FGV DAPP
DEPARTMENT OF PUBLIC POLICY ANALYSIS

BOTS, SOCIAL NETWORKS AND POLITICS IN BRAZIL

Cases of unlawful interference by automated profiles in the public debate

Vol.2

August 2018

Ficha catalográfica elaborada pela Biblioteca Mario Henrique Simonsen/FGV

Bots, social networks and politics in Brazil [recurso eletrônico] : cases of unlawful interference by automated profiles in the public debate / Marco Aurélio Ruediger, coordinator. – Rio de Janeiro : FGV DAPP, 2018.
1 recurso online (53 p.) : PDF, il. – (Caderno de referência; 2)

Dados eletrônicos.
Inclui bibliografia.
ISBN: 978-85-68823-85-9

1. Políticas públicas. 2. Eleições. 3. Redes sociais on-line. 4. Boatos (Opinião pública). 5. Internet. 6. Robos. 7. Computação humana. I. Ruediger, Marco Aurélio, 1959- . II. Fundação Getulio Vargas. Diretoria de Análise de Políticas Públicas.

CDD — 351

How to cite
RUEDIGER, M. A. (Coord.). Bots, Social Networks and Politics in Brazil: Cases of unlawful interference by automated profiles in the public debate [Caderno de referência] .Vol. 2. Rio de Janeiro: FGV DAPP, 2018.

FGV DAPP

BOTS, SOCIAL NETWORKS AND POLITICS IN BRAZIL

Cases of unlawful interference by automated profiles in the public debate

Volume 2

Department Of Public Policy Analysis Fundação Getulio Vargas

Volume
2

Bots, Social Networks and Politics in Brazil

Cases of unlawful interference by automated profiles in the public debate

Marco Aurelio Ruediger
Coordinator

Rio de Janeiro
FGV DAPP
2018

Editorial staff

FGV Founded in 1944, Fundação Getulio Vargas was created to promote social and economic development in Brazil, through the education of qualified administrators in both public and private spheres. Over time, FGV has expanded its activities to other areas of knowledge, such as social sciences, law, economics, history and, most recently, applied mathematics, which is a benchmark in quality and excellence in all of its eight schools.

Office
Praia de Botafogo 190, Rio de Janeiro RJ - Zip code 22250-900
PO Box 62.591 Zip Code 22257-970
Tel (21) 3799-5498 | www.fgv.br

Founding Chairman
Luiz Simões Lopes

President
Carlos Ivan Simonsen Leal

Vice-Presidents
Francisco Oswaldo Neves Dornelles (on leave)
Marcos Cintra Cavalcanti de Albuquerque (on leave)
Sergio Franklin Quintella

FGV DAPP **Director**
Marco Aurelio Ruediger
DAPP
(21) 3799-4300
www.dapp.fgv.br | dapp@fgv.br

Coordination
Marco Aurelio Ruediger

Researchers
Amaro Grassi
Ana Freitas
Andressa Contarato
Danilo Silva
Kaizo Beltrão
Lucas Calil
Lucas Roberto da Silva
Polyana Barboza
Rachel Bastos

Graphic Design
Gabriela Lapadula
Luis Gomes
Rodrigo Cid

Editorial Review
Ana Freitas
Ana Guedes

Translation
Carolina Aleixo
Dalby Dientsbach

Summary

Executive Summary **06**

Introduction **10**

I. BOTS **12**

1.1 Background information 14
1.2 Bot typology 20
1.3 Refined methodology 26
1.3.1 BotOrNot 26
1.3.2 Correlation 28
1.3.3 DAPP Methodology 30

II. CASE STUDIES **32**

2.1. Queermuseu exhibition 34
2.2. The pre-election debate in Paraguay 36
2.3. Former President Lula's trial 38
2.4. Political actors 42

III. IMPLICATIONS FOR POLITICS AND THE 2018 ELECTIONS **48**

Bot glossary **52**

References **53**

EXECUTIVE SUMMARY

- The influence of automated profiles on the debate on social networks threatens public debate's integrity, since bots can distort the pros and cons scenarios relative to a given topic or actor.

- Due to their harmful manipulative potential, we should identify bots, especially during events of great political relevance, such as electoral campaigns.

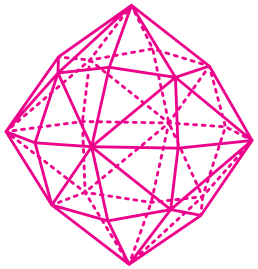
- Novel techniques gradually enhance bots' sophistication and provide them with more and more human-like traits, which makes their behavior more complex and thus hinders their identification.

