Journal of the Brazilian Computer Society, 2025, 31:1, doi: 10.5753/jbcs.2025.3747

Endorsement Networks in the 2022 Brazilian Presidential Elections: a Case Study on Twitter Data

Alejandro Dinkelberg [**University Carlos III de Madrid** | ***alejandro.dinkelberg@uc3m.es***]
Brenda Salenave Santana [**Universidade Federal de Pelotas** | ***bssalenave@inf.ufpel.edu.br***]
Ana L. C. Bazzan [**Universidade Federal do Rio Grande do Sul** | ***bazzan@inf.ufrgs.br***]

Instituto de Informática, Universidade Federal do Rio Grande do Sul, Caixa Postal 15064, Porto Alegre. RS, 91501-970, Brazil.

Received: 11 October 2023 • **Accepted:** 18 December 2024 • **Published:** 10 April 2025

Abstract Online social networks take an increasingly substantive role in modern elections. For example, Twitter (renamed X) provides an online platform to present, discuss and confirm/endorse each other's opinions. In regard of the 2022 Brazilian Presidential elections, we examine the endorsement networks from supporters of potential presidential candidates to reveal characteristic structural properties of those networks. Their construction has involved a data set with over 20 million tweets; their analysis depicts the development of the online supporters within three months of the 2022 presidential election. Within the networks we look into the roles of the most essential supporters and the presence of social bots. In order to uncover changes between election years, we compare our results to the previous election in 2018. This descriptive network analysis provides an overview of the political engagement of Twitter users during the Brazilian Presidential elections.

Keywords: Endorsement networks, 2022 Brazilian Elections, Network analysis, Social bots, Twitter data

1 Introduction

Like many others, Brazilian elections have been highly influenced by social media since at least the 2018 presidential election. Online platforms provide opportunities to observe political engagement and endorsement with political actors, especially during elections. Chen *et al*. [2021], for instance, applied a network approach to uncover alignments among parties and politically salient topics in the 2019 Finnish elections and utilised endorsement networks to evaluate polarisation within these topics. Their results show that Finnish stances on climate politics and immigration are highly aligned. The authors used retweet networks as endorsement networks, in which they link Twitter users through their retweeting behaviour and draw mutual agreement between them.

While we remark that there are several works dealing with social networks and elections, in the present article we focus on those that incorporate elements of graph/network theory in their analysis. For instance, Soares *et al*. [2019] analyse the relationship between social media and the electoral performance of candidates in the 2018 Brazilian presidential election. Specifically, they used data from Brazilian news outlets to understand how they have influenced political discussions on Twitter during the 2018 presidential campaign. They identified an asymmetric polarisation using social network analysis methods (modularity, in-degree, etc.).

The increasing spread of disinformation on social media during the election periods has been a topic of some studies. For example, regarding the 2016 U.S. presidential election, Badawy *et al*. [2018] explore the effects of such disinformation spread by taking a closer look at users who re-disseminated posts on Twitter that were produced by the Russian troll accounts; later, they were revealed by U.S. Congress investigation. Resende *et al*. [2019] deal with spreading disinformation in WhatsApp groups, covering, among others, a period corresponding to the Brazilian presidential campaign in 2018. The authors built a network based on the user interactions and activities in WhatsApp groups and uncovered similarities in the disseminated content within those groups. WhatsApp data was also used in de Freitas Melo *et al*. [2019] to evaluate the spreading of misinformation. They used an epidemiological model and WhatsApp data collected in Brazil, India, and Indonesia to assess the impact of limiting virality in the corresponding networks, showing that some efforts deployed by WhatsApp could significantly delay the spread of information. However, they are ineffective in blocking the propagation of misinformation campaigns through public groups when the content is highly viral.

In Brazil, elections reoccur in a cycle of four years to determine the president, the Congress (both the lower house and the senate), the governors of the 26 states (plus the federal district of Brasilia), and the state-level parliaments. These elections occur at the beginning of October (the first round for president and governors and all houses). A potential run-off election between two candidates for president and/or governor occurs if no candidate can gather more than 50% of the votes. A plethora of parties form the Brazilian parties system, each tagged with a specific number. Of interest in this article, we have the left-wing oriented "Partido dos Trabalhadores" (PT, No. 13), which was present in both run-off rounds of the presidential elections in 2018 and 2022, the "Partido Social Liberal" (PSL, No. 17), the "Partido Liberal" (PL, No. 22), PMDB (15), and PDT (12).

Here, we focus in particular on the Brazilian 2022 presidential election. Nevertheless, the run-off round of the previous election in 2018 shows parallels. Given that, in both elections, a polarised concentration of votes between the same two political fronts occurred, we also included data from 2018 for the candidates in the run-off rounds. Thus, we are interested in comparing the endorsement networks of the potential presidential candidates in both election years. To address this, we analyse the structural patterns in the networks of endorsers of given presidential candidates.

In the 2022 election, eleven candidates were running for president in a polarised scenario. While many candidates tried to brand themselves as centre-left or centre-right alternatives, these were never really perceived as "taking off" candidatures, visible in the polls prior to the election. Even as early as August 2022, which also marks the start of the official campaigning period, a polarised run-off between Jair Bolsonaro and Luis I. Lula da Silva (Lula) was a constant in the public imaginary. Nevertheless, in this article, we present and analyse data that refer not only to these two but include the four best-positioned candidates in the polls, i.e., Ciro Gomes and Simone Tebet. These two candidates took an active part during the campaigning period and participated in TV debates similar to those of Lula and Bolsonaro. Each of them united at least 5% of the voting intention in polls.

In 2018, there were 13 candidates running for president, again in a polarised scenario among these. The election ended in a run-off between Fernando Haddad (PT) and Jair Bolsonaro (in 2018, a member of the PSL), who we, therefore, included in our analysis.

