



# Political polarization on Twitter during the COVID-19 pandemic: a case study in Brazil

Pedro Brum<sup>1</sup> · Matheus Cândido Teixeira<sup>1</sup> · Renato Vimieiro<sup>1</sup> · Eric Araújo<sup>2</sup> · Wagner Meira Jr<sup>1</sup> · Gisele Lobo Pappa<sup>1</sup>

Received: 21 February 2022 / Revised: 19 July 2022 / Accepted: 31 July 2022 / Published online: 23 September 2022  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

## Abstract

The debate over the COVID-19 pandemic is constantly trending at online conversations since its beginning in 2019. The discussions in many social media platforms is related not only to health aspects of the disease, but also public policies and non-pharmacological measures to mitigate the spreading of the virus and propose alternative treatments. Divergent opinions regarding these measures are leading to heated discussions and polarization. Particularly in highly politically polarized countries, users tend to be divided in those in-favor or against government policies. In this work we present a computational method to analyze Twitter data and: (i) identify users with a high probability of being bots using only COVID-19 related messages; (ii) quantify the political polarization of the Brazilian general public in the context of the COVID-19 pandemic; (iii) analyze how bots tweet and affect political polarization. We collected over 100 million tweets from 26 April 2020 to 3 January 2021, and observed in general a highly polarized population (with polarization index varying from 0.57 to 0.86), which focuses on very different topics of discussions over the most polarized weeks—but all related to government and health-related events.

**Keywords** Political polarization · Twitter · Covid-19 · Bots

---

Pedro Brum and Matheus Cândido Teixeira have contributed equally to this work.

✉ Gisele Lobo Pappa  
glpappa@dcc.ufmg.br

Pedro Brum  
pedrobrum@dcc.ufmg.br

Matheus Cândido Teixeira  
matheus.candido@dcc.ufmg.br

Renato Vimieiro  
rvimieiro@dcc.ufmg.br

Eric Araújo  
eric@ufla.br

Wagner Meira Jr  
meira@dcc.ufmg.br

<sup>1</sup> Computer Science Department, Universidade Federal de Minas Gerais, Av. Pres. Antônio Carlos, 6627, Belo Horizonte, MG 31270-901, Brazil

<sup>2</sup> Computer Science Department, Universidade Federal de Lavras, Aqueanta Sol, Lavras, MG 37200-900, Brazil

## 1 Introduction

Social media is part of society, and has proven its value during many global events and catastrophes even before the COVID-19 pandemic started in 2020. The amount of opinions, statements and news spread and shared in social media turned it into one of the main sources of information regarding the outcomes and results of public policies about the coronavirus pandemic (Gallotti et al. 2020). It is known that opinions spread out across social media platforms through users' connections. Their reach and impact depends on a few aspects as the personality traits of those who receive the information and the strength of the connections between people in the network (Araújo et al. 2020). The opinions spread about the current pandemic can help people understand the disease and behave well regarding the proper health protocols recommended by reliable sources. On the other hand, the spread of messages in social media can lead to a massive misinformation based on conspiracy theories, as well as unfounded panic (Bhattacharya et al. 2021; Depoux et al. 2020; Cinelli et al. 2020; de Mello Araújo et al. 2017).

Although social media such as Twitter has been previously used to understand and predict the spread of other endemic diseases (Albinati et al. 2017), the COVID-19

pandemic brought a strong political component to it. The political polarization in many democracies contributed to a heated debate by non-experts over what measures are acceptable or not to contain the virus spread and treat infected people. The mismatch between health agencies, scientist and state leaders was brought to social media having an impact on people's opinions over the use of masks, social distancing, lockdown, and even about the guidelines for preventing, treating or curing from the virus (Graham et al. 2020). The alignment with specific groups turned the debate into a wrestling between scientific discoveries and leaders' opinions based on personal experience or unreliable sources. In this war, political instability can be considered as a fertile soil to the spread of false or misleading information (Lazer et al. 2018), preventing the effective adoption of public health recommendations (Waszak et al. 2018).

More specifically, since February 2020, days after the first cases of COVID-19 were reported in Brazil, president Jair Bolsonaro has been spreading misinformation online, defending the use of hydroxychloroquine and being against measures such as the use of mask or social distancing (Ball and Maxmen 2020).

This paper aims to introduce a computational methodology to evaluate and measure the political polarization of users on Twitter during the COVID-19 pandemic in Brazil, taking into consideration the relation between the country's political scenario, the evolution of the pandemic and people's polarized opinions. We also take into consideration the Brazilian president's political positioning and how it impacted the polarization of users, as well as the presence of bots in the topics discussed in the network, mostly aligned with the Bolsonaro's opinions.

Many papers have previously dealt with political analysis and polarization on Twitter (Kušen and Strembeck 2018; Moreira et al. 2020), even during the pandemic (Jiang et al. 2020). In this work we perform a preliminary bot detection approach based solely on tweets about the pandemic before doing text analysis. It is an important step in our methodology as the number of users made the collection of all their published messages infeasible. It is also relevant to evaluate the role of the bots in conversations, as bots, activists and fake accounts are often used to influence political discussions by falsely fueling some political positions and creating the perception of "false consensus" (Tucker et al. 2018). After detecting bots, we performed a set of analysis to identify pro-government and anti-government users, and defined a polarization index to measure users polarization. Based on these results, we also capture the differences between what is discussed by both polarized bots and polarized non-bot accounts.

We collected a corpus containing over 100 million tweets from over 7 million users posted from April 2020 to January 2021. The results show that users' polarization on Twitter

reflects the current political scenario in Brazil, with a polarity index varying from 0.59 to 0.86 during this period (in a scale from 0 to 1). An analysis of the content of the tweets posted by both pro-government and anti-government users showed they differ significantly in the content of the discussions for most weeks with a higher polarization index. The analysis of the content produced by bots also showed a different pattern in the tweets, with pro-government bots being more generic in their posts. We also observed the presence of political activists in the discussions, real users with similar behavior to the bots also intending to influence the political discussions online.

## 2 Related works

The vast majority of the literature in social media and the COVID-19 pandemic analyze how socioeconomic and social media data relates to public health and the political scenario of different countries worldwide (Gallotti et al. 2020; Graham et al. 2020; Charron et al. 2020). In general, these works are interested in how users react towards the pandemic and government policies to deal with it. In these analysis, users' behavior is characterized using their activity on Twitter, which includes messages, URLs and hashtags they share as well as their connections.

Researchers also perceived a correlation between social media activity and the number of cases and deaths of coronavirus in different locations (Charron et al. 2020). To this end, studies monitor Twitter activity and collect messages (i.e. tweets) using words commonly used in the medical discourse about COVID-19. Most of the studies consider English written messages and, consequently, capture the behavior of the English-speaking users. In this study, we deal with messages in Portuguese regarding the COVID-19 pandemic. The most significant work we found that works with Portuguese tweets and performs a similar analysis to ours is the work of Ceron et al. (2021), although their main focus is the spread of fake news.

In this paper, we analyze how Brazilians reacted to public policies implemented by president Bolsonaro's government concerning the COVID-19 pandemic by measuring the political polarity of the Brazilian Twittersphere. Graham et al. (2020), for instance, conducted a similar analysis in Australia. They analyzed the political debate regarding the measures taken by Victoria State Government to control the pandemics during the second wave of the disease in Australia. Jiang et al. (2020) looked at how people reacted to government policies in different US states. They also investigated the political polarization of US users looking at a retweet network, but using clustering methods to define user profiles instead of other labeling schemes (e.g., we used hashtags and user retweets to find out their opinion regarding the government).

As pointed out by Graham et al. (2020), coordinated activity was used to criticize the Australian government. Before that, social bots were already pointed out as sources of misinformation about specific subjects during presidential campaigns on Twitter in order to benefit a politician or a political party (Ferrara et al. 2016).

*Bots:* A social bot is defined as a social media account that is controlled in some degree by an algorithm (Ferrara et al. 2016). These bots evolve over time and present an increasingly sophisticated behavior to avoid account automation detection mechanisms (Cresci et al. 2017).

The bot detection literature is vast (Ferrara 2020), and a bot is usually identified based on their overall posting behavior during a specified period of time. In the past years, the Twitter developer API defined a policy that enforces bots to self-identify themselves in their screen-name or profile<sup>1</sup>. This helps users identify “the good bots” and automatic methods can take advantage of that.

On the other hand, several factors have limited the application of traditional techniques previously developed in the literature. For example, changes on Twitter policies regarding misinformation<sup>2</sup> allow several special actions to be taken regarding the content produced to prevent the spread of misinformation, such as reporting accounts or messages so they can be removed from the platform in case of violation of the rules and policies. These new policies make the process of recovering data from accounts labeled as bots more difficult, and can lead to inaccurate results of traditional methods developed in the literature, since the previous conditions cannot be reproduced. In addition, policies against disinformation and the narrative used to disseminate disinformation depend on cultural and legislative factors in each country.