- In order to follow advancements in automation, FGV DAPP often improves its guidelines for identifying automated profiles, by employing statistical methods that provide an accurate check-up process.

- Enhanced resources for such an identification allow a more precise and accurate recognition of patterns, groups engaged in spreading content, and evidence of content replication on the web.

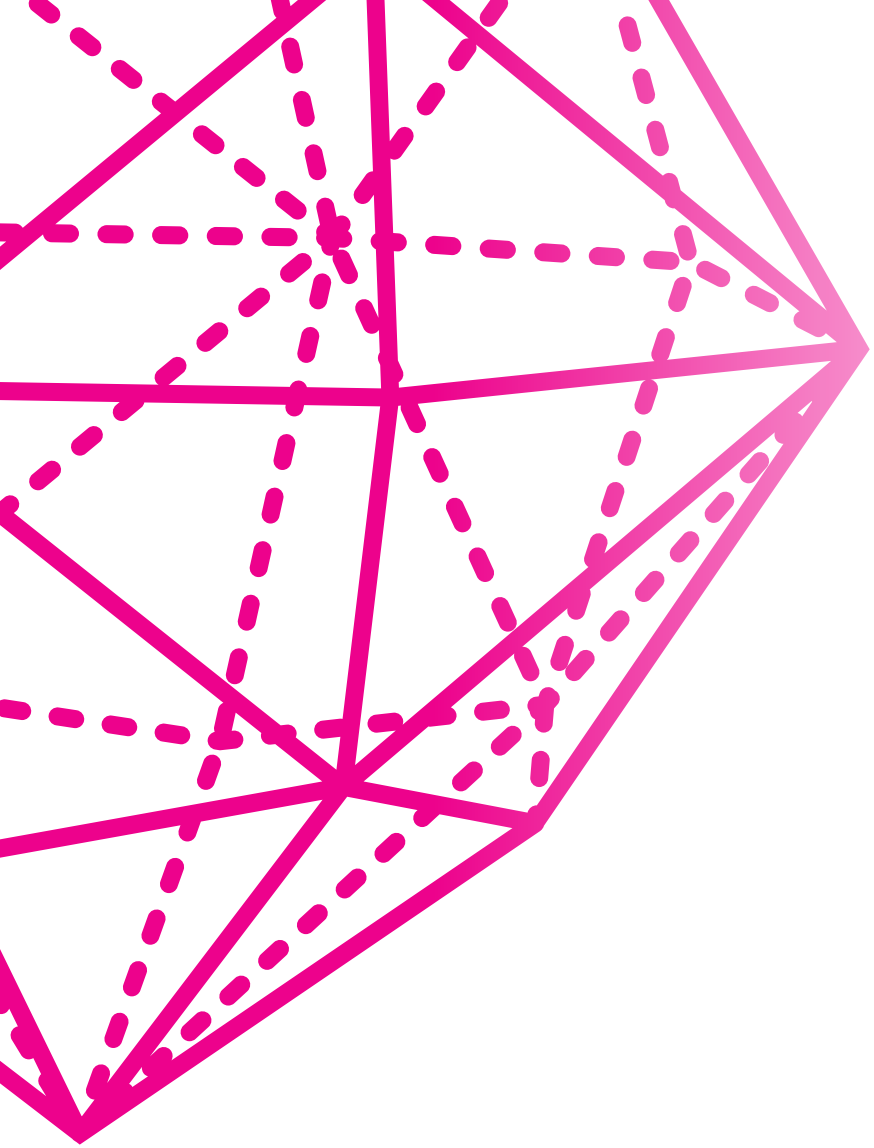
- Besides describing DAPP methodology for identifying automated profiles, this paper presents its application in four cases: the Queermuseum exhibit, the pre-electoral debate in Paraguay, debates on eight Brazilian political actors, and the trial of former-president Lula. In all cases, we could identify a substantial use of bots.

- Applied studies emphasize the challenge faced by institutions that have been in charge of monitoring the correctness of electoral processes, in order to safeguard their transparency and promote their accountability.



INTRODUCTION

> The management of automated accounts takes advantages of different textual and programming resources to avoid being recognized and subject to deletion.



INTRODUCTION

The first reference book published by FGV DAPP on bots acting in the public debate, Bots, social networks, and politics in Brazil (RUEDIGER, 2017a), had two purposes.