The core element of our analysis is the endorsement networks, which are built on collected tweets from Twitter[1] that contain a set of hashtags. Instead of dealing with the actual text of the tweets (except for extracting hashtags), we recover the retweeting connections of the users and focus on who retweeted a tweet. The assumption is that a retweet signals an agreement and endorsement with the tweet's author (as done, e.g., by Chen *et al*. [2021]). By linking users through a shared agreement based on retweeting, we form the endorsement network to characterise the structural characteristics of the presidential candidate supporters on Twitter.

Our analysis also considered the influence of bot-like activities in such endorsement networks. We used the automation-detection tool *Botometer* [Sayyadiharikandeh *et al*., 2020], a supervised machine-learning tool designed to detect social bots on Twitter [Yang *et al*., 2022], to indicate non-human behaviour in the online interactions observed. The Botometer considers over 1,000 features covering account profiles, content, social networks, and other metrics when analysing an account.

This article is organised as follows. The following section explains how the endorsement networks are built. Then, Section 3 contains the structural characterisation of the endorsement networks for the both elections as well as our analysis of them. A discussion on whether social bots play a different role in the endorsement networks appears in Section 4. Section 5 contains concluding remarks. Further details on the data collection and the datasets are presented in the appendix, in order to focus on the resulting endorsement networks.

[1]We remark that, although Twitter has been renamed to X, we keep referring to a name that is predominantly used in the literature throughout the text.

2 Construction of the endorsement networks

Our interest lies in analysing the structural patterns that occur in the network of endorsers, arising from their retweeting behaviour and the usage of given hashtags. Therefore, this section briefly describes the construction of the endorsement networks from Twitter and the decisions involved in the data collection process.

In previous research (Barberá *et al*. [2015]; Garimella *et al*. [2018]; Chen *et al*. [2021]), uncommented retweets (i.e., a retweet where no additional information is added by the user) were used to build endorsement networks from online social networks such as Twitter. In line with this, Metaxas *et al*. [2021] examined the role of retweets and the intended signalling by their use. Hence, retweets by a user signal agreement, especially when hashtags from a political context are involved. Therefore, we build our data sets by collecting uncommented retweets that contain at least one hashtag from a specific set of hashtags within a given time span.

Briefly, these hashtag sets were identified in two ways (for details about the processes of identifying hashtags, please refer to the appendix). First, we monitored each candidate's official account and extracted the hashtags used in their tweets. Second, we manually selected further hashtags that were used by people tweeting. This means that these did not necessarily originated in the official candidates' accounts. However, such hashtags indicate a positive association with a candidate. For example, for Bolsonaro *#BolsonaroReeleito2022* (Eng.: Bolsonaro reelected), for Lula *#LulaNoPrimeiroTurno* (Eng.: Lula in the first round), for Simone *#SimonePresidente* and for Ciro *#PrefiroCiro* (Eng.: I prefer Ciro).

The period for the data collection for both election years spanned three months, divided (roughly) into monthly periods representing the *beginning, middle* and *end* of the campaign. The starting date for the last period (*end*) was the day of the first election round (see Table 1). To simplify, we refer to *beginning*, *middle* and *end* periods as August, September and October, respectively. Further details on the data collection can be found in the appendix.

The retweet networks reflect endorsement networks, as aforementioned, and form here undirected graphs between Twitter users (vertices). The edges in these retweet networks indicate common endorsements between two users. In our networks, two users are linked if they have been retweeted by the other user at least once. This allows the analysis of structural patterns of the users, in regard of their online engagement with the presidential candidates.

In order to provide an overview, the process that leads to the formation of the endorsement networks is described schematically in Figure 1. In the top left corner, there are three candidates, distinguishable by colors, who post tweets via their accounts on Twitter. Their tweets containing hashtags that promote them are then retweeted (see the green user in the top part of the figure). This (re)tweet is then again retweeted by the magenta user. Thus, an edge is formed

Table 1. Periods of data collection.

Election	**Period**		
	Beginning (Aug)	**Middle (Sep)**	**End (Oct)**
2018	01/08 – 31/08	01/09 – 07/10	08/10 – 28/10
2022	01/08 – 31/08	01/09 – 02/10	03/10 – 30/10

