing 1 203 164 vertices (Twitter users) and 8 504 687 arcs, each representing one or more retweets.

Due to the size of our dataset, the choice of the community detection method is quite restricted. For these reason, we used the community detection approach based on the multilevel algorithm of Blondel *et al* [3]. Computer simulations on large networks show that the complexity of this algorithm is linear on sparse graphs [3]. The period used by the community detection algorithm corresponded to all tweets published from 17 September to 29 October 2022. The election day (30 October 2022) was excluded because it presented a highly uncommon pattern compared to the rest of the period; thus, excluding it, we guarantee that last-hour binary voting decisions on the final data collection day will not affect the communities formed during the period. However, the data collected during second round of elections was used in the rest of the study.

The three largest user groups identified by the community detection algorithm are presented in table 2, ordered by their size. They are easily identified with three large digital communities. The largest one corresponds to left-leaning users. The second one corresponds mostly to news media and press followers. The third one contains right-leaning users. Together these three groups cover 86.3% of the users set. There were many other groups whose size was significantly smaller. User memberships in these groups were also considered in our analysis, but specific results are not presented for these communities.

We notice that Fortunato and Barthélemy [15] studied the resolution limits in community detection methods that use modularity optimization, pointing out that detected modules with a size $\sqrt{2L}$ or smaller can result from an arbitrary merge of smaller structures, where L is the total number of links in the network, i.e. there may be issues when detecting modules whose size is of the order of $\sqrt{L}$ or smaller. Modularity optimization may fail in these cases. However, Fortunato and Barthélemy also show that a module whose size is significantly larger than $\sqrt{2L}$ is not likely to conceal substructures, because this only happens if all hidden submodules are very fuzzy communities.

In our study, we are interested in communities that represent a significant proportion of the number of nodes in the network: the size of these modules is much larger than $\sqrt{2L}$. In our case, $\sqrt{2L} \approx 4000$, while the three larger modules we identified have hundreds of thousands of nodes and cover 86.3% of the network. A different partitioning of smaller communities should not bring significant changes to the results of this study.

3.2. Modularity and contribution of groups to polarization

For a given division of the network's vertices into some collection of groups or communities, modularity [30] reflects the concentration of edges within groups compared with random distribution of edges between all vertices regardless of the groups. We use the modularity of the underlying undirected graph of $G(V,A)$ for measuring the polarization of the network we built with the collection of communities obtained in the previous section.

More formally, let us suppose that a graph $G(V,E)$ with $|V| = n$ vertices and $|E| = m$ edges, have its vertex set V partitioned into k disjoint groups $\{A_1, A_2, \ldots, A_k\}$. The modularity Q is defined as:

$$Q = \frac{1}{2m} \sum_{u,v \in V} \left(a_{uv} - \frac{d(u)d(v)}{2m} \right) \cdot \delta_{g_u g_v}, \tag{1}$$

where $d(v)$ is the degree of node $v \in V$; g_v the index of v's group; $a_{uv} = 1$ if there is an edge between nodes u and v, 0 otherwise; and δ_{ij} is the Kronecker delta.

While modularity can measure the overall level of polarization of the retweet network for some time window, it is not designed to evaluate the contribution of each group to network's polarization. Note that if we evaluate the modularity of the subgraph induced by some group A_i (i.e. a subgraph that contains all vertices in A_i and all edges between vertices in A_i), then the modularity value Q for this single group will be equal to zero [29]. A previous study showed that there is a lack of consolidated group-level measures for evaluating polarization [21].

For evaluating the contribution of some specific group to the overall network modularity, we developed an approach called d-modularity. Since $\delta_{g_u g_v} = 1$ if and only if the vertices u and v are from the same group (and is zero otherwise), we can rewrite the equation (1) that defines modularity Q as follows:

$$Q = \frac{1}{2m}\sum_{i=1}^{k}\sum_{u,v\in A_i}\left(a_{uv} - \frac{d(u)d(v)}{2m}\right). \quad (2)$$

Now let us suppose that one specific group A_i is chosen. Let us consider one addend in equation (2) that corresponds to the group A_i (the same reasoning applies to each group $A_i = A_1, A_2, \ldots, A_k$). Note that this term $\sum_{u,v\in A_i}(a_{uv} - d(u)d(v)/2m)$ considers only pairs of vertices u and v that both belong to A_i, and $\sum a_{uv}$ represents the actual number of in-group edges. Moreover, $\sum d(u)d(v)/2m$ represents the expected number of A_i's in-group edges after rewiring or randomizing the edges in the network while preserving the degree of every vertex (randomization known as the configuration model). If group A_i gives no more within-community edges than would be expected by random chance, then the whole term $\sum_{u,v\in A_i}(a_{uv} - d(u)d(v)/2m)$ will be nullified. On the other hand, if group A_i gives significantly more within-community edges than would be expected by random chance, then the contribution of this group to network modularity will be large.

Therefore, we propose the following way to measure the contribution of a specific group A_i to the overall modularity Q of the network. Let the contribution Q_i of the group A_i to network's modularity be defined in the following way:

$$Q_i = \frac{1}{2m}\sum_{u,v\in A_i}\left(a_{uv} - \frac{d(u)d(v)}{2m}\right).$$

Note that, as expected, $\sum_{i=1}^{k} Q_i = Q$, i.e. the sum of the contributions of all groups produces the overall network modularity. Now, the d-modularity d_i of G respect to the group A_i is defined as:

$$d_i = \frac{Q_i}{Q},$$

In short, the d-modularity d_i reflects the relative contribution of some specific group A_i to the overall modularity of the network.