Ferrara (2020) was one of the first works that tackled the problem of identifying bots tweeting about the pandemic. They use a set of four indicators collected from Twitter accounts, and approximately 300 features extracted from the content posted by users to classify the account as bot or not. We analyze a similar scenario to theirs, but with two main differences: (1) we collect only tweets related to COVID-19, and (2) do not restrict posts from any class of users, that is, regardless of whether they are bots or not. After some unsuccessful experiments in classifying bots using traditional methods due to the restrictions limiting how much of the total content published by a user could be collected (only related to COVID-19), we chose to use Botometer<sup>3</sup> (Shao

et al. 2018; Xu and Sasahara 2021; Sayyadiharikandeh et al. 2020; Yang et al. 2020) – a Twitter bot detector provided as a web-service and popular for bot detection in the literature. Botometer returns a score that reflects the probability of an account being a bot. It generates the scores using a supervised classifier, Random Forest, on approximately 1,000 features extracted from the 200 most recent tweets published by the analyzed user (Davis et al. 2016). Botometer assigns users to one out of four classes: fake\_follower (bots purchased to increase follower counts), self\_declared (bots from botwiki.org), spammer, and others (a miscellaneous of other bots that do not follow in the previous categories.) However, as with any supervised classifier, its quality is highly influenced by the quality of the datasets, which can become obsolete fast. This is because bots are constantly evolving, with a behavior increasingly similar to that of authentic users.

There are not many dedicated works to follow their development and change of performance over time (Ferrara et al. 2014). Among the few works dedicated to optimizing existing methods, Sayyadiharikandeh et al. (2020) propose an ensemble of specialized classifiers to allow Botometer to follow the evolution of bots: each classifier is trained to identify a specific type of bot. The results showed that this method has a greater generalization capacity than the traditional Botometer (v3) and was adopted in the subsequent version of the tool (v4).

A more in-depth characterization of bot behavior is presented in Varol et al. (2017), where the authors observe several difficulties in classifying modern bots: (1) binary classification is increasingly difficult due to the increasingly similarity of bot behavior to genuine users, (2) there are different classes of bots created with different purposes and, in some cases, their difference to humans is not clear, and (3) features based on user metadata and content are more valuable than those related to the relationships that the user has in a social network (friends, network or sentiment-based features).

Aiming to tackle these problems, Rodríguez-Ruiz et al. (2020) propose a one-class classifier able to identify anomalies. The classifier is trained with samples from genuine users and anomalies are classified as bots. Varol et al. (2017) estimates that the bot population is around 9-15% of users, so most users are genuine and therefore, in theory, it is easier to train the classifier with genuine users and detect anomalies. This model also has the advantage of “automatically capturing” the evolution of bots –although it can be affected by changes in the behavior of genuine social network users.

Despite these recent advances in understanding the behavior of bots and in new techniques for detecting them, traditional methods, such as the Botometer, continue to be used in the literature given that they focus on a fixed time window, that is, the studied dataset does not keep evolving

<sup>1</sup> [https://blog.twitter.com/developer/en\\_us/topics/community/2020/twitter\\_developer\\_policy\\_update.html](https://blog.twitter.com/developer/en_us/topics/community/2020/twitter_developer_policy_update.html).

<sup>2</sup> <https://help.twitter.com/en/rules-and-policies/medical-misinformation-policy>.

<sup>3</sup> <https://botometer.osome.iu.edu/>.

over time (Shao et al. 2018; Xu and Sasahara 2021; Sayyadharikandeh et al. 2020; Yang et al. 2020).

**Polarization Analysis:** Investigations on polarization are also subject of complex network studies, as a mean to understand the propagation of sentiments, opinions and behavior through connections between people (Moreira et al. 2020; Prasetya and Murata 2020). Despite the fact that there are many studies in the field of opinion polarization of Online Social Networks (OSNs) users, there is no consensus about a quantitative measure for it (Schmitt 2016). With the purpose of finding users' opinions, Moreira et al. (2020) calculate the polarity of OSNs users using retweets networks, since retweets act mostly like endorsements of an opinion, which seems to be a more meaningful basis for computing the user's polarity. Having the polarity of users, they calculate the polarization index proposed by Morales et al. (2015), which takes into account the probability density distribution of the opinions of individuals to quantify the segregation within a population. One of the main advantages of this approach is that it does not require the network structure to calculate the final polarization, since it is based on the density distribution of polarity values. Obtaining the network structure is an expensive process, and even harder in our context, where tweets are restricted to COVID-19-related messages.

### 3 Data collection

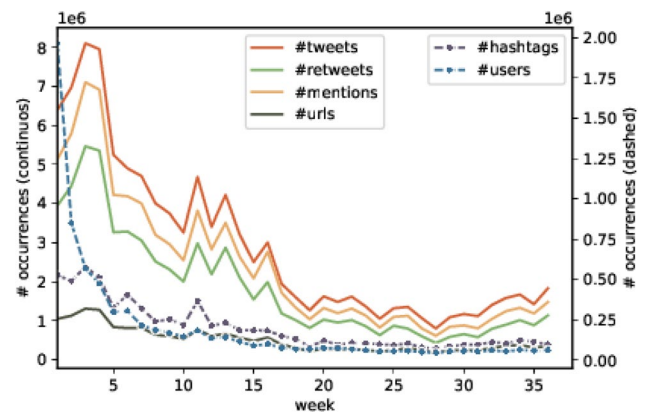
We collected a dataset of Portuguese tweets through the public Twitter Stream API using 13 keywords, in the period from April 2020 to January 2021. The keywords are: “corona”; “covid”; “coronavírus” (*coronavirus*); “covid19”; “quarentena” (*quarantine*); “hidroxicloroquina” (*hydroxychloroquine*); “cloroquina” (*chloroquine*); “confinamento” (*confinement*); “distanciamento social” (*social distancing*); “aglomeração” (*crowding*); “aglomerações” (*agglomerations*); “sars”; and “covid-19”. The language was identified using the Twitter *lang* field.

We chose these terms because they comprised possible treatment (e.g., “hidroxicloroquina”, “cloroquina”), name variations of either the virus or disease (e.g. “corona”, “covid”), and preventive measures (“distanciamento social”). Within this context of COVID-19, we checked which hashtags were concerned with politics. We used this data intersection to perform political polarity analysis, i.e., look at the political position of people while tweeting about COVID.

We collected approximately 7.1 million users and 104 million tweets posted between April 26, 2020, and January 4, 2021, in a total of 36 weeks. Table 1 presents the overall number of tweets, retweets, mentions, URLs and unique users analyzed. Figure 1 shows these same statistics over

**Table 1** Main statistics of the dataset collected

Time period	April 2020 to January 2021
Number of tweets	104,113,713
Number of retweets	66,099,002
Number of hashtags	7,289,188
Number of URLs	18,247,641
Number of mentions	8,6061,269
Number of unique users	7,146,271



**Fig. 1** Main statistics of the tweets collected over the 36 week-period

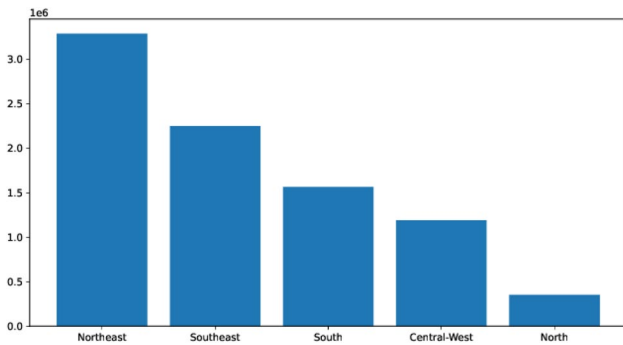
time. This set of tweets covers 66,099,002 (63.49%) retweets, 2,912,273 unique URLs and 389,296 hashtags.

We have also looked at some basic demographics of the users. First, we used a simple gender inference method based on a list of names created by Filho et al. (2015) to infer gender. With this approach, 45% of gender users were identified, 52% being women and 47% men. Next, we looked at user location. From all the tweets collected, 8,637,950 (8.3%) had information about location (*geo* field) in Brazil.

Figure 2 shows the distribution of users in the five main regions of the country: North, Northeast, Central-West, Southeast, and South. Note that the region with the most tweets over time is the Northeast, although the state with the most tweets is São Paulo, in the Southeast region. Although we cannot guarantee the remaining tweets were posted from Brazil, currently Brazil has the 4th largest population of Twitter active users<sup>4</sup>, with no other Portuguese speaking countries appearing in the top-10 ranking.

In order to perform a temporal analysis of polarity, we split the dataset into epidemiological week time slices. The dataset was divided into 36 parts, each corresponding to a period of one epidemiological week. Figure 3a shows the total number of tweets over the studied period. Note that

<sup>4</sup> <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>.



**Fig. 2** Number of geo-tagged tweets per Brazilian geographical regions

the number of tweets has its peak in the third week (10 May 2020 to 16 May 2020), with 8,070,223 tweets, where 5,441,738 (67.43%) are retweets. The first half of the period of analysis, i.e., between the 1st and 16th weeks (from 26 April 2020 to 15 August 2020), concentrates most of the tweets.