The first one was to warn about the repeated interference of automated profiles with relevant events in politics in Brazil.

The second one was to establish an initial methodological effort to approach different strategies and functions related to the activity of these bots, with the design of a monitoring and identification typology of such activity.

Bots, however, are only a vertex of interaction manipulation via social networks. Alongside the growing concern about the impact of automated profiles on democracy and politics, not only in Brazil, but also in Europe and in the United States, there has been close attention to the so-called "fake news".

Fake news try to distort opinions, make value judgments about people and agendas, and undermine the traditional journalism.

The first studies and reflections on the pragmatic result of bots and fake news on networks are foreign, especially in the US, where the fog covering the Russian participation in Donald Trump's election as president remains dense. Still under investigation in the country, the use of fake profiles and spread of messages favorable to Trump by the Russians accelerated the response of different sectors of democracy to the problem.

At the end of 2017, the European Union created a working group dedicated to prevent false news and interactions on the web with members of the academic community, press and government. In Brazil, the Superior

Electoral Court formed an advisory board dedicated to the same task, of which FGV DAPP is member.

However, the correlations between bot production and use of false information to manipulate elections and the public debate are unclear. A study published on January 9, 2018 (GUESS et al., 2018) on the audience of fake news by Trump's and Hillary Clinton's voters during the 2016 elections showed a strong overlap between reading fake news and predisposition of voters who read the news to vote for the favored candidate. That is, the access to distorted content is predominant among those who are already inclined to support the candidate benefited by the distortion.

Similarly, there is not a strong evidence collection indicating the widespread use of bots to share false information on social networks. FGV DAPP's preliminary studies on the activity of automated profiles on Twitter, even among the examples in this second book, do not lead to this conclusion, but restate the main problem: bots on social networks threatens any scenario of support, criticism, reach, and interest of the civil society in a country's major matters due to artificial engagements.

Another problem identified in the analyses by FGV DAPP is the component of adaptation of artificial activity on Twitter. Bots are not only difficult to identify, but also the management of automated accounts takes advantages of different textual and programming resources to avoid being recognized and subject to deletion. For this reason, we searched for additional verification parameters with the contribution of statistical methods that apply rigor to the checking process.

The methodological improvements we made in this book aim to solidify the first version of the research, with new referrals that handle "different bots" in action on the web. With better resources to identify them, it is possible to recognize patterns, scattering groups, and characteristics of content replication on the web more precisely and accurately.

In short, it is easier to find them, even if they hide very well; but it is also easier to have tools to help protect the democratic debate on the internet.

BOTS

Bots are getting more and more human-like characteristics, which make their behavior more complex and hinder the identification of their virtual origin.

1.1 Background information

The first study on bots published by FGV DAPP identified the presence of automated accounts in political relevant moments in Brazil in recent years. For example, the general strike in April 2017, when more than 20% of interactions identified on Twitter among users in favor of the strike were caused by this type of account. Another episode in which the presence of bots was identified was during the 2014 presidential elections, when they generated more than 10% of debate.

The virtual world has allowed the adaptation of old defamation and manipulation political strategies in the public debate, now on a larger scale and without apparent transparency. The initial FGV DAPP study aimed to establish a first step to assess the communication modes and behavior among automated accounts.

As the use of bots has grown in sophistication and popularity, the identification of these automated accounts and their collection assumed increasingly important roles in the political debate during the 2018 elections. For that reason, studying the behavior of bots in the Twitter ecosystem and the interaction between human and bot users is imperative.

According to the purpose for which the bot was built, it will present a profile and behavior based on its interaction with human profiles. With the development of new techniques, bots are getting more and more human-like characteristics, which make their behavior more complex and hinder the identification of their virtual origin.

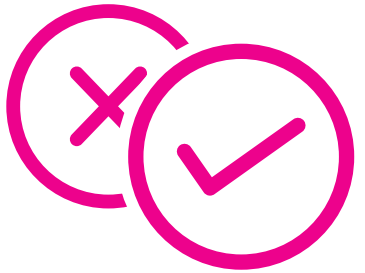
It is important to remember that bots can have the behavior that the programmer wishes. Despite the classification of types of bots, programmers are free to define the behavior of each bot. Technically, automated accounts are not able to act on their own. When the user uses a social network, it gives life to the account, and the same can be said for the applications that control the profiles.