Figure 2 shows an example of a graph with 12 vertices partitioned into three groups, whose overall modularity is 0.402. The three communities are equally-sized, with four vertices each. Note that each red or blue vertex has exactly two in-group adjacent vertices, and one adjacent vertex from a different group. Each black vertex, however, has exactly three in-group neighbors, and at most one adjacent vertex from a different group. Correspondingly, the group A_1 composed of black nodes contributes with $Q_1 = 0.180$ to modularity, thus $d_1 = 0.448$, or 44.8%. On the other hand, the contribution of the other two groups A_2 and A_3, composed of red and blue vertices, respectively, is quite smaller ($Q_2 = Q_3 = 0.111$, and $d_2 = d_3 = 0.276$, or 27.6%).

To illustrate other possible d-modularity values that might occur in practice, suppose that we have a graph $G(V,E)$ with strong community structure (modularity values in the range from 0.3 to 0.7 [30]), several groups, and one specific group A_i is chosen. We highlight four possibilities regarding the d-modularity value d_i obtained using the described procedure:

1. If $d_i \approx 1$, then $Q_i \approx Q$, and the network's modularity is strongly affected by the connections of vertices in A_i. This fact indicates that the contribution of A_i to the polarization of the network (A_i's distribution of within-group and out-group edges) is decisive for overall network's strong community structure.
2. If $d_i > 0$ (but is neither too close to 1 nor too close to 0), then the contribution of A_i to the polarization of the network is not negligible. Still, neither is it the only one that affects the network's modularity.
3. If $d_i \approx 0$, then $Q_i \approx 0$, and the network's modularity is not affected significantly by A_i's vertices connections. In this case, the chosen group's contribution to network polarization is neutral, and the distribution of within-group and out-group edges in A_i does not differ too much from that expected by random chance, implying that the strong community structure in the graph is mostly influenced by the contributions Q_j of the rest of groups, with $j = 1, \ldots, k, j \neq i$.
4. If $d_i < 0$, then the contribution of A_i to the network's modularity is negative. This case is improbable to happen in real networks, neither occurred in our case study, since it implies heterophily in the group's vertex connections.

The analysis of the retweet network's modularity and the d-modularity values was performed day-by-day, that is, considering, at each step, the interactions performed during each 24 h period. The goal was to

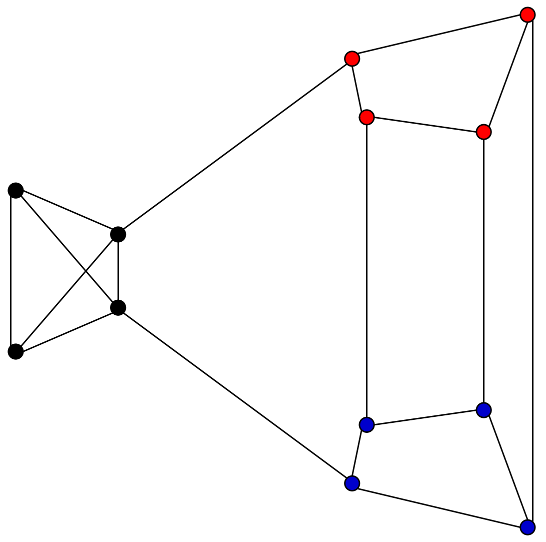

Figure 2. Example of a graph with three groups that contribute unequally to network's modularity $Q = 0.402$. The contribution Q_1 of the black group is 0.180 ($d_1 = 0.448$, or 44.8%), while the contribution of red and blue groups are $Q_2 = Q_3 = 0.111$ ($d_2 = d_3 = 0.276$, or 27.6%).

measure the evolution of network polarization during the analyzed period, as well as the contribution of groups to overall Twitter network polarization.

3.3. Domination in interaction graphs

In this section, we are interested in studying coordinated information spread both in the interaction network among all users and in each group separately, as well as differences in information spread among the groups. For this purpose, we used the concept of domination. The mathematical study of domination in graphs comes from the 1960s. More recently, the use of domination for the study of coordinated communication in social networks was studied by Campan *et al* [5]. The reader is referred to the book by Haynes *et al* [19] for a more detailed review of domination in graphs.

In this work we intend to find some (expected to be small) subset of information spreaders that is capable of reaching (*dominating*) all or most of the users that participate in a discussion. Our research hypothesis is that in real life, small subsets of information spreaders may be used for the task of reaching (or communicating with) a large number of users in some network or in some specific group.

More formally, the problem can be stated in the following way. Let V be the set of n users, and $D \subseteq V$ be the set of spreaders, i.e. users whose tweets were shared or retweeted at least once. We are interested in finding the smallest subset S^* of D such that the number of users reached from S^* is at least $\rho \cdot n$, where ρ is a parameter that reflects the minimum proportion of users to be reached. We may study this problem considering, for example, $\rho = 100\%$, $\rho = 75\%$, or $\rho = 50\%$ of a given network or group. A variant of this problem (which will not be analyzed in this section) proposes finding the smallest subset S^* not of D, but of some specific subset D', representing a specific category of spreaders.

In the case when $\rho = 100\%$, the abovementioned problem can be formulated as the directed dominating set (DDS) problem previously studied, for example, by Habibulla *et al* [18]. A directed graph $G(V,A)$ is built, where V is the set of vertices that represent the users, and A is the set of arcs representing relations between the users. The smallest subset $S^* \subseteq V$ that reaches each vertex represents a group of spreaders that dominates the interaction network. An instance of the classical dominating set (DS) problem, known to be NP-hard, can be easily reduced to an instance of DDS problem by replacing each undirected edge (u,v) with a pair of directed arcs $\overrightarrow{uv}$ and $\overrightarrow{vu}$. Consequently, DDS is also NP-hard, that is, there is no efficient algorithm for finding the optimum solution of this problem unless $P = NP$.