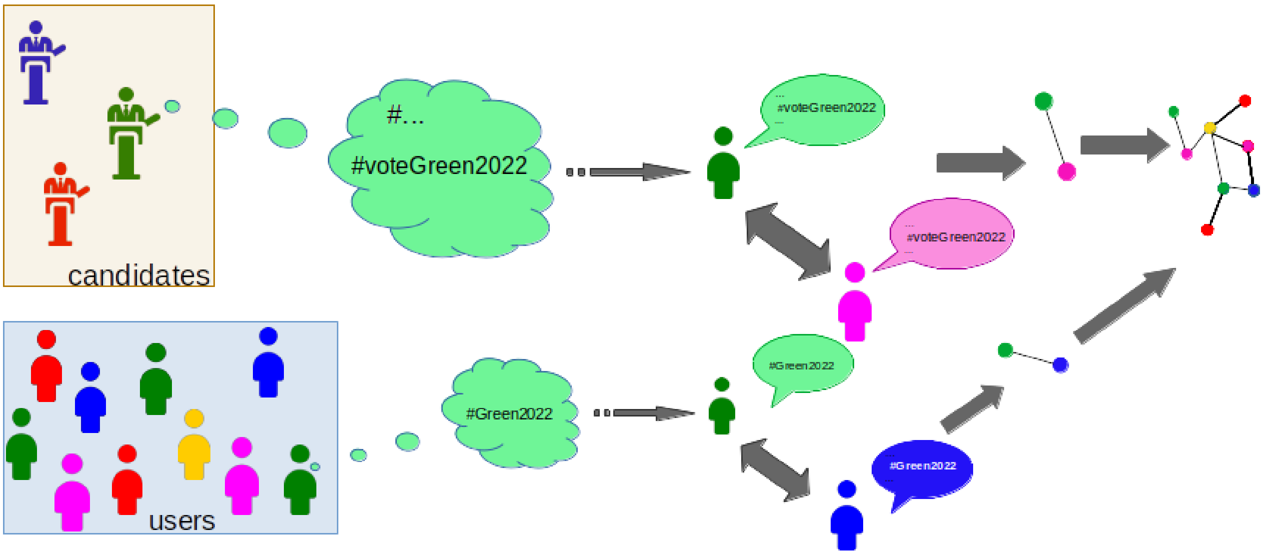

Figure 1. Creation of an endorsement network for a given candidate based on retweeting tweets that contain a given hashtag; the top of the figure represents this process based on candidates' official accounts, while the bottom includes any hashtag created by the users themselves.

among the green and the magenta users. Similarly, a network is created among these and other users (top right side of the figure), for endorsers of given hashtags that promote a certain candidate. The bottom part of the figure shows a similar process, but now the hashtags are created by the users themselves (not by the candidates), who may include their own hashtags in their tweets. When these get retweeted, a link is formed between retweeting users, as depicted in the bottom part of the figure. This way, further edges are inserted in the endorsement network of a candidate.

3 Network analysis

3.1 Endorsement network characteristics

The structural characterisation of networks is helpful in recognising patterns of interest, such as how close two vertices are, what the maximum among the shortest distances between any two vertices is, etc. One question that may be answered this way is whether such patterns unveil non-obvious relationships between vertices, e.g., showing their importance. Further, metrics such as the number of vertices and edges can be used to compare networks.

These measures are meant to characterise a graph as a whole. However, there are other measures that relate to individual vertices. Degree (or degree centrality) of a vertex i corresponds to the number of direct connections i has. Since here we deal with hundreds of thousands of vertices, we discuss degree centrality as an aggregated measure, i.e., we present mean and maximum values for each network.

In the discussions ahead, we show the number of vertices, the number of edges, and the mean and maximum degrees of vertices for the networks of each candidate. The number of vertices indicates the amount of endorsers using at least one of the hashtags of the corresponding candidate. Depending on the election year, these figures can be found in tables 2 and 3. For the sake of illustration, let us take a closer look at the line referring to Lula's network in October (Table 2). Those values reveal that roughly 300,000 endorsers were responsible for around 1,000,000 retweets, providing an insight into the reach of Lula's campaign. The degree values, i.e., the number of connections in the networks, show that the most connected of the endorsers was linked to 83,615 others; on average, endorsers were connected to 6.68 others.

Activity during the 2022 election

Regarding the 2022 election (see Table 2), we discuss the networks of four candidates. However, recall that just two of them were still running in October.

Simone's networks experienced a significant change in September, characterised by a substantial reduction in the number of vertices and edges. This change in network size is a pivotal marker of a shift in her campaign's dynamics. It's crucial to emphasise that despite this decrease, there is a discernible surge in user activity, as demonstrated by the increase in mean degree from 3.37 to 5.84, indicating a more interconnected network of endorsers.

The endorsement networks from Ciro in the two time periods vary only marginally, showing the slightest changes in all networks. We observe a slight increase in vertices and edges from August to September. Moreover, the degree values indicate a strong interaction between Ciro's endorsers,

Table 2. Main characteristics of the endorsement networks in 2022 (candidates in alphabetical order).

Candidate	Period	Vertices	Edges	Degree	
				Mean	Maximum
Bolsonaro	Aug	81 559	478 475	11.73	13 856
	Sep	127 998	893 159	13.96	40 699
	Oct	175 096	976 040	11.15	60 973
Ciro	Aug	9 014	72 193	16.02	4 737
	Sep	10 424	77 314	14.83	3 115
Lula	Aug	216 960	812 161	7.49	86 440
	Sep	246 745	1 034 881	8.39	60 903
	Oct	300 619	1 004 031	6.68	83 615
Simone	Aug	4 707	7 921	3.37	1 428
	Sep	1 615	4 719	5.84	624

although it decreases slightly from August to September. In fact, his networks have the most substantial interaction (i.e., highest mean degree) compared to all other candidates.

For Bolsonaro's and Lula's networks, we observe an increase in the number of vertices throughout the entire period, with more retweets containing Lula's hashtags. The number of edges shows a significant increase for Bolsonaro, particularly from August to September. It also rises in the same period for Lula, with a slight decline from September to October. The number of edges in September and October are comparable for both candidates. Given that Lula had a significantly larger number of endorsers in the same period, this suggests a much higher level of interaction among Bolsonaro's endorsers. This trend is also evident in the mean degree values. While the mean degree of Bolsonaro's networks fluctuates between 11.15 and 13.96, the mean degree of Lula's networks ranges from 6.68 to 8.39. Both cases show an increase from August to September, followed by a decrease in October when the mean degree was the lowest for both candidates.

For general visualisation purposes, we present here two network instances from September: the smallest one in terms of number of edges and vertices, which related to Simone Tebet (see Figure 2), as well as a network related to Lula (see Figure 3), which is the second largest in number of vertices, but has the largest number of edges. Due to their sizes, the visualisation of these large networks is of limited use when drawing quantitative conclusions. In this regard, we need to recur to the information given in Table 2 (and its equivalent for 2018).