We also looked at the number of new cases and deaths from COVID-19 over the studied period to assess whether online conversations were somehow correlated to them. The curves are also depicted in Figures 3a and b, where the y axis in the right shows the number of cases and deaths, respectively. The number of tweets and new cases/deaths started to follow the same trend around the 10th week (June 28th, 2020 and July 4th, 2020). These numbers reached their lowest values in week 28 (November 2nd, 2020 and November 8th, 2020). From this point, the numbers started to increase again. This confirms to a certain extent that the online conversations resound the pandemic situation in the country.

We have calculated the Pearson correlation for the 36 weeks period between the number of tweets and the number of COVID cases, which is 0.44. If we divided the data into two partitions according to data volume, i.e., up to the 10th

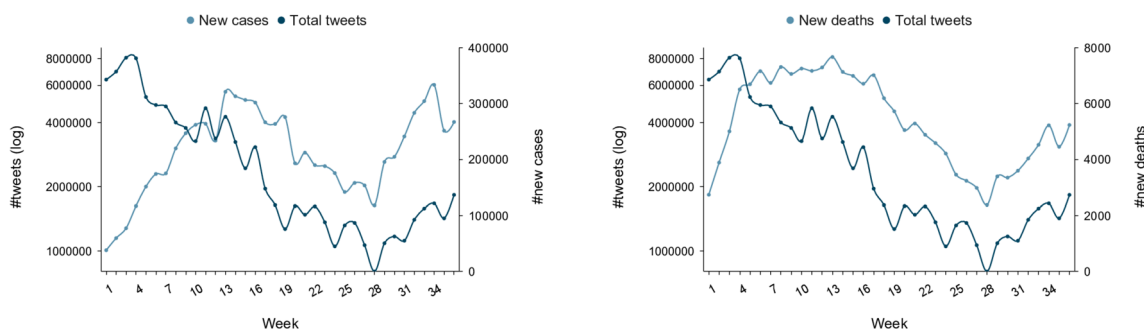
week and from the 11th week onwards, the correlation is then -0.86 and 0.55, respectively. The t-statistic shows that, in all cases, there is a linear correlation between the number of tweets and the number of cases, considering a 0.05 level of confidence.

### 4 Bots detection

Bot literature classifies users according to the degree of automation. For example, users fully controlled by some computer program are called bots, while those partially automated are called cyborgs. The cyborgs are users controlled by a human and a computer alternately (Chu et al. 2010; Varol et al. 2017). In this paper, we do not distinguish between these two groups, both are considered bots.

To find bots in our dataset, we use Botometer (Shao et al. 2018; Xu and Sasahara 2021; Sayyadiharikandeh et al. 2020; Yang et al. 2020). It is a web service that calculates the probability that a given user is a bot. It generates a probability score using a supervised classifier (Random Forest) on approximately 1,000 features extracted from the 200 most recent tweets published by the analyzed user and other account information (Davis et al. 2016). The classifiers are trained through a labeled dataset formed by the results of works that classify bots manually or through the feedback of papers that used the tool to classify bots.

When the User ID is sent to the tool, Botometer returns the complete automation probability (CAP), that indicates the probability that the account is automated. An user with a CAP equal to  $\theta$  and overall score equal to  $\alpha$  can be considered a bot if  $\alpha \geq \theta$  with probability  $\theta$  of certainty. The  $\theta$  parameter can be chosen according to the desired degree of precision (we chose  $\theta = 0.82$ , as these values provide a quantity of approximately 10,000 bots). In addition to returning the CAP value, the Botometer returns the  $S_e$  and  $S_i$  scores, which indicate the probability of the user being



(a) Number of COVID-19 cases (right y axis).

(b) Number of COVID-19 deaths.

**Fig. 3** Number of tweets related to COVID-19 (left y axis) over time in Brazil



**Table 2** Features used to flag a user as a potential bot, followed and ordered by the percentage of users in the dataset that present the indicator

---

Self-declared bots by screen name (0.01%)
Proportion of symbols per tweet is greater than 0.95 (0.01%)
Description has a url pointing to a Github page (0.03%)
Self-declared bots by name (0.04%)
Screen name contains terms such as “bot”, “robot”, “robo”, “conta-reserva” (Portuguese for backup account)
Description contains expressions like “automatically retweet ...” (0.08%)
Proportion of tweets related to COVID-19 is greater than 0.95 (2.61%)
Has predominantly more followers than follows, i.e. $\#friends/(\#friends + \#followers)$ (3.02%)
Number of followers or friends is zero (3.50%)
Proportion of urls per tweet is greater that 0.95 (3.60%)
Proportion of urls per tweet is greater than 0.95 (7.11%)
Age of the account (calculated by $e^{(-0.046 * \text{age in days})}$ ) is greater than 0.5 (8.40%) <sup>a</sup>
Proportion of mentions per tweet is greater than 0.95 (41.40%)
More than 95% of account tweets are retweets (46.07%)
Uses default profile, i.e. not customized (67.95%)
Has no coordinates or places (89.39%)
Unverified account (92.88%)
Unprotected account (93.03%)

---

<sup>a</sup> 0.046 gives a score of 0.25 if the account age is 30 days

a bot based on the textual content of the tweets produced. The difference between  $S_e$  and  $S_t$  is that the former is language dependent – the tool assumes that tweets are in English– while the latter is not.

A major drawback of Botometer, as stated in Sayyadiahari-kandeh et al. (2020), is the constant need to retrain the classifiers and update the training set to keep the service capable of detecting new bots that come up each time. The service is commonly used in works whose focus is to rather analyze bots’ behavior (like ours) than simply detecting them,

since the latter is a difficult task, as bots evolve every day to bypass detection mechanisms and present a more human-like behavior (Cresci et al. 2017).

Botometer has both free (with restrictions to the number of requests) and paid plans. The free web-service plan imposes a restriction on the limit of requests that can be made in a 24-hour period (500 requests per day). Therefore, this work adopts a heuristic to optimize the success of requests receiving a user with a higher probability of being a bot.

---

### Algorithm 1 Bots Detection

---

**Input:** set of tweets  
 $D_0 = \emptyset$   
 $U$  = Organizes data by users, where each user  $u$  in  $U$  has a subset of *tweets*  
 Calculates a set of 18 features for each user in  $U$   
 $R$  = Ranks users according to the 18 features using borda count  
 $S$  = Selects the top-scored users according to a predefined criteria  
**for**  $i = 1$  to  $k$  **do**  
    $U = U - S$   
   Send  $S$  for Botometer to label  
    $L$  = Labeled users returned by Botometer  
   Extracts a set of 115 features for users in  $L$   
   Normalizes data  
   Performs feature selection with mutual information  
    $D_k = D_{k-1} \cup L$   
    $model$  = Classifier trained with  $D_k$   
   Apply  $model$  to  $U$   
    $S = 500$  user with highest confidence of being bots according to the model  
**end for**

---

Algorithm 1 describes the process followed to identify users as bots. Given a set of tweets, we organize these data by users and store it in  $U$ . For each user in  $u$ , we extract a set of 18 simple features selected from the literature – commonly used to flag a user as a bot (Gilani et al. 2017; Stringham et al. 2010; Rodríguez-Ruiz et al. 2020; Ferrara et al. 2016, and listed in Table 2. After extracting these features, users are ranked using the borda count algorithm creating the set  $R$ . At the end of this process, we expect users at the top of the ranking to have a higher probability of being bots.

Following Alg. 1, the top  $N$  users from set  $R$  form set  $S$ , and are sent to Botometer to be classified.  $N$  was set as the number of users with a score in the 90 percentile of the set  $R$  and who have published at least 3 tweets related to COVID-19. Next, Botometer returns a score that indicates how likely users are to be bots. As the tweets are mostly written in Portuguese, the  $S_i$  score is used, as it is language-independent (Yang et al. 2022).

As the end of this process, we obtain the first ground truth dataset  $L$  of users ranked by Botometer. As already mentioned, we use the  $S_i$  score to classify users, and users with scores greater than or equal to 0.82 are marked as bots and those with lower scores are marked as non-bots. These users will be used as an initial dataset to train a classifier.

For all users in  $L$ , we extracted a set of 115 features from both the user accounts and the tweets they published. This may seem like a large number, but it is important to emphasize that many of them reflect different statistics about the same indicator, e.g., mean, median, and standard deviation of text size in tweets published by a user  $u$ . As some features may change over time – such as the number of followers – for numerical features we have a time series - summarized by the mean, standard deviation, first and third quantiles, median, maximum and minimum. For non-numeric features, such as URLs to external sites, boolean features were created to indicate if the event occurred at least once.