In the more general case when $\rho \leqslant 100\%$, however, the problem requires a specific formulation, which we call partial DDS (PDDS) problem. An instance of the PDDS problem is composed of a directed graph $G(V,A)$ and a parameter ρ. The goal is to find a DS that reaches at least $\rho|V| = \rho n$ vertices. This problem is similar to the variant of the set cover problem called partial cover [25]. The PDDS is also NP-hard, since it contains the DDS problem as a special case ($\rho = 1$).

We note that there is also another common formulation for problems involving partial covering. In this variant, instead of ρ, there is an integer k, such as the number of users reached from S^* must be at least k. It is easy to show that this versions are equivalent to our formulation by setting $k = \lceil \rho \cdot n \rceil$.

Since the problem is NP-hard, and the considered instances are very large, a greedy algorithm is used. Algorithm 1 presents the greedy heuristic for the PDDS problem. The algorithm is an adaptation of an existing greedy heuristic for the DS problem studied, for example, by Parekh [31]. It is known to be the best in terms of the approximation ratio for DS unless $P = NP$ [7]. In the algorithm, $W(v)$ denotes the set of yet uncovered vertices among successors of the vertex v, also called the span of v.

The algorithm starts setting the current solution S^* to an empty set (line 1), and the number of covered vertices f^* to zero (line 2). Initially, the span of each vertex v is initialized by the set of their out-neighbors $\mathcal{N}^+(v)$ (line 3). In this notation, the set $\mathcal{N}^+(v)$ includes v. The algorithm performs a number of iterations (lines 4–9). In each iteration, the algorithm picks the vertex with the largest span, i.e. the spreader that reaches the maximum number of uncovered users, updating the current solution S^*, the number of covered vertices f^*, and the span $W(v)$. When the number of covered vertices f^* reaches at least $\rho|V|$ vertices, the algorithm ends returning S^* (line 10).

Algorithm 1. Greedy algorithm for partial directed dominating set problem.

Input: $G = (V, A)$, **constant** ρ
Output: S^*
1: $S^* \leftarrow \emptyset$;
2: $f^* = 0$;
3: $W(v) \leftarrow \mathcal{N}^+(v) \quad \forall\, v \in V$
4: **while** $f^* < \rho|V|$ **do**
5: $\quad u \leftarrow \mathrm{argmax}_{v \in V} |W(v)|$;
6: $\quad S^* \leftarrow S^* \cup \{u\}$;
7: $\quad f^* = f^* + |W(u)|$;
8: $\quad W(v) \leftarrow W(v) \setminus W(u) \quad \forall\, v \in V$
9: **end while**;
10: **return** S^*;

Note that, since any instance of the DS problem in undirected graphs can be linearly transformed in a PDDS instance with $\rho = 1$ by replacing each undirected edge (u, v) with a pair of directed arcs $\overrightarrow{uv}$ and $\overrightarrow{vu}$, it can be easily verified that the solutions are in a one-to-one correspondence. Therefore, as shown in [31], the greedy algorithm (nor any other one) cannot do better than $\mathcal{H}(\delta + 1)$ times the size of an optimal DS, where δ is the maximum out-degree in the graph, and $\mathcal{H}(\delta + 1) = \sum_{i=1}^{\delta+1} 1/i = \Theta(\log \delta)$, and there is no $o(\log \delta)$-approximation for this problem unless $P = NP$.

In this work, several parameters are explored when building instances for the PDDS from our data. Different values of ρ are used. Moreover, we study the domination in the subgraphs that represent the interactions inside user groups, as well as in the whole network.

4. Results

This section presents the main findings of our study, which provide several insights into the evolution of polarization, influence, and domination during the 2022 Brazilian elections. The results bring evidence of the extreme isolation of specific communities and the ease of accomplishing coordinated communication inside these groups.

Figure 3 exemplifies the generated interaction networks, presenting the graph of interactions that occurred on 2 October 2022, the day of the first round of the elections. We recall that the communities represented by red, blue, and yellow vertex colors, among others, were generated considering all interactions over the analyzed period, not only those that happened during this specific day. These colors indicate membership in the left-leaning, right-leaning, and news media or press followers communities, respectively. The visualization of the network was generated using only 5% of randomly chosen user–user interactions among the almost 700 000 generated on 2 October.

4.1. Polarization dynamics and isolation of groups

The daily analysis of the retweet network's modularity and d-modularity values is presented in this section. Our goal is to measure the evolution of network polarization, as well as the contribution of groups to overall polarization of the network of retweets, during the analyzed period. While some social networks show a massive presence of strongly polarized communities, in others, echo chambers may not arise [11]. We observe that, in practice, modularity values for networks with strong community structure typically fall in the range from about 0.3 to 0.7, and higher values are rare [30].