In both figures, we take the value of the mean degree (see Table 2) and use it as a threshold to determine the vertices colors. Vertices whose degree is above the threshold are depicted in orange, otherwise in blue. For the edges, those that are connecting vertices whose mean degree is twice the threshold are depicted in brown; otherwise, they are depicted in very light colors. We do this to put the focus on the most important edges, thus making the picture less crowded. In short, the idea of the figures is to focus on vertices and edges that are associated with high degree values. We remark that, since the mean value differs for each candidate and period, the color code in a figure is specific to the network depicted there.

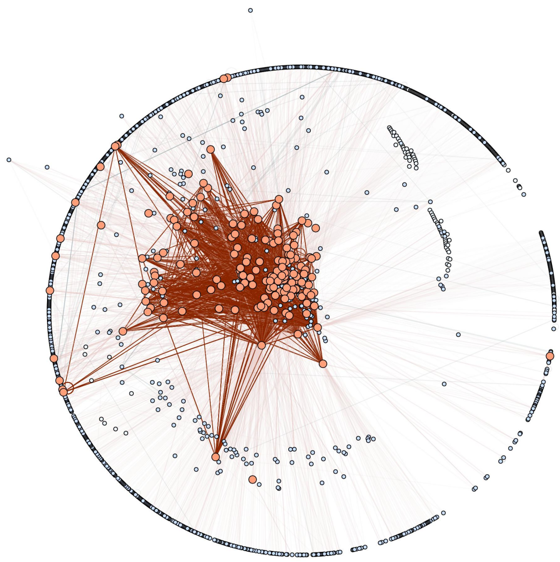

Figure 2. Endorsement Network of Simone Tebet in September 2022; blue color refers to endorsers with smaller degrees or edges connecting these vertices; brown refers to endorsers with higher degrees.

It is also important to remember that the networks were built from the tweets using the set of hashtags we defined previously. Thus, the observations above are specific to the network of users related to these hashtags; not all candidate's endorsers are active on Twitter. Obviously, the observations do not extrapolate to candidates endorsers on other social media or in general. Kreiss and colleagues *apud* Brito *et al.* [2019] stated that "researchers should refrain from automatically generalizing the results of single-platform studies to social media as a whole".

Activity during the 2018 election

As previously mentioned, we have also gathered data regarding the two candidates who participated in the 2018 run-off to compare the structural properties to those related to the 2022 election. While the candidates were not identical in the

Table 3. Main characteristics of the endorsement networks in 2018.

Candidate	Period	Vertices	Edges	Degree	
				Mean	Maximum
Bolsonaro	Aug	103 677	357 869	6.90	13 425
	Sep	69 057	344 293	9.97	19 833
	Oct	82 536	531 453	12.88	27 543
Haddad	Aug	46 883	155 712	6.64	19 675
	Sep	63 032	276 651	8.78	18 408
	Oct	168 445	565 575	6.72	48 933

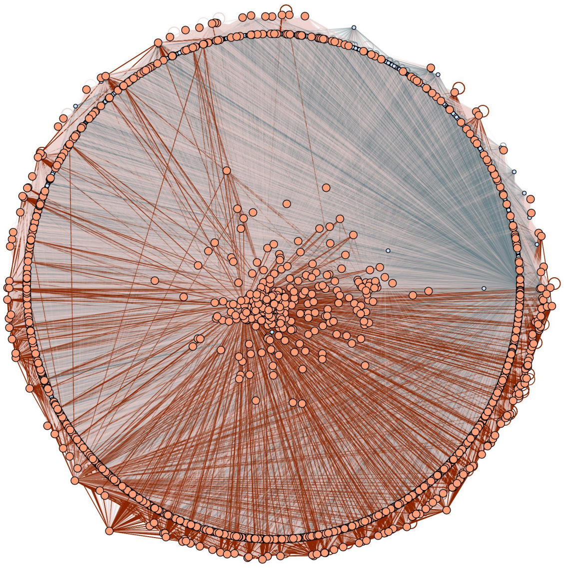

Figure 3. Endorsement Network of Lula in September 2022; blue color refers to endorsers with smaller degrees or edges connecting these vertices; brown refers to endorsers with higher degrees.

2018 and 2022 run-offs, the two opposing political fronts remained the same, making the comparison reasonable. The corresponding results are shown in Table 3.

For the two fronts present in the 2018 run-off, we observe that the number of vertices for Bolsonaro endorsers decreases during the whole period, with a sharper decrease from August to September, followed by a slight increase in October. In other words, the number of users (re)tweeting using hashtags referring to him decreases in the period (from 103 677 to 82 536, according to Table 3).

The number of Haddad's endorsers is lower than Bolsonaro's in August and September but sharply increases each month. This results in a network size of 168 445 endorsers for October, much larger than the number of Bolsonaro endorsers in the same period. The number of edges in Bolsonaro's networks is similar in August and September, while the number of vertices decreases. This indicates more interaction between those endorsers in September, i.e., more retweets. In October, when the number of vertices increases, the number of edges also increases. Besides that, the number of edges is higher in October than in August, although the number of vertices is lower. This also indicates increasing user interaction during the analysis period, especially in the last month. Regarding the number of edges in Haddad's networks, we also observe an increase, but less pronounced compared to the increase in the number of vertices, especially from September to October. This indicates much less interaction between users who promote hashtags related to him.

The observations above regarding the interaction of users can also be observed by looking at the mean degree values for Bolsonaro's and Haddad's networks. We see a clear increase in the mean degree of Bolsonaro's networks, indicating that users are interacting more and, consequently, the corresponding vertices have more connections on average. For Haddad's networks, we see an increase in the mean degree from August to September, followed by a decrease in October, falling to roughly the same value as in August.