The features are pre-processed as follows: (1) we remove features that have constant values for all examples; (2) we normalize the data using min-max, and (3) we perform feature selection using the mutual information measure, which calculates how much the feature  $X_i$  contributes to predict the target variable  $Y$  and generates a score, which indicates the importance of the feature. Features that score lower than the median of the set of scores are discarded. This reduces the dimensionality of the dataset from 115 to 37, i.e., 32% of the features are kept. At this point, we have a dataset  $D_0$ .

Having this labeled dataset, we model the problem as a supervised learning classification problem and train a classifier with  $D_0$ . After that, we start an iterative process, inspired by both a self-training and an active learning approach (Chapelle et al. 2006): we use the trained classifier to score all users  $U$  in the dataset, select the top-500 users – now from  $D_1$ , generated with our classifier trained over

$D_0$  – classify them with Botometer and generate an updated ground truth dataset  $\mathcal{D} = \cup_{i=1..k} \mathcal{D}_i$ , where  $k$  represents the  $k$ -th iteration, and so the cycle repeats. This updated dataset included all users classified by Botometer, spammer or genuine.

In this work, the top-scored users selected to be in  $S$  were defined according to two criterion: their score was in the 90 percentile and they had published at least 3 tweets related to COVID-19. 33,026 users met these criteria and were therefore given as input to Botometer. From these, 9,307 users had their accounts suspended or deleted within the period of collection and the Botometer experiment (which occurred about 8 months after tweets were collected), which serves as an indication of the misuse of the account. As a result, 23,719 top-ranked users were scored by Botometer, where 10.7% (2,548) were classified as bots. Following the Botometer classification scheme, 1,342 bots are from the “others class”, 757 of the “self\_declared class”, 433 of the “fake\_follower” class, and 16 of the “spammer” class.

The effectiveness of this initial method is proven through the results that show that approximately 10% of verified users (23,719) are really bots (2,548). Varol et al. (2017) estimates that around 9-15% of Twitter users are in fact bots and our results agree with this proportion. However, the users analyzed in this paper are only those who posted something related to COVID-19, so the proportion may not really be the same and more studies should be carried out to quantify the participation of bots in issues related to the pandemic.

Since ranking users by indicators extracted from posting behavior is a new method for optimizing bot detection on large datasets, we verified if this method really has any impact on the result: a chi-square test was performed to verify if the users obtained through the ranking by indicators are really significant or if the users were randomly selected it would generate the same result. The contingency table in Table 3 shows how the groups are distributed. The result shows that the chi-square statistic is 122.86 ( $p$ -value  $< 0.00001$ ) and hence significant at  $p = 0.05$ . As such, we conclude that the use of the ranking by indicators is simple and effective to perform to flag potential bots.

With this initial labeled dataset, we compared the performance of five classifiers to train *model*: Ada Boost (AB), Decision Tree (DT), Random Forest (RF), Support

**Table 3**  $\chi^2$ -test measuring the difference of using ranking or testing users at random

	Bots	Genuine	$\Sigma$
Rank	2,548	21,396	25,391
Random	23	1,424	1,447
$\Sigma$	2,571	22,820	52,229

**Table 4** Comparison between classifiers used to identify users as bots

Classifier	F1-score	Precision	Recall	ROC AUC
AB	0.6157	0.7367	0.5337	0.9179
DT	0.5678	0.6131	0.5374	0.7386
KNN	0.5917	0.7442	0.5079	0.8346
LR	0.5551	0.7550	0.4429	0.8877
RF	0.6536	0.8225	0.5521	0.9331
SVC	0.5476	0.7940	0.4208	0.8658

AB Ada Boost DT Decision tree, LR linear regression

KNN k-Nearest neighbours, RF random forest, SVC Support Vector Classification

Vector Classification (SVC) K-Nearest Neighbors (KNN) and Logistic Regression (LR) classifiers (Rodríguez-Ruiz et al. 2020). The hyperparameters of each classifier were chosen through grid search, where the search interval is around the default value assigned by Scikit-Learn. The performance measures evaluated were precision, recall, F1-score and AUC-ROC as they are commonly used in the literature (Rodríguez-Ruiz et al. 2020). Metrics were calculated using a stratified K-fold cross-validation, as our dataset is unbalanced, and the final score is equal to the average score obtained in each fold. The results are shown in Table 4 and, as can be seen, the RF classifier presents the best performance among the classifiers tested and was the classifier selected to compose the process represented in Algorithm 1.

After 20 iterations, we were able to evaluate 10,404 users with Botometer, 7,760 (74.59%) of them classified as bots. In the last but one iteration, the dataset was trained with 34,123 instances, 7,470 (21.9%) bots and 26,653 (78.1%) genuine users. The difference to the complete set regards users labeled by Botometer in the last iteration. The complete experiment took about 68 days and evaluated over 34,000 users<sup>5</sup>.

It is important to mention that the objective of this experiment was to identify bots and verify what type of content was being disseminated by them. The proposed method was used as a way to prioritize users and save resources. The objective was achieved, as 10,404 (30.5%) users were classified as bots, with 62,41% belonging to the ‘others class’, 28.94% to the ‘self\_declared class’, 8.36% to the ‘fake\_follower’ class, and 0.29% to the class spammer.

## 5 Political polarity

Besides identifying bots, it is important to define a polarity measure for political positioning and identify how the polarity score of the Twitter population evolves over time.

<sup>5</sup> We have also evaluated users that had their accounts suspended or deleted.

The definition of polarization used in this work is the one proposed by (Morales et al. 2015). Morales et al. (2015) consider that *a population is perfectly polarized when it is divided into two groups of the same size that share opposite views about a subject*. The overall polarization depends on the individual level. In our work, we consider that the polarity of an individual user is related to their position about the government and, more specifically, the president: we have users (1) pro-government and (2) anti-government.

### 5.1 Users polarity

The approach used to calculate individual users’ polarity follows three main steps:

1. Classify selected hashtags related to politics as anti-government or pro-government;
2. Label tweets and users who authored these tweets according to the two categories of hashtags; and
3. For those who did not use the selected hashtags, classify them based on their retweet behavior towards other classified users. For each week, we build a retweet network based on the concept of homophily (Solomon et al. 2019).

We detail each of these steps below.

#### 5.1.1 Hashtags selection

In order to capture political opinions effectively, we manually selected and classified hashtags<sup>6</sup> as in favor or against the government. Table 5 presents a subset of these hashtags. We selected in total 59 anti-government and 42 pro-government hashtags. The hashtags include opinions about the government and the current and former presidents.

#### 5.1.2 Classifying users based on hashtags

In order to perform a temporal analysis of polarity, users were labeled according to their position during the period that data was collected. Initially only the users who posted tweets using the selected hashtags were classified. Given  $U$  as the set of all users in the dataset, for each selected user  $u$ , their tweets were labeled in 2 categories: anti-government ( $|M^+|$ ) and pro-government ( $|M^-|$ ). For each week, users where  $|M^+| > 0$  and  $|M^-| = 0$  are labeled as in favor of the government (+). On the other hand, users where  $|M^-| > 0$  and  $|M^+| = 0$  are labeled as against the government (-). The

<sup>6</sup> [https://docs.google.com/spreadsheets/d/1WZ73ED8MC7ksBIktoQTvjziNQBvxsqkzDREQ\\_wndXSG/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1WZ73ED8MC7ksBIktoQTvjziNQBvxsqkzDREQ_wndXSG/edit?usp=sharing).



**Table 5** Hashtags related to the public political opinion regarding the government in the tweets dataset

Anti-government Hashtags
#ForaBolsonaro( <i>Bolsonaro Out</i> )
#BolsonaroGenocida ( <i>Bolsonaro is a Genocide</i> )
#ImpeachmentDoBolsonaro Urgente ( <i>Impeachment Now</i> )
#StopBolsonaro Mundial ( <i>Stop Bolsonaro Worldwide</i> )
#BolsonaroAcabou ( <i>Bolsonaro is Over</i> )
Pro-government Hashtags
#BolsonaroTemRazao( <i>Bolsonaro is Right</i> )
#BrasilComBolsonaro ( <i>Brazil with Bolsonaro</i> )
#DireitaComBolsonaro ( <i>Right Wing with Bolsonaro</i> )
#EuApoioBolsonaro ( <i>I Support Bolsonaro</i> )
#BolsonaroReeleito ( <i>Bolsonaro Re-elected</i> )

remaining users were not labeled at this step because their position was not clear from their messages. Then, the set of users  $U$  is divided into two subsets:  $U_{labeled}$  and  $U_{unlabeled}$ , where the opinion of labeled users is already known.

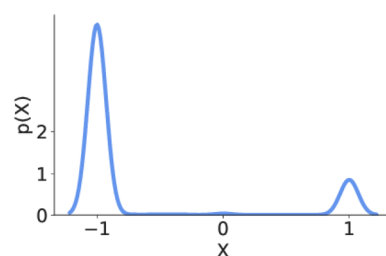
### 5.1.3 Labeling users based on homophily

Next, we deal with the unclassified users from the  $U_{unlabeled}$  set. The data was split in weekly slices, and for each week we build a retweet network to calculate polarity values for unlabeled users. Each network corresponds to a weighted, directed graph  $G = \{V, E\}$ , where  $V$  is the set of vertices (users) and  $E$  the set of edges. Users  $u$  and  $v$  are connected by an edge  $e(u, v)$  if  $u$  retweeted a post from  $v$  and the weight of the edge  $w$  is the total number of tweets from  $v$  retweeted by  $u$ . Isolated vertices represent users who made only original tweets (i.e. authorial tweets) and were not retweeted.