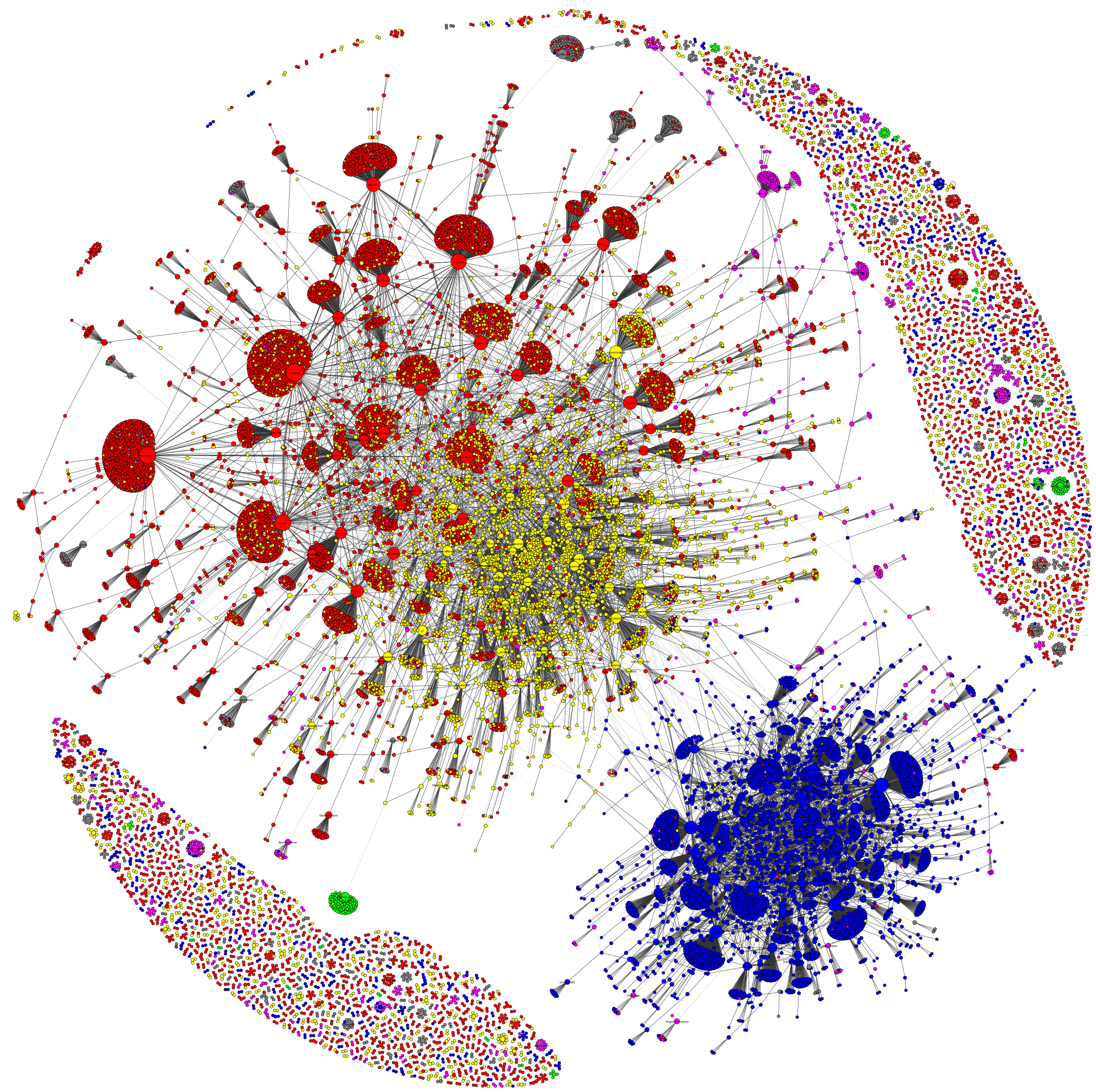

Figure 3. Network that represents the interactions during 2 October 2022, the day of the first round of the elections. Red, blue, and yellow vertex colors indicate membership in the left-leaning community, right-leaning community, and media followers group, respectively.

Figure 4 shows the evolution of polarization and group isolation before and during the Brazilian elections in 2022. Despite identifying an unexpected decreasing modularity trend in the whole network, when analyzing the contribution of specific groups, we see a different picture. While the overall modularity decreased due to more frequent interactions between users from different groups, right-leaning group contribution Q_r to the network's modularity stayed stably high during the entire analyzed period. Left-leaning group reached the higher value of Q_l during the first round of elections day (when the curves meet each other), slightly surpassing the right-leaning group contribution Q_r, also getting very close to right-leaning group's curve between 23 and 24 October, a week before the second round. However, on all other analyzed days, right-leaning group's contribution Q_r was higher.

Note that the right-leaning group is significantly smaller than the group of left-leaning users (see table 2). However, it is contribution to network's modularity is higher than the contribution of the left-leaning, significantly larger, group.

On the other hand, figure 5 shows the evolution of d-modularity values for right-leaning and left-leaning groups. Since there was a stable and high modularity over whole the period (0.510 on average, with standard deviation of only 0.036), the shape is very similar to that of figure 4. The increasing right-leaning d-modularity trend shows that, despite the overall decreasing modularity, the right-leaning group's polarization stayed high and even increased its degree of isolation from other groups.

On average, during all the analyzed period of 44 days, 44.1% of the modularity value is explained by right-leaning group's contribution (standard deviation 5.1%), while the contribution of left-leaning larger