3.2 Comparison of the two election cycles

The following facts are salient when comparing the structure of the endorsement networks for 2022 and 2018, focusing on the two candidates who participated in the run-off. While in 2018 the two candidates have something between 50 and 80 thousand endorsers (again, considering the set of hashtags used), in 2022 these values roughly doubled. Regarding the number of retweets, there is a steep increase (see columns labelled Edges in tables 2 and 3). The change in the degree values shows that the more active endorsers (last columns in both tables) retweeted much more in 2022, even if the values for average degree lie in the same order of magnitude (6–14) in both years.

Highest reach users

As seen in Table 2 (and similarly in Table 3), there is a massive difference between the maximum and mean degree of the endorsers). Thus, it is interesting to look at the top endorsers and compare them throughout candidates and/or years.

We start with the 2018 election, which involves just two candidates. Figure 4 shows the top 25 maximum degrees for each period and candidate in 2018[2]. Here, two things can be highlighted: (i) the steep decrease from the first to the fifth, and (ii) that there was a spike in the degree of the most active endorsers of Haddad towards the end of the campaign (brown curve).

In 2022, two of the four candidates (Simone Tebet and Ciro Gomes) have endorsers with much smaller degrees (2

[2]Henceforth, the plots identify candidates by the first letter of their respective given names.

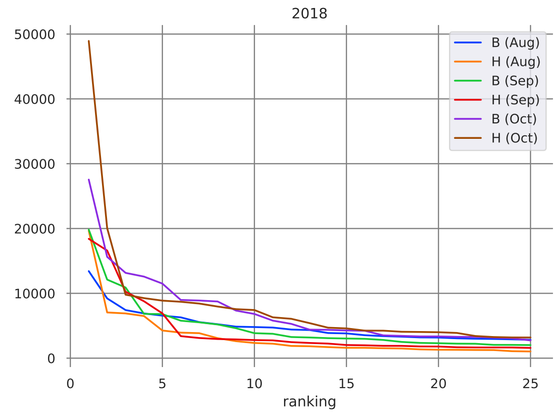

Figure 4. Top 25 vertices, with the highest degrees (sorted in descending order), for each period and candidate in 2018. We compare here the distribution of the degrees among the 25 vertices with the highest degrees, with the vertices' ranking on the x-axis and the degree on the y-axis.

to 3 orders of magnitude smaller). These can be seen in Figure 5, where each plot refers to a given period. It is worth mentioning the huge difference in degrees of the first top endorsers regarding Lula, when compared to Bolsonaro, both in August. However, this changes in September and October, following the development of the presidential election that peaked in a very close run-off between these two candidates.

As Bolsonaro was the only candidate present in both elections, we compare the development of the vertices with the highest degrees in Figure 6. In 2018, the highest-degree vertices had fewer connections to other network vertices than in 2022. In August of 2022, at the beginning of the election campaigns, the users' engagement was at a similar level as in 2018 regarding the number of the top 25 users. However, their engagement in September and October of 2022 was much higher than in 2018.

Generally, the distributions of the 25 highest-degree vertices are similar in both election years. Nevertheless, there are exceptions in the differences in the highest degrees of some candidates, such as in October 2018 (27 543 and 48 933 for Bolsonaro's and Haddad's networks, respectively) and in August 2022 (13 856 and 86 440 for Bolsonaro's and Lula's networks respectively). Although in these cases the networks of Haddad and Lula include the endorsers with the highest degrees, it is worth mentioning that, as seen in the previous section, in Bolsonaro's networks, there is a higher general interconnection.

4 Occurrences of social bots

Online social networks, particularly Twitter, have become key battlegrounds for shaping public opinion and disseminating information during political campaigns [Santana and Vanin, 2020; Chen *et al*., 2022]. On Twitter, human and social bot accounts coexist and a substantial share of accounts are run by social bots (or bots) — up to 15%, Varol *et al*. [2017]. Ferrara *et al*. [2016] (p. 96) describe these social bot accounts as "computer algorithm[s] that automatically produces content and interacts with humans on social media, trying to emulate and possibly alter their behaviour". Savvopoulos *et al*. [2020] have analysed the role of automated chatting in bots' activities, concluding that advanced communications skills contribute to more message interactions between social bots and other accounts. Also, conversational social bots are present in the Twittersphere, thus potentially having impact on political discussion and opinion formulation. In the context of the 2016 U.S. Presidential election, for instance, Bessi and Ferrara [2016] demonstrated that social media bots can negatively affect democratic political discussion rather than improve it, which in turn can potentially alter public opinion and endanger the integrity of the election.

To examine the presence of social bots in our networks, we check the bot-like behaviour of a subset containing the 100 accounts with the highest degrees in each period for each candidate. A widely-used tool to detect bots on Twitter is the Botometer[3] (Yang *et al*. [2022]). The Botometer estimates the likelihood of Twitter accounts acting as bot- or human-like accounts with values ranging from 0 (human) to 5 (bot); however, there is no exact threshold from which an account can be determined as a bot or not (Ferrara *et al*. [2016]). We, therefore, display the probability density plot for the Botometer scores for each network of the 100 accounts with the highest degrees (see Figures 7 and 8).

Figure 7 shows the distribution of Botometer scores for Jair Bolsonaro and Fernando Haddad for each period in 2018. The distributions follow the same pattern, with accounts most likely having a Botometer score of around 1 (bell-curve shaped), and a small proportion of the users are classified as 2.5 or higher. In previous research Broniatowski *et al*. [2018], makes the point that scores of 4 or higher points to an account most likely being a bot. Using this guideline, for our both candidates, only five or fewer accounts are likely to be a bot.