One must keep in mind that the account alone is not a bot, but it presents automated behavior due to the application that controls it. If it is disabled, the profile will no longer be able to post anything. Therefore, we chose to use the term "bot" to indicate accounts that presented automated behavior in the period of the analysis.

Anyone can automate its own account. Some web tools, such as Twitter Feed, Social Oomph, Dlvr.it, and Tweetdeck, provide many valid resources for account management. The user can also develop an application for that with a public Twitter API (Application Programing Interface) and knowledge of a programming language, such as Ruby, Javascript, or Python.

The public Twitter API gives access to all actions an account can perform. Although there is an anti-spam policy in terms of API use, it is flexible enough to allow an account to post frequently and without risk of being blocked by Twitter when the content is not pornographic, with malicious links (which lead to pages that steal data, for example), etc. In the clause on the automation of profiles, Twitter even encourages the creation of automated accounts if they follow the following rules of use of the API:

Publish information to create useful solutions automatically, like the accounts @LeiSecaRJ and @ponterionit;



Run creative campaigns that auto-reply to users who engage with your content, as the accounts @WhatTheFare and @ArnaldoPode;

Build solutions that automatically respond to users in Direct Messages (DMs);

Try new things that help people (and comply with our rules), as @censusAmericans and @_grammar;

Make sure your application provides a good user experience and performs well — and confirm that remains the case over time.

"The Twitter Rules", which discourages two operations particularly: the development of automation without the use of the API, by means of a script in the website, for example; and profiles that bother other users with frequent or unsolicited messages. Applications developed based on the API are monitored constantly. When Twitter suspects that the application has violated any provision of the terms, it is blocked; to reverse the situation, an e-mail contact is required, describing the purposes of the software to Twitter evaluators.

According to Chu et al. (2010), Twitter accounts may be classified as bots, humans, or cyborgs. Bot accounts are fully controlled by an application, that is, they follow a specific code. Human accounts are controlled by a person without automation codes. Cyborg accounts have characteristics of human and bot accounts. They are controlled by humans most of the time, but they have some automation attributes, such as the campaigns "donate a tweet", in which the user allows an institution to publish with his/her profile, for example, or through account management tools, such as Tweetdeck.

Bots may have highly artificial behavior (spam) or be more similar to humans (social bots). Those called spam are usually used to develop marketing strategies, share videos, photos, and music, and boost the dissemination of a hashtag, link, or post. With clearly standardized behaviors, these accounts are easily detected by social networking platforms and run the risk of being blocked, unless they are placed in the White List. For example, they usually do not respond to a

1 All automation rules can be found at: https://help.twitter.com/en/rules-and-policies/twitter-automation

White List

Twitter allows users to warn the platform they will adopt an automated behavior, in the case of marketing strategies, for example. In this case, the profile becomes part of a list of accounts that are considered "reliable" even with a standardized behavior.

direct message. When they do, they give a simple and automatic response.

In contrast, social bots try to imitate human behavior on the social network, with profile and cover photos and bio, that is, with a complete profile. The tweeting pattern of social bots is also less coordinated than the spam bot.

When a large number of automated accounts is used for a coordinated action and controlled by a single actor – a botmaster –, they form a botnet (VAROL et al., 2017). Abokhodair et al. (2015) show how a botnet flooded discussions in Syria with hashtags unrelated to the civil war as a strategy to outshine the hashtags about the conflict. Apparently, the network was used to divert attention from the debate about the war.

As explained previously, the use of automated accounts on social platforms will increase not only its complexity, but also its number. Therefore, prompt and effective identification of these accounts and their collection have become increasingly important. However, the humanization of bot behavior is a challenge that tends to further hinder this process.

1.2 Bot typology

By observing the behavior of automated accounts and the way they communicate, it is clear they pursue specific goals and adopt strategies that depend on their purpose. That is, the account changes its behavior and text in order to achieve its purpose. For this reason, there are different types of automated accounts with a variety of features that depend on their specific goal.

As mentioned above, according to Chu et al. (2010), Twitter accounts may be classified as bots, humans, or cyborgs. Their study states that humans have a complex behavior regarding frequency in tweeting, while bots and cyborgs may present clearly patterned behaviors. By examining tweets posted by bots, researchers also found that a high proportion contains spam content. Other possible factors to detect automated accounts consider the external URL and the post generator used to tweet, as discussed by Ruediger (2017a).