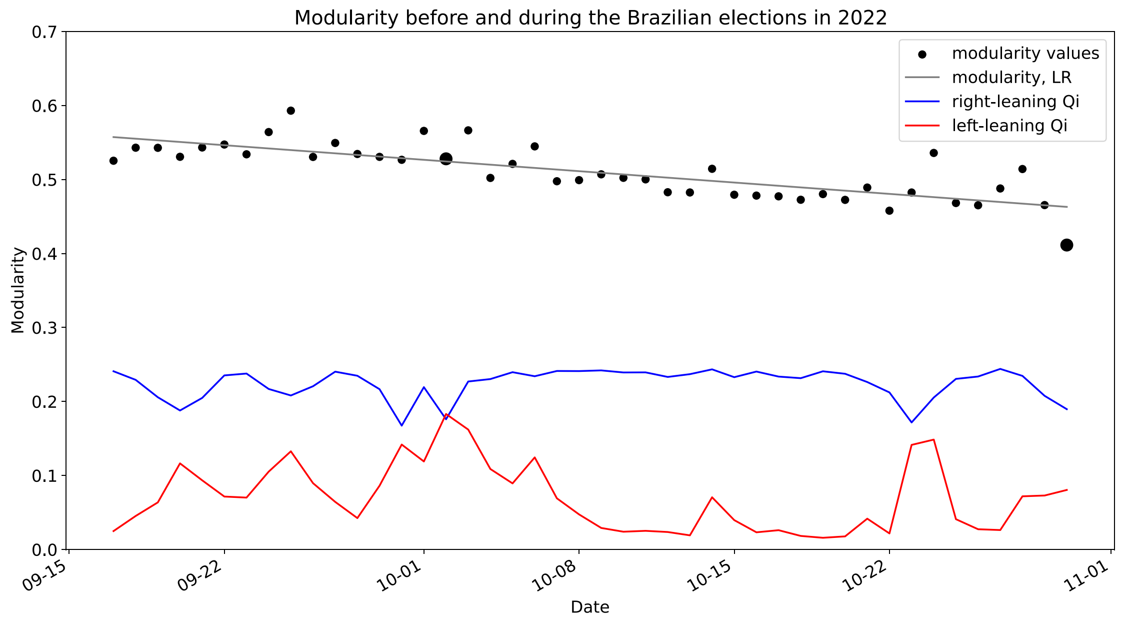

Figure 4. Evolution of polarization and group isolation before and during the Brazilian elections in 2022. Black points represent modularity values for each day, with two bigger points representing the two rounds of voting days. Decreasing gray line shows the modularity trend obtained using least square regression (LR). Blue and red lines represent the contributions Q_i of, respectively, right-leaning and left-leaning groups, to network's modularity (see section 3.2).

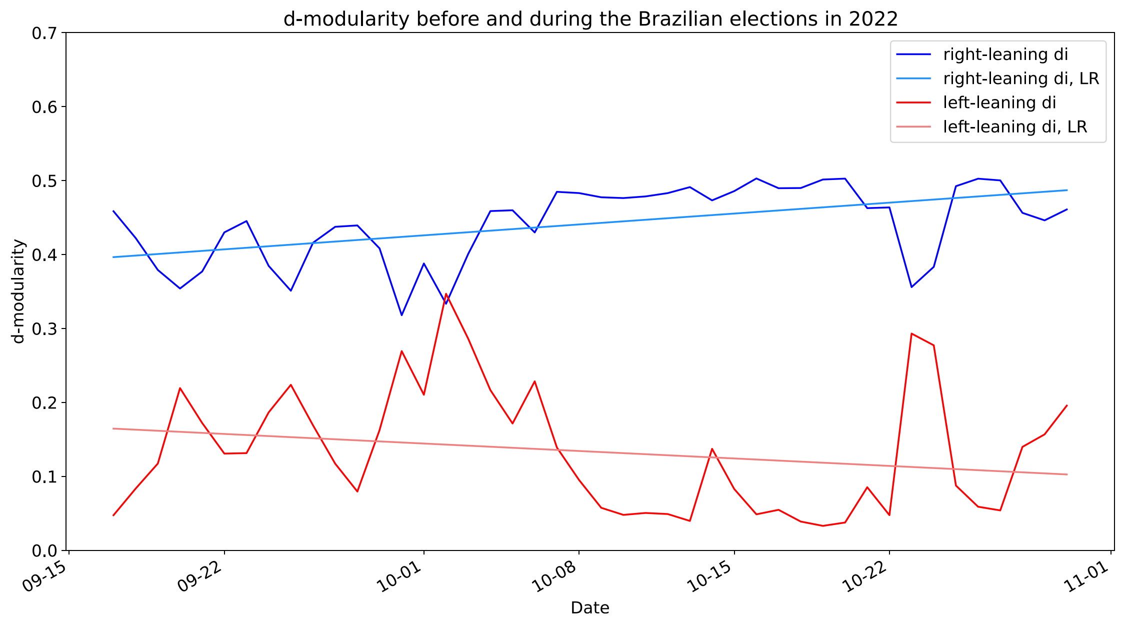

Figure 5. Evolution of d-modularity before and during the Brazilian elections in 2022. Blue and red lines represent the d-modularity values d_i for, respectively, right-leaning and left-leaning groups. Light blue and light red lines show the d-modularity trends obtained using least square regression (LR) over right-leaning and left-leaning d-modularity values, respectively.

group is 13.4% on average (standard deviation 8.4%). We can conclude that the network's modularity is much less affected by the connections of left-leaning users when compared to the right-leaning ones, since the left-leaning group is much more integrated with the other groups in the network.

4.2. Dominating the interaction network during the 2022 Brazilian election

For measuring domination during the 2022 Brazilian election, we consider two classes of instances. In the first class, the goal is to cover vertices in a single group using spreaders from this group alone. That is, the instance of the PDDS problem is a subgraph whose vertices belong to one single community. In this case, we measure the capability of reaching all users in the considered group, as well as the ability to communicate