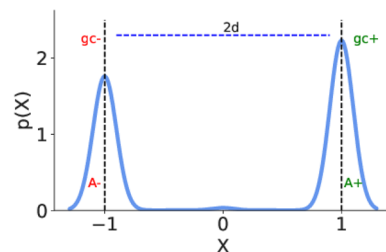
These networks were used to calculate the political polarity of unlabeled users based on the concepts of homophily: users that retweet messages from pro-government users tend to be pro-government and vice-versa. Given that an unlabeled user  $u$  retweeted messages from a set of  $n$  labeled users, they have  $n$  edges  $E_u = \{e_1, e_2, \dots, e_n\}$ . Equation 1 shows how the number of retweets made by user  $u$  is calculated.

$$M_u = \sum_{e_i \in E_u} w_{e_i} \tag{1}$$

Recall that users polarity  $x \in \{+, -\}$ , and edges that intercept user  $u$  can be divided into two groups: those that connect  $u$  to users pro-government (+) and those that connect  $u$  to users anti-government (-). The total number of retweets  $M_u^x$  made by user  $u$  from users of position  $x$  is equal to  $\sum_{e_i \in E_u^x} w_{e_i}$ . From these values it is possible to calculate the



(a) Week 1 (20/04/26-20/05/02)



(b) Week 24 (20/10/05-20/10/11)

**Fig. 4** Probability density function for Twitter political polarities in Brazil during Weeks 1 (a) and 24 (b). 4b shows the variables of the population of opposing opinions ( $A^-$  - against government and  $A^+$  - pro-government), the centers of gravity for each population ( $gc^-$  and  $gc^+$ ), and the distance between the centers of gravity ( $d$ )

proportion of retweets  $r_u^x$  from unlabeled users on users of different opinions, as shown on Equation 2.

$$r_u^x = M_u^x / M_u \tag{2}$$

Following this approach, we calculate the polarity value for each non-labeled user  $u$ ,  $p_u = r_u^+ - r_u^-$ , i.e., the difference between the proportions of retweets made by a given user  $u$  over the set of labeled users' posts for each position. The polarity value comprises the values of the interval  $[-1, 1]$  and represents the level of proximity of a user to a given opinion.

Values close to +1 indicate that the user tends to be in favor of the Brazilian government and the current president. On the other hand, values close to -1 indicate that the user tends to be against the Brazilian government and the current president. Given the polarities values, we calculate the overall population’s polarization index for the given period.

### 5.2 Polarization index

We use a probability density function (PDF) to define the overall polarization considering all the labeled individual users. Figure 4 presents the polarization distribution for Weeks 1 and 24 from the dataset. This metric considers the size of the populations with opposite opinions, that is, those in favor of the government ( $A^-$ ) and those against the government ( $A^+$ ), the gravity centers of each population ( $gc^-$  and  $gc^+$ ) and the distance between these gravity centers ( $d$ ).

The polarity  $X$  of each individual is defined as a value in  $[-1, 1]$ , and we calculate the size of the populations with diverging opinions by taking into account the PDFs of polarities  $p(X)$  for the set of studied individuals. Thus, the polarization index for the population with negative opinions  $A^-$  ( $X < 0$ ) is calculated by integrating the distribution  $p(X)$  in the interval  $[-1, 0]$  (Equation 3). The polarization index for the population with positive opinions  $A^+$  ( $X > 0$ ) is then calculated by integrating the polarity distribution in the interval  $[0, 1]$ , as shown in Equation 4.

$$A^- = \int_{-1}^0 p(X)dX = P(X < 0) \tag{3}$$

$$A^+ = \int_0^1 p(X)dX = P(X > 0) \tag{4}$$

Then we calculate the absolute difference between the size of populations  $\Delta A = |A^+ - A^-|$ . This variable represents the level of imbalance between groups of users with opposite opinions. As the values  $A^+$  and  $A^-$  lie in the  $[0, 1]$  interval, the normalized difference  $\Delta A$  is also restricted to the range  $[0, 1]$ . If one population is much larger than the other, this metric will have a large value and the probability distribution will have the shape of a unimodal distribution.

Another variable related to the polarization index is the distance  $d$  between the gravity centers, which quantifies the level of divergence between the opposite populations. In other words, this variable measures how different the opinions of the two populations are. The gravity centers of the positive and negative populations can be seen in the Equations 5 and 6 respectively.

$$gc^+ = \frac{\int_0^1 p(X)XdX}{\int_0^1 p(X)dX} \tag{5}$$

$$gc^- = \frac{\int_{-1}^0 p(X)XdX}{\int_{-1}^0 p(X)dX} \tag{6}$$

The distance  $d$  is computed as the normalized difference between the gravity centers  $gc^-$  (Eq. 6) and  $gc^+$  (5), as shown by Equation 7:

$$d = \frac{|gc^+ - gc^-|}{|X_{max} - X_{min}|} = \frac{|gc^+ - gc^-|}{2} \tag{7}$$

where  $X_{max}$  represents the maximum polarity value for the positive population ( $X_{max} = 1$ ) and  $X_{min}$  represents the minimum polarity value for the negative population ( $X_{min} = -1$ ). If  $d = 0$  then the gravity centers of the populations are equal and the individuals share the same opinion. Otherwise, if  $d = 1$ , the populations opinions are extremes and perfectly opposed. Finally, from the difference  $\Delta A$  between the population sizes and the distance  $d$  between the gravity centers, the polarization index  $\mu$  for the set of users is calculated, and shown in Equation 8.

$$\mu = (1 - \Delta A)d \tag{8}$$

The polarization index  $\mu$  lies on the range  $[0, 1]$ . When  $\mu = 1$ , then the population is perfectly polarized. This is the case when opposing populations have the same size and their polarities are centered in the extreme values (-1 and 1). Otherwise, if  $\mu = 0$ , then the opinion distribution is not polarized at all. In this case, the probability distribution of polarities takes the shape of an unimodal distribution and the difference between the populations sizes is equal to 1 ( $\Delta A = 1$ ), implying that the population is centered at a neutral opinion or one of the extremes. Figure 4b shows as example for when the value of  $\mu$  is equal to 0.85, as the values of the difference of populations ( $\Delta A$ ) and the distance between the gravity centers ( $d$ ) are equal to 0.057 and 0.90, respectively.

### 5.3 Topics analysis

Finally, we have also performed a topic analysis over the tweets posted by users pro-government and anti-government to compared whether the polarity was reflected in the text. For each week of the dataset we run Biterm Topic Model (BTM) (Yan et al. 2013), a generative method conceived to deal with short-text. BTM deals with the whole set of tweets as a single document, and models the collection as a mixture of topics.

For each time interval  $t$ , we used BTM to generate a set of  $k$  topics, where  $k$  is a user defined parameter, and each topic is represented by its most representative terms. After that, we followed the approach proposed by Moreira et al. (2020), where a similarity graph  $G_t = \{V_t, E_t\}$  of the topics

**Table 6** Network metrics for retweet graphs

Property	Max	Min	Avg
Vertices	2,277,658	4,385,12	1,060,410.19
Edges	4,691,621	377,321	1,640,000.64
# Isolated vertices(%)	869,826 (43.2)	174,239 (24.7)	328,435.42 (32.3)
Avg. degree	4.63	1.72	2.78
Avg in-degree	2.32	0.86	1.39
Largest connected component	25,533.00	61	4,777.53
Density	2.15E-06	6.83E-07	1.47E-06
# components	2,256,438	438,178	1,054,625.28
# maximal cliques	6,573,259	538,957	1,985,270
Largest clique	21	6	8.08

produced over time is created. In this graph, nodes represent topics and edges the similarity between two topics. The similarity is calculated using the Jaccard coefficient between the words that describe the two topics, i.e., their proportion of common words.

From the graph  $G_t$  we extract a set of “super-topics” by merging topics with a Jaccard coefficient lower than a predefined threshold. We used the value of 0.35 after preliminary tests. Then, the relevance of a super-topic  $T_x$  at time interval  $i$  is given by its popularity (i.e., the frequency it appears in the tweets) in the period.

### 6 Results and discussion

We proposed a user political polarity method that works in three phases. First we label users according to the hashtags they posted. In this phase, 74,503 users were labeled as positive (pro-government) and 74,368 as negative (anti-government). In the second phase, we built the retweet networks for each week. Table 6 shows a summary of these networks. As observed, around a third of users (32.3% of isolated vertices) did not retweet or were retweeted by any other user in the network – and were discarded at this stage. The networks are also very sparse, with an average density of  $1.47 \times 10^6$ .