Observations by Lee et al. (2011) characterize bots as content polluters and classify them into four groups:

Duplicate Spammers: they post nearly identical tweets with or without links;

Duplicate @ Spammers: they are similar to Duplicate Spammers, but they also abuse Twitter's @username mechanism by randomly inserting a legitimate user's @username, even if the user does not follow the bot;

Malicious Promoters: they tweet about business, marketing, etc. They are more sophisticated than other content polluters, and, among promoted tweets, they post legitimate tweets such as greetings;

Friend Infiltrators: they seem legitimate, but they abuse the reciprocity in relationships on Twitter. As relationships on the social network are strongly based on reciprocity – if user A follows the user B, then user B typically will follow user A –, this type of content polluter waits until it has a large number of followers to engage in spam activities.

However, not all bots have a polluting activity, and, according to Chu et al. (2012), some may be legitimate. The so-called "good bots", for example, publish news and update feeds. Usually companies use these types of harmless automated accounts on their social networks. Among legitimate bots we can find:

Chatbots: bots that simulate conversations with users with the use of natural language processing models to understand and create dialogs. An example is Siri/Cortana in our mobile devices. There are also different types of chatbots:

- **Service: bots that act as facilitators when one accesses services or information, makes transactions, seeks the customer service, etc.**
- **Commerce: bots that act as facilitators in the sales process, customer acquisition, generation of qualified offers, among other possibilities.**
- **Assistente Pessoal: bots that act as facilitators in education, finances, time management, weather forecasting, and self-confidence.**

Spider Bots: bots used to follow links and index the internet; they generate data that feed search engines.

Trading Bots: bots that operate in the financial market making transactions automatically.

Media and Entertainment Bots: bots that act as facilitators when one accesses news, services or information related to television, radio, news, media, means of communication in general, events, parties, celebrities, etc.

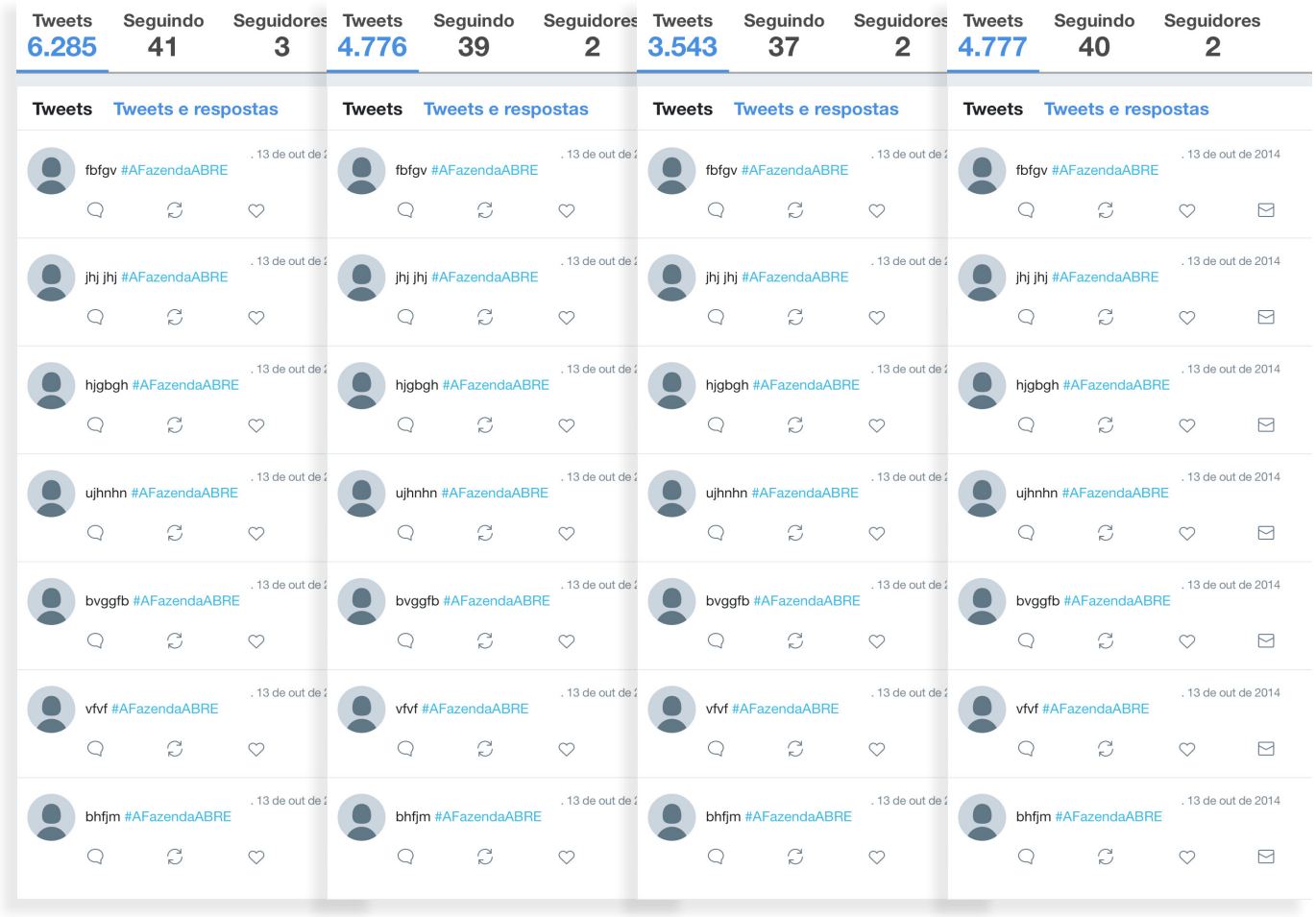
Chu et al. (2012) also mention different types of cyborgs, which can be classified as bot-assisted human or human-assisted bot.