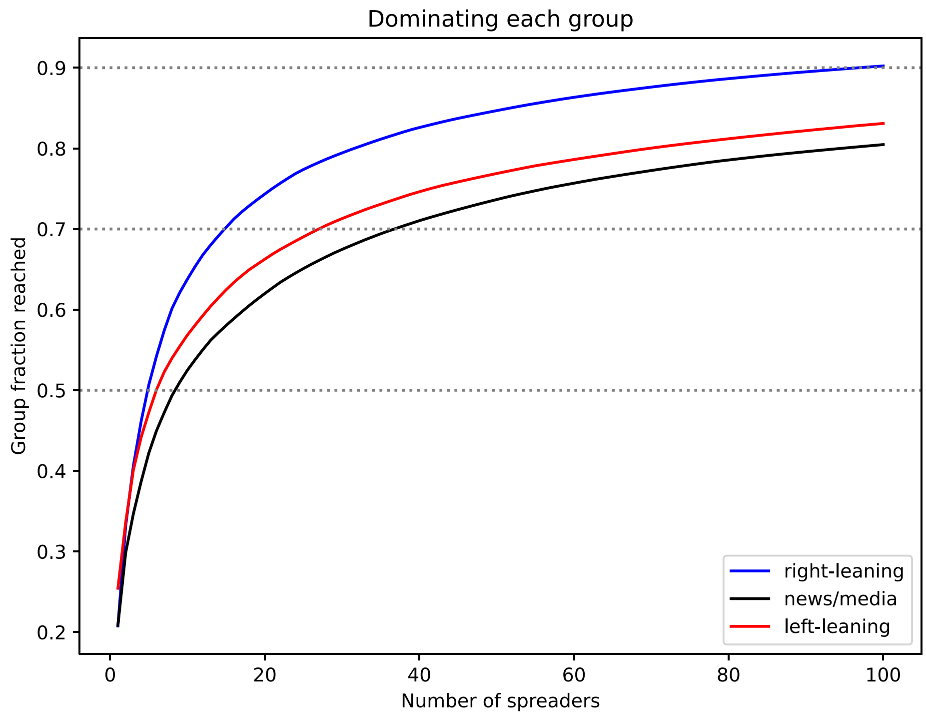

Figure 6. Results for in-group domination. Most of the group can be reached using a small group of spreaders. For example, the number of spreaders necessary for reaching 70% of the group are 15, 27, and 37 for right-leaning, left-leaning, and media groups, respectively. Right-leaning spreaders are able to dominate up to 90% of their group using few spreaders.

quickly within the group. For example, if there is a small subset of spreaders that can reach the majority of the group, the ability of quick communication inside the group is high.

In the second class of instances, the goal is to cover all vertices in the network by using spreaders from one single group. Therefore, the capability of some subset of information spreaders (e.g. right-leaning or left-leaning ones) to reach a substantial segment of users that participate in the discussion is measured. Note that in this case a slight modification of the PDDS problem is used, and we attempt to find the smallest subset S^* of some specific subset of spreaders A_i.

Figure 6 shows the results for the first class of instances, where the goal is to cover vertices in a single group by spreaders from this group. Figure 7 shows the results for the second class of instances, where the goal is to cover the whole network by using spreaders from one single group. We present the results for the left-leaning, right-leaning, and also for the media followers group, taken in order to obtain the domination profile of a more neutral in the ideological spectrum community. In each case, the presence of a point (x_0, ρ_0) in the graph means that the greedy Algorithm 1 described in section 3.3 returned x_0 spreaders in the DDS with algorithm's parameter ρ equal to ρ_0.

Figure 7 shows that in the whole network the left-leaning high-influential information spreaders dominated, reaching a substantial fraction of the vertices with less spreaders. This outcome is, in a way, expected when considering the 2022 Brazilian election results. However, the magnitude of the difference between the groups is very large. For example, using at most 100 spreaders, media and right-leaning influential accounts can reach approximately 30% and 20% of users, respectively, very far from the almost 50% reached by 100 left-leaning spreaders. This confirms that right-leaning spreaders are confined to their group, which represents approximately 20% of the users.

However, when comparing domination inside the groups, we surprisingly found that the results are essentially the inverse. As figure 6 shows, right-leaning high-influential users dominate their communities using much less vertices than in the other analyzed groups. To perceive it, we can draw an horizontal line at, for example, $\rho = 0.7$, and compare the number of spreaders necessary for reaching 70% of the group obtained by intercepting the curves. The obtained values are 15, 27, and 37 for right-leaning, left-leaning, and media groups, respectively. Therefore, the differences among the groups are not negligible, and the right-leaning spreaders show as the most capable of accomplishing coordinated communication inside their group.

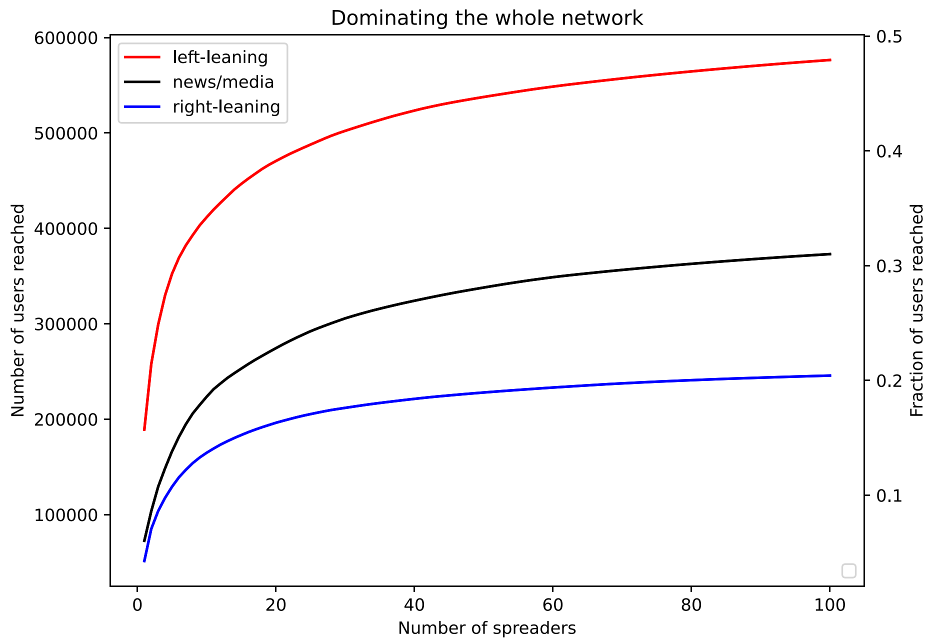

Figure 7. Network domination results. Left-leaning spreaders are able to reach a large portion of the network, while right-leaning spreaders may reach a small portion of the network, comparable to the size of their group.