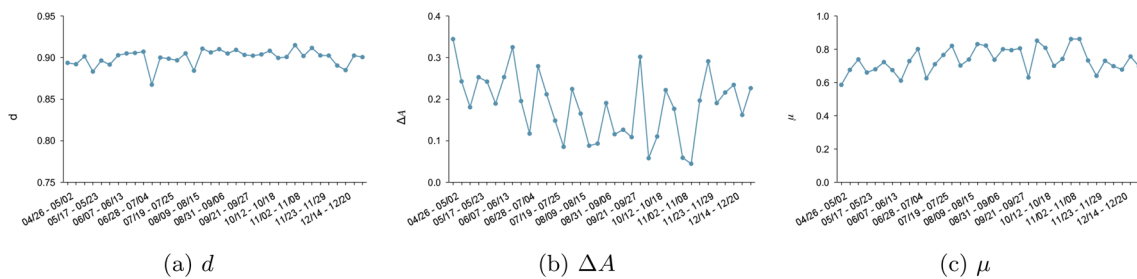
There is also a large number of connected components with few users. Then, the 66% of remaining users that belong to a connected component with at least one edge had their polarization value computed according to the methodology described in Sect. 5.

Figure 4 shows the PDFs of the individual polarity values for first and 24th weeks. Note that in both weeks most of the users are clearly concentrated on divergent groups, with a small number of users having polarity values close to 0. In the first week (Fig. 4a), most users have polarity values less than 0 and, consequently, have opinions more similar to those against the Brazilian government and the current president. On the other hand, in the 24th week (Fig. 4b), most users have a polarity value greater than 0, and consequently, have opinions more similar to those who support the government and the president.

We computed the polarization index  $\mu$  of the users for each time slice from the PDFs. Figure 5 shows the temporal evolution of  $\mu$  and its related variables: the difference of populations  $\Delta A$ , and the distance between gravity centers  $d$ .

The results show that the polarization index has high values over the entire studied period, which reveals that Twitter users were highly polarized over the weeks regarding the political scenario in Brazil and its current president. The distance between gravity centers ( $d$ ) remains almost uniform (average of 0.90 and standard deviation of 0.01) over time, and hence the polarization index was mostly affected by oscillations in the differences between populations sizes ( $\Delta A$ ).  $\Delta A$  has a minimum value of 0.04 and a maximum value of 0.34.

Note that the polarization index peak ( $\mu = 0.86$ ) occurs in two consecutive weeks, from 2 November 2020 and 15 November 2020 (weeks 28 and 29). It is interesting to note that the polarization index has its peak when  $\Delta A$  is close to 0, indicating that groups of supporters and non-supporters of the government have the same density, and hence the polarity value is largely determined by the distance between the gravity centers. The polarization index recorded its minimum value in the first week, between the 26 April 2020 and 02 May 2020 ( $\mu = 0.59$ ). Since there was no apparent change in  $d$  over the weeks, we can assume that the variation



**Fig. 5** Time evolution of polarization index ( $\mu$ ) and its related variables ( $\Delta A$  and  $d$ )



(a) Anti-government users. (b) Pro-government users. (c) Anti-government bots. (d) Pro-government bots.

**Fig. 6** Word clouds for genuine users and bots pro- and anti-government in the least polarized week (week 1)



(a) Anti-government users. (b) Pro-government users. (c) Anti-government bots. (d) Pro-government bots.

**Fig. 7** Word clouds for genuine users and bots pro- and anti-government in the most polarized weeks (weeks 28 and 29)

of  $\mu$  is related to the variation on  $\Delta A$ . Observe that the largest difference of populations ( $\Delta A$ ) occurs in the first week.

These numbers coincide with major events that occurred in the country. First, the peak occurred in the weeks preceding and during the municipal elections that happened on November 15<sup>7</sup>. These elections were marked by a highly polarized dispute between candidates endorsed by the president, that often defended similar actions as those implemented by the federal government (to ease social distancing and other restrictions, and argued that the economic impact is worse than the virus itself), and those in favor of strictly following the recommendations of the World Health Organization. Early in November 2020 the country also registered the lowest 14-day moving average of cases since the beginning of the pandemic. And, finally, in this same week, Pfizer offered Brazil a deal for millions of vaccines<sup>8</sup>.

We also checked the polarity of users flagged as bots. From the set of 4,638 bots, 1,060 are labeled as pro-government and 600 as anti-government. Looking at their polarity index over the weeks they remain stable, showing the clear position of the bots. Figures 6 and 7 show the word clouds of the most frequent terms appearing in tweets anti-government and pro-government for both genuine users and bots for the least and most polarized weeks. We show them in Portuguese and discuss the most relevant terms together with their

translation. First, note that the clouds are not too different, including similar terms such as “bolsonaro”, “presidente” (president), “mortes” (deaths) and “saude” (health). We observe differences in frequency of some relevant words, e.g., “casa” (home), which is more frequent for the anti-government public, supporters of the “stay home, stay safe” campaign. For bots, it is interesting to observe that the pro-government bots tend to use random words in their posts (e.g., “maconha” (marijuana) and “teclado” (keyboard)) while the anti-government bots use terms more related to the pandemics, including “gripezinha” (harmless flu), a term adopted by the president to define COVID-19 since the first infections were detected in Brazil.

For the most polarized week 29 (11/09 - 11/15) we see a much clearer bias on subjects discussed by genuine users anti- and pro-government. For the anti-government, the main subject is “vacina” (vaccine), given that during this period Pfizer offered Brazil millions of doses of the vaccine. For the pro-government users, one of the most frequent words is “vassoura” (broom). This is related to a campaign made by the restaurant Burger King for Halloween. In the middle of the pandemic, the restaurant invited the public to go to the drive-through in a witch broom in exchange for a free sandwich. Stores were packed with people, starting a “war” between people pro- and anti-government, as the former still defend the “herd immunity policy” and claims that using masks do not change the pandemic scenario. The bots, during this period, tweeted similar topics. Again, “vassoura” (broom) appears in both clouds. Also observe the use of

<sup>7</sup> <https://bit.ly/3uUS2OT>.

<sup>8</sup> <https://reut.rs/3LFJ6Tg>.



**Table 7** Words describing the top-5 most relevant topics discussed by users in week 29 (the most polarized week)

Id	Anti-government	Pro-government
1	Days, people, chloroquine, quarantine, now	Acai, hair, unemployment, shakira, mobile
2	Cases, Brazil, deaths, raise, pandemics	Quarantine, wave, home, day, second, Brazil, people
3	Vaccine, health, ministry, efficacy, pfizer, coronavac	WHO, against, pandemics, president, airplane
4	Hair, green, shakira, unemployment, joao, mobile, acai	Brasil, Argentina, millions, causes, cases, deaths
5	Treatment, disease, social, precocious, masks, distance	Treatment, precocious, doctor, symptoms, air, feel

the word “cadela”, which is a Brazilian slang used in social media to say that someone really *admires* someone/something else. For all word clouds we also see the presence of “Recife” – a Brazilian big city that had a heated municipal elections at the time (two cousins from a traditional political family were running against each other) – and João Dória, the governor of the state of São Paulo, a center-right wing politician who supported the development and administration of the Chinese CoronaVac vaccine. Bolsonaro said he would not buy “João Dória’s Chinese vaccine” in October 2020, claiming that the population was not going to be treated as guinea pigs<sup>9</sup>.

Going beyond the frequency of the most discussed words, we performed a topic analysis according to the methodology previously described and considering  $k=10$  topics for anti- and pro-government users in week 29. Table 7 shows the words describing the most discussed topics. Observe that users anti-government are discussing the use of chloroquine (topic 1), the increase in the number of deaths (topic 2), vaccines (topic 3), and measures of prevention against covid19 (topic 5). Topic 4 seems random. User pro-government, in contrast, also have a first topic with random words (very similar to the ones used by anti-government users) followed by a topic discussing their despise for measures such as quarantine (topic 2), the positions of WHO and the president regarding the pandemics (which are conflicting) (topic 3), the situation of the pandemics in Brazil and Argentina (topic 4) and, finally, types of precocious treatments for covid, which could be offered as soon as you have problems breathing (represented by the word “air”). The topics only corroborate the information from the words clouds but in a more informed way. A temporal topic analysis of what was discussed during the 36 weeks of the dataset can be found in Appendix A.

In general, we observe that the increase in the value of the polarity index in certain weeks can also be explained by certain events that occurred during that specific period. For instance, the abrupt resignation of the Minister of Health Nelson Teich<sup>10</sup> after less than a month (15 May 2020- 3rd week), and the release by the Brazilian Supreme Court of

a video of a cabinet meeting<sup>11</sup> (22 May 2020 - 4th week). In addition, it is possible to observe the variation in polarization index according to events related to the COVID-19 pandemic, such as the expansion of the use of hydroxychloroquine in Brazil<sup>12</sup> (20 May 2020 - 4th week) and the significant increase in the number of deaths in the country<sup>13</sup> (20 May 2020 - 4th week). This demonstrates that discussions and comments regarding the COVID-19 pandemic on Twitter are strongly related to the health policies and governmental declarations in Brazil.