However, the focus of this study is those behaviors that may influence the political debate on social networks; and manipulating the public opinion does not depend on the bot type, but on its kind of activity. Those potentially harmful behaviors to the public debate are exposed next.

Spam – Frequent hashtag posting: Bots use hashtags to inflate a debate. There are variations of this strategy; for example, some bots search for the most popular hashtags at any given time and tweet with random letters along with the hashtag found. The hashtag can also be chosen for a particular purpose.

This bot can inflate a debate about a hashtag even with messages without any content. The goal of this type of account is to create volume on a subject or specific person. In creating more volume, the bot causes an impression that the online community is talking significantly about this issue or person. These types of profiles are used to increase a person's image exposure and to divert attention from a subject to another (Figure 1).

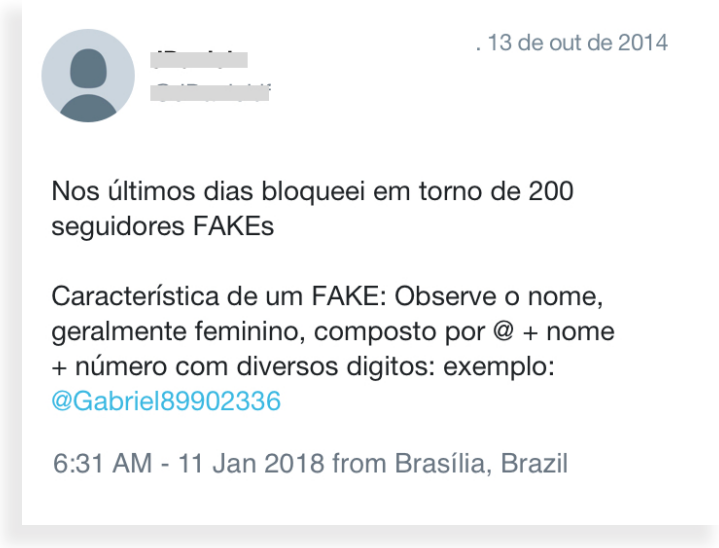
FIGURE 1. Post example



Source: elaborated by FGV DAPP, based on data from Twitter.

Spam bots are not built to interact with a human profile, as they are easily identified as an automated account. Human users commonly report the fake accounts to Twitter, as the example bellow:

FIGURE 2. Post example



Source: Twitter

Disseminators

Retweeting from webpages:

Some bots follow webpages and retweet their posts. According to DAPP's studies, human users do not typically interact with this type of bot, and they usually do not have a large number of human followers. Unlike humans, disseminators follow a determined patter when they retweet from webpages. For example, the programmer can choose which webpages will be retweeted, while humans choose arbitrarily. Furthermore, the retweet frequency has similar patterns, while retweets from human accounts are more random.

FB

Follow back:

Looking for more exposure for their posts, some accounts use hashtags and keywords to signal they will follow back those who follow their profile. With this method, an account can get thousands of users "organically". An account can reach over a thousand followers with some tweets and likes. A bot that aims to propagate news or quickly gain audience can use such hashtags (like #SegundaDetremuraSDV (#MondayDetremuraFB) and #fb) to build a large network of followers who will be exposed to its posts. Another way to get followers fast is to hire marketing agencies specialized in selling likes, followers, and retweets.