Table 3. Tweets with the overall highest number of retweets in the analyzed period.

Author's user name	Tweet ID	Date	# of retweets
@Casimiro	1584293297841111040	23 October 2022	189 181
@g_agurgel	1586727126656602114	30 October 2022	33 238
@odontinho	1576601689523560449	2 October 2022	31 306
@SF_Moro	1577313831449141249	4 October 2022	29 728
@RomeuZema	1577322128231350273	4 October 2022	28 862

A curious but revealing fact is that the most influential tweet from our dataset was made by Casimiro [6], a non-political Brazilian journalist, presenter, sports commentator, and streamer. As shown in table 3, it reached 189 181 retweets in the analyzed period, more than five times the second most influential tweet.

5. Conclusions and implications

In this study, we investigated the evolution of network and group polarization, influence, and domination in online interaction networks, using Twitter data collected before and during the 2022 Brazilian elections as a case study.

Partially dominating sets composed by influential profiles from different groups may reveal an internal organization of these communities and their ability to perform coordinated communication with the users. By comparing the reach of Partially dominating sets, we may evaluate their ability to dominate the groups and segments of the entire network during some specific time window. We found that in the whole Brazilian elections' retweet network the left-leaning high-influential information spreaders dominated, reaching a substantial fraction of the network with few spreaders. However, inside the groups, the results are inverse, and right-leaning high-influential users dominate their communities more easily (with fewer spreaders) than in the other analyzed groups, facilitating coordinated communication inside the groups.

Our work also presents a methodology called d-modularity used for evaluating the contribution of specific groups to network's polarization by using the well-known modularity measure. We show how different communities may contribute unequally to the network's modularity. In our case study, we found that, on average, 44.1% of the daily modularity of the interaction network was explained by right-leaning group's contribution, while the contribution of left-leaning group was only 13.4%. This result is specially interesting since the right-leaning group is significantly smaller than the group of left-leaning users. The contribution of the right-leaning group to network's modularity is higher than the contribution of the left-leaning, significantly larger, group.

These results bring evidence of extreme isolation and the ease of accomplishing coordinated communication that characterized right-leaning groups during the 2022 Brazilian elections. Based on the presented evidence, we may hypothesize that the isolation and the ease of domination inside this community influenced the coup events that occurred on 8 January 2023, in Brasília, which included the invasion of the Supreme Federal Court, the National Congress building, and the Planalto Presidential Mansion.

Among the limitations of our study, we can mention that the analyzed period ends on 30 October 2022, and does not cover further developments, nether 8 January 2023 events. Consequently, the causal relationship between the isolation and the ease of domination of specific communities during the elections, and the coup events that occurred on 8 January 2023, in Brasilia, may be better established in future research.

We also intend to explore in future work the possibility of using other community and polarization detection methods that can be used for large networks, such as those based on the spectrum of the Laplacian matrix [14].

Our findings have broader implications and could be applied to data from different social networks, providing insights into how highly polarized and easily manipulated communities arise over time. In particular, we shed light on the potential for specific online communities to become increasingly isolated, even if the overall network modularity decreases on time: the decreasing modularity trend in online interaction networks does not exclude radicalization. The connection between the ease of coordinated communication and the rise of easily manipulated communities may facilitate real-world large attacks against other groups or institutions and coordinated efforts to delegitimize elections.

Data availability statement

The data that support the findings of this study will become openly accessible on Mendeley Data after a period of time no longer than 6 months (before 1 October 2023). The data that support the findings of this study are also available upon reasonable request from the authors. http://doi.org/10.17632/x7ypgrzr3m.1.

Acknowledgments

Ruben Interian was supported by research Grant 2021/12456-5, São Paulo Research Foundation (FAPESP). Francisco Rodrigues acknowledges CNPq (Grant 309266/2019-0) and FAPESP (Grant 19/23293-0) for the financial support given for this research. Research carried out using the computational resources of the Center for Mathematical Sciences Applied to Industry funded by FAPESP (Grant 2013/07375-0).

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ORCID iDs

Ruben Interian https://orcid.org/0000-0001-5878-6462
Francisco A Rodrigues https://orcid.org/0000-0002-0145-5571

References

[1] Kaleel S B and Abhari A 2015 Cluster-discovery of Twitter messages for event detection and trending *J. Comput. Sci.* **6** 47–57
[2] Baumann F, Lorenz-Spreen P, Sokolov I M and Starnini M 2021 Emergence of polarized ideological opinions in multidimensional topic spaces *Phys. Rev.* X **11** 011012
[3] Blondel V D, Guillaume J-L, Lambiotte R and Lefebvre E 2008 Fast unfolding of communities in large networks *J. Stat. Mech.* p 10008
[4] Bonow Soares F, Recuero R and Zago G 2019 Asymmetric polarization on Twitter and the 2018 Brazilian presidential elections *Proc. 10th Int. Conf. on Social Media and Society* (*SMSociety 2019*) (*New York, NY, USA*) (Association for Computing Machinery) pp 67–76
[5] Campan A, Marius Truta T and Beckerich M 2015 Fast dominating set algorithms for social networks *Proc. 26th Modern AI and Cognitive Science Conf.* pp 55–62
[6] Casimiro Miguel Vieira da Silva Ferreira 2022 Most influential tweet during the 2022 Brazilian elections (available at: https://twitter.com/Casimiro/status/1584293297841111040)
[7] Chlebík M and Chlebíková J 2004 Approximation hardness of dominating set problems *Algorithms—ESA 2004* ed S Albers and T Radzik (Springer) pp 192–203
[8] Conover M D, Goncalves B, Ratkiewicz J, Flammini A and Menczer F 2011 Predicting the political alignment of Twitter users *2011 IEEE 3rd Int. Conf. on Privacy, Security, Risk and Trust and 2011 IEEE 3rd Int. Conf. on Social Computing* pp 192–9
[9] Cota W, Ferreira S C, Pastor-Satorras R and Starnini M 2019 Quantifying echo chamber effects in information spreading over political communication networks *EPJ Data Sci.* **8** 35