Analogous to Figure 7, Figure 8 shows the distribution of the Botometer scores for the 100 highest-degree accounts for each period and candidate in 2022. Although the distributions' curves are similar to those in 2018, they shift towards less or more bot-like accounts. For Lula and Bolsonaro, most accounts have scored around 1.2; five accounts are classified as bots (Botometer score 4 or higher). Significantly, for Simone and Ciro's cases, Botometer scores are notably lower, around 0.8 or less, in stark contrast to the high bot-like activity in Lula and Bolsonaro's accounts. Furthermore, none of Simone accounts are likely to be a bot. This stark contrast underscores the diversity of online strategies employed by these candidates. In summary, their 100 highest-degree accounts demonstrate a markedly lower level of bot-like activity than those in the Lula and Bolsonaro networks.

No visible anomalies were found in detecting social bots among the 100 highest-degree accounts of the networks. However, in 2022, it was observed that the candidates who made it to the run-off, namely Lula and Bolsonaro, had a higher overall bot-like activity in their networks, as compared to those who dropped out after the first election round. It is important to note that Botometer's scores are affected by a wide range of parameters, including the user activity on Twitter. Therefore, the accounts' scores can vary. Although Botometer can provide valuable insights and flag suspicious

[3] For a detailed description of the application of the Botometer, see Yang *et al*. [2022].

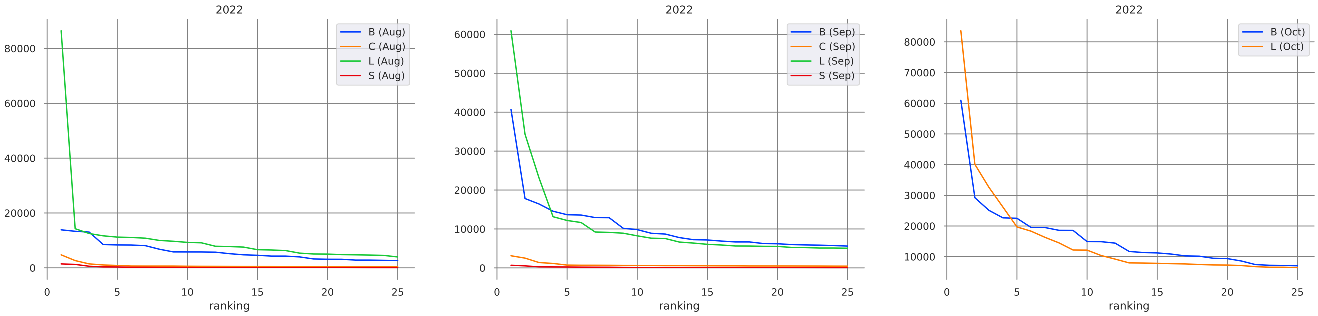

Figure 5. Top 25 vertices, with the highest degrees (sorted in descending order), for each period and candidate in 2022. Similar to Figure 4, we observe here the development of the rankings over the three time points.

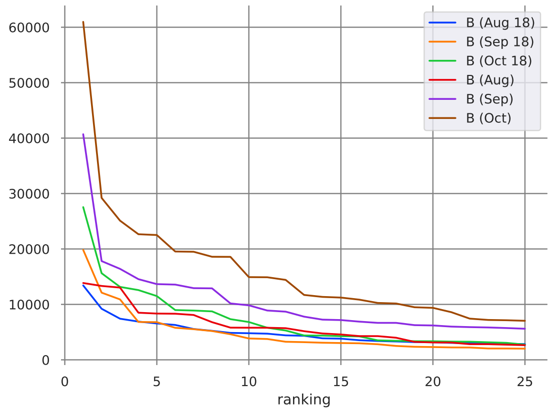

Figure 6. Top 25 vertices, with the highest degrees (sorted in descending order) for Bolsonaro in each period in 2018 and 2022.

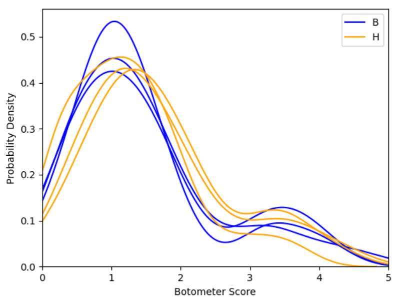

Figure 7. Probability density function of the Botometer scores. For each period and candidate in 2018, we displayed the distribution of the Botometer score for the 100 users with the highest degree in the network.

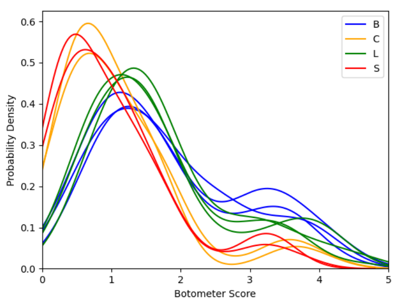

Figure 8. Probability density function of the Botometer scores. For each period and candidate in 2022, we displayed the distribution of the Botometer score for the 100 users with the highest degree in the network.

accounts, its scores should be taken carefully. Social bots constantly evolve; sophisticated bots can mimic human behaviour to evade detection. Additionally, Botometer relies on publicly available data and algorithms that may have limitations and biases.

5 Concluding remarks

Twitter as a social media platform has become an essential tool for sharing opinions, especially during political elections. Twitter can be seen as a monitor for endorsement in the next elections. For instance, in 2024, there will be elections for mayors and city councils in Brazil. More importantly, in 2026, elections for president are scheduled, where there is a high chance that the two polarised fronts mentioned here will compete against each other, very likely with the participation of candidates drawn from those same fronts that have participated in the run-offs both in 2018 and 2022.