Influencers

Tweeting original messages:

This bot simulates a human trying to become an influencer by creating text and actively participating on Twitter. Human users often enjoy and retweet this type of profile, since, if done right, it may be difficult to identify it as a bot.

1.3 Refined methodology

Currently, several researchers have sought to develop bot detection methods on social networks. Different approaches have been used, such as machine learning and text analysis. We will present comparative results on bot identification in six databases using three different methods.

The first method is based on an article by Varol et al. (2017), and it is called BotOrNot. It analyzes Twitter accounts through a machine learning algorithm that verifies recent data history of an account – including mentions – and calculates the likelihood of this account to be a bot. Currently, Twitter has an API that allows the use of this method for checking accounts automatically.

The second method, described in articles by Chavoshi et al. (2016a; 2016b), verifies the existence of bots in a group of Twitter accounts through a post correlation analysis for each possible pair of users within this group of accounts. There is also an API that enables the automatic use of this method to verify a set of Twitter accounts.

The third method, called DAPP methodology (RUEDIGER, 2017a), developed by FGV DAPP, uses a metadata called generator – information indicating the device that generated the tweet – and the frequency of tweets in the same second by the same account to classify it as a bot.

Based on these three methods, FGV DAPP analyzed six tweet databases collected in moments of intense political discussion: the first and second round of the 2014 Brazilian presidential elections (2014 1R and 2014 2R, respectively), the 2016 pro-impeachment demonstrations, the 2016 municipal elections in São Paulo (Municipal Elections SP), the general strike in Brazil on April 28, 2017 (General Strike) and the Senate vote on the Labor Reform on July 11, 2017 (Labor Reform).

1.3.1 BotOrNot

The BotOrNot method (VAROL et al., 2017) is based on users' relationships to search, via the Twitter Search API, the last

200 public tweets of each user and the last 100 public tweets that mention him/her. This method considers more than a thousand variables, separated by classes, to assess the existence of bot characteristics in each account. The analysis shows, for each class, the likelihood of the account to be a bot and, considering all the classes, the universal probability (UP).

Because of this dynamic, this process is slow, taking about 15 minutes to analyze a database of 180 users, for example. Therefore, we opted to collect samples of users in each of the cases studied. As suggested in the literature, it is important to ensure that the sample is representative, that is, with the same basic features of the phenomenon studied on average (BUSSAB and MORETTIN, 2017). Therefore, we adopted a Simple Random Sampling plan (SRS) without replacement. Such method selects units for the sample considering equal selection probability. To set the sample size that would be enough to ensure a maximum of 5% error and 99% confidence, we estimated proportion using SRS, which returns the required sample size conservatively. See below the results of this method applied to samples of the six themes:

Table 1 shows the result obtained with the method BotOrNot in the six themes. The algorithm did not verify all the accounts in the sample analyzed. This is because some users may have had their accounts suspended by Twitter itself or may have deleted them.

TABLE 1: BotOrNot Results

			Verified		Non Verified		Identified Bots (%)				
	Total of Account	Sample	total(	%)	total(	%)	UP 0,5	UP 0,6	UP 0,7	UP 0,8	UP 0,9
2014 Elections - 1st R.	320.091	663	646	97,44%	17	2,56%	4,83%	3,02%	1,51%	0,60%	0,30%
2014 Elections - 2nd R.	286.452	663	648	97,74%	15	2,26%	5,73%	2,56%	1,51%	0,45%	0,15%
Pro-Impeachment	383.469	662	581	87,76%	81	12,24%	4,83%	1,66%	1,06%	0,45%	0,00%
São Paulo Elections	251.423	651	631	96,93%	20	3,07%	5,38%	2,76%	1,23%	0,77%	0,15%
General Strike	383.469	663	627	94,57%	36	5,43%	4,37%	1,36%	0,45%	0,00%	0,00%
Labour Reform	76.614	659	631	95,75%	28	4,25%	4,86%	1,67%	1,06%	0,46%	0,00%

Source: elaborated by FGV DAPP

Despite recent attempts to identify bots with machine learning, Cresci et al. (2017) indicate that is not the ideal technique to develop this task, because it cannot detect new bots. Such phenomenon occurs because of the contrast between the constant evolution of bots, which try to circumvent detection mechanisms, and the rigid structure of building a machine-learning model, which depends on a set of fixed data to find patterns to classify accounts.