## 7 Conclusion and future work

In this paper we collected and analyzed Twitter data related to the COVID-19 pandemic. We focused on messages written in Portuguese in the Brazilian political context. We looked at the presence of bots and at the polarity of users towards the Brazilian government and its policies to deal with the pandemic.

The data was collected from April 2020 over 36 weeks. The dataset contains over 100 million tweets and 7 million Twitter users. Although we cannot guarantee all messages came from Brazil – and this is a limitation of our method – we know Brazil is the 4th country with most active Twitter users in the world, and messages related to Brazilian politics are most likely to come from Brazilian users. Likewise, the terms used for data collection were related to Covid, and then messages filtered by hashtags related to politics. If we had collected political terms and then filtered by Covid, the results could have been different. This is a bias of the proposed approach.

We used Botometer, a web-service dedicated to detecting bots on Twitter, to identify bots in the dataset used in this article. We also developed a method of ranking users so that users at the top of the rankings are more likely to be bots than other users. This contributes to the scalability of bot detection algorithms or services because it allows finding

<sup>9</sup> <https://bbc.in/34ZUoB4>.

<sup>10</sup> <https://bit.ly/3rSCUQ6>.

<sup>11</sup> <https://bbc.in/3sIlapM>.

<sup>12</sup> <https://reut.rs/3sP0rRi>.

<sup>13</sup> <https://reut.rs/3JzF5Ox>.



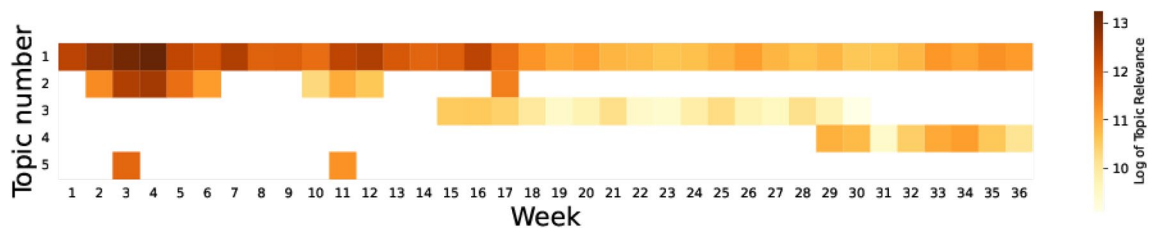


Fig. 8 Top-5 most relevant super-topics discussed by users pro-government over time

Table 8 10 most relevant words describing the topics discussed by pro-government users over time

Id	Top-5 Terms
1	Governors, deaths, hospital, corona, Nise, combat, WHO, doctors, government, now
2	WHO, Lancet, deaths, science, patients, saves, Raoult, symptoms, notice, azithromycin
3	Quarantine, ramalho, watch, lives, broom, Recife, joao, sorry, peak, weak flu
4	Uai, mobile, acai, caraio, hair, joao, iphone, shakira, igor, smile
5	Covid19, positive, result, chloroquine, exam, STF, Jair, test, president, negative

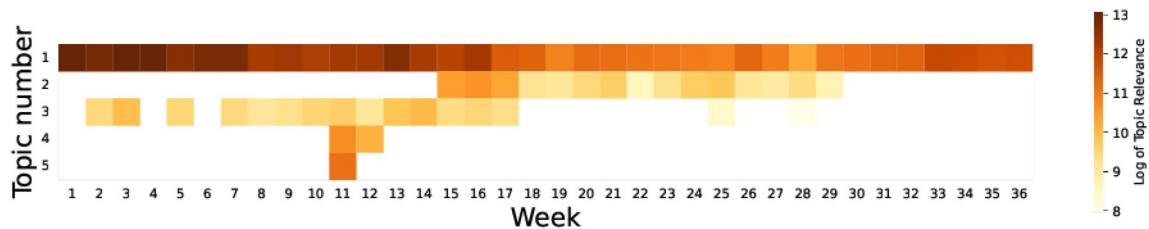


Fig. 9 Top-5 most relevant super-topics discussed by users anti-government over time

Table 9 10 most relevant words describing the topics discussed by anti-government users over time

Id	Top 5 Terms
1	President, latin, Pazuello, hydroxychloroquine, brand, workers, Brazil, vaccine, variant, ozone
2	Recife, quarantine, cachorra, weed, watch, joao, cadela, computer, weak flu, vassoura
3	Roberto, followtrick, corno, keyboard, computer, paulo, fuck, iphone, cachorra
4	Do, video, against, hydroxychloroquine, Bolsonaro, hospital, chloroquine, medicine, use, covid19
5	Advertisement, positive, PF, Bolsonaro, covid, chloroquine, now, fake, covid19

a significant amount of bots in fewer evaluations than randomly evaluating users.

With these bots and the set of genuine users, we performed a polarity analysis. We observed that the polarity index varied from 0.59 to 0.86 during this period, in a scale from 0 to 1. An analysis of the content of the tweets posted by users pro- and anti-government showed they differ significantly in the most polarized weeks. The variations on the polarity index can be explained according to the political and health-related events that happened in the analyzed weeks.

An analysis of the content of bots according to their polarization also showed they have a different bias, with

bots pro-government being more generic in their posts. The results could be better if there was an easier and faster mechanism for detecting bots. We also consider that the mechanism of retweeting can be two-folded, that is, people can retweet and comment on messages of others on the opposite side of the discussion to show disagreement. In those cases, we require a better analysis of the texts, which would lead the work to a better mapping of the users' polarization position. As future work, we intend to look at the dissemination of misinformation, and the role of bots and activists in this process.

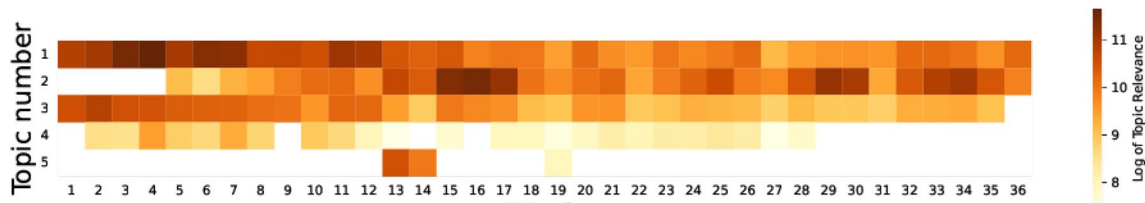


Fig. 10 Top-5 most relevant super-topics posted by bots over time

Table 10 10 most relevant words describing the topics posted by bots over time

Id	Top 5 Terms
1	quarantine, covid19, left, STF, york, States, total, stayhome, year, anvisa
2	iphone, bus, cachorra, green, naruto, quarantine, shakira, sorry, sensible, bots
3	yes, hospital, quarantine, days, torcidas, covid19, Rio, take, world, social
4	porn, unemployed, spree, quarantine, sorry, finished, pop, followtrick, bts, gay
5	broom, shit, smile, jaehyo, naruto, eat, all, trump, quarantine, fuck

### Topic analysis

This section presents a topic analysis performed according to the methodology introduced in Sect. 5.3. We used BTM to extract 10 topics for each of the 36 weeks, and then merged these topics according to the methodology to better follow the topics along time.

Figures 8 and 9 show the top-5 most relevant super-topics to each of the 36 weeks considered in the analysis, while Tables 8 and 9 list the most relevant terms for the topics discussed by users pro-government or against-government, respectively. The relevance of a topic is calculated according to the number of tweets in the dataset that have that topic. From the 360 initial topics, 278 super-topics were obtained for users pro-government and 159 for users anti-government, showing the users that support the president discussed a wider range of topics. For bots, we had 83 super-topics.

The heatmap in Fig. 8 shows topic 1 (first line) was the most discussed along time. It is a general topic, and includes terms like governors, combat, covid, deaths. The most specific term is “nise” (Nise Yamagushi), the doctor that advised Bolsonaro in the beginning of the pandemics and that was in favor of using chloroquine as a treatment. The second most present topic along time is more concentrated at the beginning of the pandemics. It talks about WHO, science, and Raoult, the French doctor that defended the use of chloroquine, something also defended by Bolsonaro’s supporters. Topic 3 concerns events during the pandemics, which Bolsonaro’s supporters always referred to as a “gripezinha” (weak flu). Topic 5 is more scattered, and concerns the times Bolsonaro himself was believed to have gotten covid. Topic 4 looks like spam, with words that do not make much sense together.