In our study, we systematically collected retweets, containing hashtags that show endorsement for the main candidates in the 2022 and 2018 Brazilian presidential elections. We constructed a network where users are linked by a retweet, thus signaling a common stance and endorsement. We have presented the structural analysis of such networks, i.e., their size and degree centrality characteristics, highlighting similarities and differences in the candidates' endorsement networks.

Obviously, such analysis has to be framed in its context, namely, our selection of hashtags (as justified in the text) and the use of Twitter (due to the retweet functionality). As mentioned in Section 3, results from one particular social media cannot be generalised (Kreiss and colleagues *apud* Brito *et al.* [2019]). However, they form general trends that can be used either by future campaign managers, or by the media (when monitoring activity in Twitter/X), or by the voters (when engaging in a candidate's network of endorsement).

Appendix: Data collection

Selecting hashtags

The hashtags are the keywords for the data collection on Twitter. They are used to filter and organise the data. For each candidate and for each election year, we generate a unique set of hashtags. These sets contain hashtags that

Table 4. Most popular hashtags for each candidate and period (in 2022).

Candidate	Period	Most popular hashtag	Count	Percentage
Ciro	Aug	#PrefiroCiro	60631	32.96%
Ciro	Sep	#PrefiroCiro	72168	39.41%
Simone	Aug	#SimonenoJN	5475	27.67%
Simone	Sep	#SimonePresidente15	9643	59.89%
Bolsonaro	Aug	#BolsonaroReeleitoEm2022	89266	13.18%
Bolsonaro	Sep	#BolsonaroNoPrimeiroTurno	349178	30.29%
Bolsonaro	Oct	#Bolsonaro22	644249	40.52%
Lula	Aug	#LulanoJN	563548	35.50%
Lula	Sep	#BrasilDaEsperanca	446140	18.14%
Lula	Oct	#BrasilDaEsperanca	231149	23.66%

indicate a positive association to the candidate in order to endorse the candidate's election, e.g., for Bolsonaro *#BolsonaroReeleito2022* (Eng.: Bolsonaro reelected), for Lula *#LulaNoPrimeiroTurno* (Eng.: Lula in the first round), for Simone *#SimonePresidente* and for Ciro *#PrefiroCiro* (Eng.: I prefer Ciro).

We identified other relevant hashtags by monitoring each candidate's official account since the official start of each official candidature and extracted the hashtags used in their tweets. Only those hashtags that were specific to the election context were selected, e.g., those supporting a candidate by engaging with a hashtag *#Lula13MeuVotoÉseu* (Eng.: Lula13 my vote is yours). One difficulty we faced was that some official accounts of the candidates did not include many hashtags in their tweets. This was the case, e.g., for Bolsonaro's official account, where virtually no hashtags were used in 2022.

Additionally, we also considered hashtags that endorsed a candidate but were not started by the official candidate's account, but rather those hashtags were created by endorsers.

Together, those sets of hashtag have sizes that range from 15 to 43 hashtags, accounting for around 20 million tweets.

The time period of interest, for both election years, spanned three months, divided into (roughly) monthly periods representing the *beginning, middle* and *end* of the campaign. The starting date for the last period (*end*) was the day of the first election round (see Table 1). In 2022, our focus was on four candidates, as mentioned, but only Lula da Silva and Jair Bolsonaro took part in the run-off. For Simone Tebet and Ciro Gomes, the data collection was performed until the candidates dropped out in the first round in 2022. Throughout the collection process, we relied on the same hashtag set per candidate and election year, despite the division into three time periods. Our analysis is based on a data set for each candidate, year and time period.

Gathering data from the Twitter API

For each presidential candidate, the collection of retweets was divided into three distinct, but consecutive, time periods. The process of data collection was identical for the two election years (2018, 2022); however, different sets of hashtags were used in each year.

For the data collection, we accessed the available data through the Twitter's application programming interface (*Twitter API v2*). With the *academic research access* for Twitter, it was feasible to search for historical public data on the Twitter platform, which includes the retweet activity from users in the past. Accessing the Twitter API, we gathered data on uncommented retweets that included at least one of the identified hashtags within a specific time period.

The Twitter API then returned the information about the retweets, such as unique identifier, user ID of the retweeting user (retweet author), the user ID of the retweeted user (original author), the text and the timestamp of the tweet. The IDs of the authors and the retweeted users for each candidate and time period build the foundation for our endorsement networks: every pair of users defines an edge in our networks.

Details about the hashtags

For the sake of illustration, we show data about the most popular hashtags (i.e., hashtags that occurred the most in the retrieved tweets) used in each period in the 2022 election, for each candidate. This can be seen in Table 4.

For Ciro Gomes, the most popular hashtag used in the August period was the *#PrefiroCiro* hashtag, which occurred 60631 times. Moreover, considering this period, the most popular day of use of this hashtag was August 25 – the day after Ciro Gomes' interview with the "Jornal Nacional" (JN)[4]. In the following period (Sep.), the most popular hashtag was again *#PrefiroCiro*, this time occurring 72168 times. The most popular day associated with this hashtag was September 1 – the day CNN interviewed Ciro Gomes.

As for Simone Tebet, the most popular hashtag in August was *#SimonenoJN*, which occurred 5475 times. The peak of its use occurred on August 27 (one day after Tebet's interview in the Jornal Nacional). In September, Simone Tebet's most popular hashtag was *#SimonePresidente15*, which occurred 9643 times; this hashtag was most used on September 12.

For Bolsonaro, we found that in August, *#BolsonaroReeleitoEm2022* was the most popular hashtag, occurring 89266 times (the most popular day for this hashtag was August 24, the day after Bolsonaro's interview with Jornal Nacional). For the second period, *#BolsonaroNoPrimeiroTurno* (Eng.: Bolsonaro in the first round) was the most popular

[4]Brazilian prime time news broadcast on open television.

hashtag (349178 times). In this period, such a hashtag was most used on October 1 – the eve of the first round of the 2022 presidential election. For Bolsonaro, considering the period between the first and the second round of the presidential election, the most popular hashtag was *#Bolsonaro22*, occurring 644249 times; the most popular day was the eve of the run-off.