[10] Dal Maso C, Pompa G, Puliga M, Riotta G and Chessa A 2015 Voting behavior, coalitions and government strength through a complex network analysis *PLoS One* **9** 1–13
[11] De Francisci Morales G, Monti C and Starnini M 2021 No echo in the chambers of political interactions on Reddit *Sci. Rep.* **11** 2818
[12] DeSilver D 2022 The polarization in today's Congress has roots that go back decades *Technical Report* Pew Research Center
[13] Dinkelberg A, O'Reilly C, MacCarron P, Maher P J and Quayle M 2021 Multidimensional polarization dynamics in us election data in the long term (2012–2020) and in the 2020 election cycle *Anal. Soc. Issues Public Policy* **21** 284–311
[14] Fortunato S 2010 Community detection in graphs *Phys. Rep.* **486** 75–174
[15] Fortunato S and Barthélemy M 2007 Resolution limit in community detection *Proc. Natl Acad. Sci.* **104** 36–41
[16] Garcia D, Abisheva A, Schweighofer S, Serdült U and Schweitzer F 2015 Ideological and temporal components of network polarization in online political participatory media *Policy Internet* **7** 46–79
[17] Garimella K and Weber I 2017 A long-term analysis of polarization on Twitter *Proc. 11th Int. AAAI Conf. on Web and Social Media* pp 528–31
[18] Habibulla Y, Zhao J-H and Zhou H-J 2015 The directed dominating set problem: generalized leaf removal and belief propagation *Frontiers in Algorithmics* ed J Wang and C Yap (Springer International Publishing) pp 78–88
[19] Haynes T W, Hedetniemi S and Slater P 1998 *Fundamentals of Domination in Graphs* (Taylor & Francis)
[20] Heading S and Zahidi S 2023 The global risks report 2023, 18th edn *Technical Report* World Economic Forum
[21] Interian R, Marzo R G, Mendoza I and Ribeiro C C 2022 Network polarization, filter bubbles and echo chambers: an annotated review of measures and reduction methods *Int. Trans. Oper. Res.* **30** 3122–58
[22] Interian R, Moreno J R and Ribeiro C C 2021 Polarization reduction by minimum-cardinality edge additions: complexity and integer programming approaches *Int. Trans. Oper. Res.* **28** 1242–64
[23] Interian R and Ribeiro C C 2018 An empirical investigation of network polarization *Appl. Math. Comput.* **339** 651–62
[24] Interian R and Rodrigues F 2023 *2022 Brazilian Election Twitter Dataset* (Mendeley Data)(https://doi.org/10.17632/x7ypgrzr3m.1)
[25] Kearns M J 1990 The computational complexity of machine learning *ACM Distinguished Dissertation Award* Harvard University
[26] Maoz Z and Somer-Topcu Z 2010 Political polarization and cabinet stability in multiparty systems: a social networks analysis of European parliaments, 1945-98 *Br. J. Polit. Sci.* **40** 805–33
[27] Markgraf M and Schoch M 2019 Quantification of echo chambers: a methodological framework considering multi-party systems *27th European Conf. on Information Systems* (available at: https://aisel.aisnet.org/ecis2019_rp/91) (Accessed 22 May 2022) pp 1–16
[28] Morales P R and Cointet J-P 2021 Auditing the effect of social network recommendations on polarization in geometrical ideological spaces *15th ACM Conf. on Recommender Systems* (ACM) pp 627–32
[29] Newman M E J 2016 Equivalence between modularity optimization and maximum likelihood methods for community detection *Phys. Rev.* E **94** 052315
[30] Newman M E J and Girvan M 2004 Finding and evaluating community structure in networks *Phys. Rev.* E **69** 026113
[31] Parekh A K 1991 Analysis of a greedy heuristic for finding small dominating sets in graphs *Inf. Process. Lett.* **39** 237–40
[32] Shang K-K, Small M, Wang Y, Yin Di and Li S 2020 A novel metric for community detection *Europhys. Lett.* **129** 68002
[33] Shi Y, Mast K, Weber I, Kellum A and Macy M 2017 Cultural fault lines and political polarization *Proc. 2017 ACM Web Science Conf.* (ACM) pp 213–7
[34] ul Haq E, Braud T, Kwon Y D and Hui P 2020 Enemy at the gate: evolution of Twitter user's polarization during national crisis *2020 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining* pp 212–6
[35] Weron T and Sznajd-Weron K 2022 On reaching the consensus by disagreeing *J. Comput. Sci.* **61** 101667
[36] Wolfowicz M, Weisburd D and Hasisi B 2021 Examining the interactive effects of the filter bubble and the echo chamber on radicalization *J. Exp. Criminol.* **19** 119–41
[37] Yang B, Shang K-K, Small M and Chao N 2023 Information overload: how hot topics distract from news—COVID-19 spread in the US *Natl Sci. Open* **2** 20220051