Turning to Fig. 9, topic 1 again is recurrent over time, but appears much stronger than the topics pro-government. The main topic here is vaccination, and governments preference to treat covid with chloroquine. It also brings the name of Pazuello, Ministry of Health for some time and that supported all of Bolsonaro’s ideas in favor of chloroquine and against lockdown. Some words that may seem unrelated, like *ozone*, do have a reason to appear. In this case, for example, people had very strong opinions about a viral video made by a politician that suggested to treat COVID by applying ozone to the anus of the patient. Topic 2, as in the pro-government topics, has also references to events during the pandemics, such as Burguer King’s action already mentioned, but also brings some spam words. Topic 3 talks about Bolsonaro’s videos and positions defending chloroquine. Following, topic 5 is related and discusses Bolsonaro’s strategy of spreading fake news regarding the pandemics. Is also shows PF, which is the Federal Police and with which he had many problems during the time.

Finally, we looked at the topics produced by bots in contrast to those produced by legitimate users. Figure 10 shows again the top-5 most relevant topics, while Table 10 shows the most relevant terms appearing in these topics. First notice that the topics are much more consistent during the whole period, with the first three topics being active during the whole time. The first topic talks about the quarantine, the left wing party and has the hashtag “stayathome”. Topics 2 and 4 are a mixture of random words, while topic 3 is related to hospitals and quarantine. Topic 5 again does not have a “structure”, but uses a set of words that were popular along the time of the pandemics to call attention to their tweets (e.g., broom).

**Acknowledgements** The authors would like to thank FAPEMIG, MPMG (project Analytical Capabilities), CNPq, CAPES, MASWEB, INCT-Cyber, CIIA-Saúde, and MCTIC/RNP for the financial support.

**Data availability** The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Conflict of interest** All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

## References

- Albinati J, Meira JrW, Pappa GL, Teixeira M, Marques-Toledo C (2017) Enhancement of epidemiological models for dengue fever based on twitter data. In Proc of the ACM Int Conf on Digital Health 66:109–18
- Araújo E, Ferro M, Silva G (2020) Disconnecting for the good: a network-oriented model for social contagion of opinions and social network interventions to increase adherence to social distancing. In Anais IX Brazilian Workshop Soc Netw Anal Min 54:142–153 (SBC)
- Ball P, Maxmen A (2020) The epic battle against coronavirus misinformation and conspiracy theories. *Nature* 581:371–374
- Bhattacharya C, Chowdhury D, Ahmed N, Özgür S, Bhattacharya B, Mridha S, Bhattacharyya M (2021) The nature, cause and consequence of covid-19 panic among social media users in india. *Soc Netw Anal Min* 11(53):1–11
- Ceron W, de Lima-Santos MF, Quiles MG (2021) Fake news agenda in the era of covid-19: identifying trends through fact-checking content. *Online Soc Netw Med* 21:100116. <https://doi.org/10.1016/j.osnem.2020.100116>
- Chapelle O, Scholkopf B, Zien A (2006) Eds. *Semi-Supervised Learning*
- Charron N, Lapuente V, Rodriguez-Pose A (2020) Uncooperative society, uncooperative politics or both? How trust, polarization and populism explain excess mortality for covid-19 across European regions
- Chu Z, Gianvecchio S, Wang H, Jajodia S (2010) Who is tweeting on Twitter. In Proceedings of the 26th annual computer security applications conference on - ACSAC '10. ACM Press, New York, New York, p. 21
- Cinelli M, Quattrocioni W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, Zola P, Zollo F, Scala A (2020) The covid-19 social media infodemic. *Sci Rep* 10(1):1–10
- Cresci S, Di Pietro R, Petrocchi M, Spognardi A, Tesconi M (2017) The paradigm-shift of social spambots: evidence, theories, and tools for the arms race. In Proceedings of the 26th international conference on World Wide Web Companion, WWW '17 Companion, pp. 963–972
- Davis CA, Varol O, Ferrara E, Flammini A, Menczer F (2016) BotOrNot. Association for computing machinery (ACM), pp. 273–274
- de Mello Araújo EF, Franke A, Hosain RW (2017) A temporal-causal model for spread of messages in disasters. In International conference on computational collective intelligence, Springer, pp. 386–397
- Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, and Larson H (2020). The pandemic of social media panic travels faster than the covid-19 outbreak
- Ferrara E (2020) What types of covid-19 conspiracies are populated by twitter bots? *First Monday* 25(6)
- Ferrara E, Varol O, Davis C, Menczer F, Flammini A (2014) The Rise of Social Bots. <https://doi.org/10.1145/2818717>. arXiv:1407.5225
- Ferrara E, Varol O, Davis C, Menczer F, Flammini A (2016) The rise of social bots. *Commun ACM* 59(7):96–104
- Filho RM, Almeida JM, Pappa GL (2015) Twitter population sample bias and its impact on predictive outcomes: a case study on elections. In 2015 IEEE/ACM International conference on advances in social networks analysis and mining (ASONAM), pp. 1254–1261
- Gallotti R, Valle F, Castaldo N, Sacco P, De Domenico M (2020) Assessing the risks of ‘infodemics’ in response to covid-19 epidemics. *Nat Human Behav* 4(12):1285–1293
- Gilani Z, Kochmar E, Crowcroft J (2017) Classification of twitter accounts into automated agents and human users. In Proceedings of the 2017 IEEE/acm international conference on advances in social networks analysis and mining 2017, pp. 489–496
- Graham T, Bruns A, Angus D, Hurcombe E, Hames S (2020) #istandwithdan versus #dictatorand: the polarised dynamics of twitter discussions about Victoria’s covid-19 restrictions. *Media International Australia: 1329878X20981780*
- Jiang J, Chen E, Lerman K, Ferrara E (2020) Political polarization drives online conversations about covid-19 in the United States. *Hum Behav Emerg Technol* 2(3):200
- Kušen E, Strembeck M (2018) Politics, sentiments, and misinformation: an analysis of the twitter discussion on the 2016 austrian presidential elections. *Online Soc Netw Med* 5:37–50
- Lazer DM, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, Metzger MJ, Nyhan B, Pennycook G, Rothschild D et al (2018) The science of fake news. *Science* 359(6380):1094–1096
- Morales AJ, Borondo J, Losada JC, Benito RM (2015) Measuring political polarization: Twitter shows the two sides of venezuela. *Chaos Interdisciplinary J Nonlinear Sci* 25(3):033114
- Moreira RC, Vaz-de Melo PO, Pappa GL (2020) Elite versus mass polarization on the Brazilian impeachment. *Soc Netw Anal Min* 10(1):1–23
- Prasetya HA, Murata T (2020) A model of opinion and propagation structure polarization in social media. *Comput Soc Netw* 7(1):1–35
- Rodríguez-Ruiz J, Mata-Sánchez JI, Monroy R, Loyola-González O, López-Cuevas A (2020) A one-class classification approach for bot detection on Twitter. *Comput Sec*. <https://doi.org/10.1016/j.cose.2020.101715>
- Sayyadharikandeh M, Varol O, Yang KC, Flammini A, Menczer F (2020) Detection of Novel Social Bots by Ensembles of Specialized Classifiers. *Int Conf Inf Knowledge Manage, Proceed*: 2725–2732. <https://arxiv.org/abs/2006.06867>
- Schmitt J (2016) How to measure ideological polarization in party systems. In ECPR Graduate Student Conference
- Shao C, Ciampaglia GL, Varol O, Yang K, Flammini A, Menczer F (2018) The spread of low-credibility content by social bots. *Nat Commun* 9(1):1–9
- Solomon RS, Srinivas PYKL, Das A, Gamback B, Chakraborty T (2019) Understanding the psycho-sociological facets of homophily in social network communities. *IEEE Comput Intell Mag* 14(2):28–40. <https://doi.org/10.1109/MCI.2019.2901084>
- Stringhini G, Kruegel C, Vigna G (2010) Detecting spammers on social networks. In Proceedings of the 26th Annual Computer Security Applications Conference on - ACSAC '10, New York. ACM Press, New York, Up. 1
- Tucker JA, Guess A, Barberá P, Vaccari C, Siegel A, Sanovich S, Stukal D, Nyhan B (2018) Social media, political polarization, and political disinformation: a review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature*

- Varol O, Ferrara E, Davis CA, Menczer F, Flammini A (2017) Online Human-Bot Interactions: Detection, Estimation, and Characterization. Technical report
- Waszak PM, Kasprzycka-Waszak W, Kubanek A (2018) The spread of medical fake news in social media-the pilot quantitative study. *Health Policy Technol* 7(2):115–118
- Xu W, Sasahara K (2021) Characterizing the roles of bots on Twitter during the COVID-19 infodemic. *J Comput Soc Sci*. <https://doi.org/10.1007/s42001-021-00139-3>
- Yan X, Guo J, Lan Y, Cheng X (2013) A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web, WWW '13*, New York. Association for Computing Machinery, NY, pp. 1445–1456
- Yang KC, Ferrara E, Menczer F (2022) jan. Botometer 101: Social bot practicum for computational social scientists. Technical Report
- Yang KC, Torres-Lugo C, Menczer F (2020) Prevalence of low-credibility information on twitter during the covid-19 outbreak. arXiv preprint [arXiv:2004.14484](https://arxiv.org/abs/2004.14484)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.