Regarding Lula, we found that in the August period, *#LulanoJN* was the most popular hashtag, occurring 563548 times, with the most popular day being August 26 – the day after Lula's interview with the Jornal Nacional. For September, *#BrasilDaEsperanca* (Eng.: Brazilian Hope, associated with Lula) was the most popular hashtag, used 446140 times by his endorsers, as well as by his official campaigning team. This hashtag had the most widespread use on September 27 – a day after the results published by a polling institute (Ipec) on voting intentions were released. For the last period (i.e., between the first and the second round of the presidential election), the most popular hashtag was also *#BrasilDaEsperanca*, which occurred 231149 times and peaked on October 26.

Declarations

Acknowledgements

The authors would like to thank people involved in early stages of this research: Prof. Marcelo de Souza and Eduardo Stuani (undergraduate student) for discussions, extensive work on data collection and interpretation of the results. Marcelo also contributed to writing and editing of a first draft of the manuscript.

Funding

Alejandro Dinkelberg was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 802421). Ana Bazzan thanks CNPq (grant no. 304932/2021-3) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil) - Finance Code 001.

Authors' Contributions

The authors confirm contribution to the article as follows: All authors contributed to the study's conception and design, data collection, analysis of results and writing and editing the draft of the manuscript. All authors have approved the manuscript's final version.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Twitter restricts publishing individualised data with personally identifiable information. We are able to provide aggregated data or networks about the users upon request.

References

Badawy, A., Ferrara, E., and Lerman, K. (2018). Analyzing the digital traces of political manipulation: The 2016 russian interference twitter campaign. In *2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, pages 258–265. IEEE. DOI: 10.1109/ASONAM.2018.8508646.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., and Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10):1531–1542. DOI: 10.1177/0956797615594620.

Bessi, A. and Ferrara, E. (2016). Social bots distort the 2016 us presidential election online discussion. *First monday*, 21(11-7). Available at:`https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2982233`.

Brito, K., Paula, N., Fernandes, M., and Meira, S. (2019). Social media and presidential campaigns – preliminary results of the 2018 brazilian presidential election. page 332–341, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3325112.3325252.

Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., Quinn, S. C., and Dredze, M. (2018). Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. *American Journal of Public Health*, 108(10):1378–1384. DOI: 10.2105/ajph.2018.304567.

Chen, L., Chen, J., and Xia, C. (2022). Social network behavior and public opinion manipulation. *Journal of Information Security and Applications*, 64:103060. DOI: https://doi.org/10.1016/j.jisa.2021.103060.

Chen, T. H. Y., Salloum, A., Gronow, A., Ylä-Anttila, T., and Kivelä, M. (2021). Polarization of climate politics results from partisan sorting: Evidence from Finnish Twittersphere. *Global Environmental Change*, 71:102348. DOI: 10.1016/j.gloenvcha.2021.102348.

de Freitas Melo, P., Vieira, C. C., Garimella, K., de Melo, P. O. S. V., and Benevenuto, F. (2019). Can WhatsApp counter misinformation by limiting message forwarding?. DOI: 10.48550/arXiv.1909.08740.

Ferrara, E., Varol, O., Davis, C., Menczer, F., and Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7):96–104. DOI: 10.1145/2818717.

Garimella, K., Morales, G. D. F., Gionis, A., and Mathioudakis, M. (2018). Quantifying controversy on social media. *ACM Transactions on Social Computing*, 1(1):1–27. DOI: 10.1145/3140565.

Metaxas, P., Mustafaraj, E., Wong, K., Zeng, L., O'Keefe, M., and Finn, S. (2021). What do retweets indicate? results from user survey and meta-review of research. *Proceedings of the International AAAI Conference on Web and Social Media*, 9(1):658–661. DOI: 10.1609/icwsm.v9i1.14661.

Resende, G., Melo, P., Sousa, H., Messias, J., Vasconcelos, M., Almeida, J., and Benevenuto, F. (2019). (Mis)Information dissemination in WhatsApp: Gathering, analyzing and countermeasures. In *The World Wide Web Conference*, WWW '19, page 818–828, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3308558.3313688.

Santana, B. S. and Vanin, A. A. (2020). Detecting group beliefs related to 2018's brazilian elections in tweets A combined study on modeling topics and sentiment analysis. *CoRR*, abs/2006.00490. DOI: 10.48550/arXiv.2006.00490.

Savvopoulos, A., Vikatos, P., and Benevenuto, F. (2020). Socialbots' first words: Can automatic chatting improve influence in Twitter? In *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '18, page 190–193. IEEE Press. DOI: 10.1109/ASONAM.2018.8508786.

Sayyadiharikandeh, M., Varol, O., Yang, K.-C., Flammini, A., and Menczer, F. (2020). Detection of novel social bots by ensembles of specialized classifiers. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, CIKM '20, page 2725–2732, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3340531.3412698.

Soares, F. B., Recuero, R., and Zago, G. (2019). Asymmetric polarization on twitter and the 2018 brazilian presidential elections. In *Proceedings of the 10th International Conference on Social Media and Society*, SMSociety '19, page 67–76, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3328529.3328546.

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., and Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. DOI: 10.48550/ARXIV.1703.03107.

Yang, K.-C., Ferrara, E., and Menczer, F. (2022). Botometer 101: social bot practicum for computational social scientists. *Journal of Computational Social Science*, 5(2):1511–1528. DOI: 10.1007/s42001-022-00177-5.