Due to the evolution of bots, automated accounts look more and more like ordinary users; then, it is difficult to detect them, even by humans (GREW et al., 2017). Algorithms that were the state of the art in the past are no longer able to detect all types of bots. This paradigm shift in the detection of automated accounts calls for new solutions that implement different techniques.

1.3.2 Correlation

We designed this methodology inspired by a method by Chavoshi et al. (2016a; 2016b) with the goal to detect not only suspected automated accounts, but also those called botnets (groups of suspicious accounts that may constitute a network). The procedures for analysis compare the time behavior of each user, that is, they calculate the correlation between the time stamps of each account activities with all the others, considering a fixed time interval. Thus, it is possible to capture possible bots working simultaneously, which have activities at the same intervals.

We divided each dataset used for comparing the methods into divisions with a fixed time interval of one hour (1h) or half an hour (0.5h), depending on the amount of accounts and their activities. The analyses applied to each division considered the Pearson correlation (COR) as a method to calculate correlations with time intervals of 20 seconds according to Chavoshi et al. (2016B), disregarding users that showed an activity lower than two tweets in each period of 1h or 0.5h. Regarding the cut-off point stipulated by Chavoshi et al.

(2016b), we only considered accounts with correlation greater or equal to 0.995. However, for this study we evaluated the results for different cut-off points, with a consistent cut of 0.9.

Since there is a great probability that a high correlation between accounts with a low number of activities is a coincidence, we performed a recurrence analysis of correlations in the different periods analyzed in an attempt to eliminate this coincidence. To do so, we considered the frequency of each pair of highly correlated users, as well as the number of distinct pairs of each user. We eliminated all pairs that had a lower score in the third quartile.

In new selected samples, we applied the correlation analysis method and, for comparative purposes, the methodology developed by BotOrNot. Tables 2 and 3 show the results.

TABLE 2. Correlation results in the sample

	Total of Accounts	Sample	Verified total(	Verified %)	Identified Bots (%)
2014 Elections - 1st R.	320.091	663	658	99,2%	2,87%
2014 Elections - 2nd R.	286.452	663	443	66,8%	7,69%
Pro-Impeachment	383.469	662	277	41,8%	5,29%
São Paulo Elections	251.423	651	611	93,8%	5,68%
General Strike	383.469	663	583	87,9%	20,10%
Labour Reform	76.614	659	613	93,0%	0,76%

Fonte: elaborado pela FGV DAPP.

BotOrNot results show that the higher the probability of an account to be automated, the lower the proportion of accounts identified. With the results of the correlation methodology, the proportion of bots detected was higher.

TABLE 3. BotOrNot results in the sample

	Total of Accounts	Sample	Verifie d total	Verifie d (%)	Non-Verified total	Non-Verified (%)	Identified Bots (%) UP ≥ 0,5	UP ≥ 0,6	UP ≥ 0,7	UP ≥ 0,8	UP ≥ 0,9
2014 Elections - 1st R.	320.091	663	633	95,48%	30	4,52%	5,13%	1,81%	0,30%	0,00%	0,00%
2014 Elections - 2nd R.	286.452	663	646	97,44%	17	2,56%	6,64%	2,26%	1,21%	0,30%	0,00%
Pro-Impeachment	383.469	662	576	87,01%	86	12,99%	4,38%	1,66%	0,60%	0,00%	0,00%
São Paulo Elections	251.423	651	632	97,08%	19	2,92%	11,37%	5,38%	0,31%	0,00%	0,00%
General Strike	383.469	663	615	92,76%	48	7,24%	3,17%	1,36%	0,15%	0,00%	0,00%
Labour Reform	76.614	659	647	98,18%	12	1,82%	2,43%	0,30%	0,00%	0,00%	0,00%

Source: elaborated by FGV DAPP

1.3.3 DAPP Methodology

The methodological procedure of bot identification here named DAPP methodology can be briefly presented based on two joint assessment approaches described in the study Bots, social networks, and politics in Brazil : (1) use of platforms that allow the automatic generation of tweets; (2) "non-human" frequency of publications (RUEDIGER, 2017a).

The first approach starts with the manual verification of the metadata called generator, which, according to Twitter Data Format , represents a platform used for tweet posting. Then, the platforms that allow the automatic production of tweets are marked. Later, for each account in this database, we verify the proportion of tweets produced using these platforms and mark as bots those accounts that produced at least 10% of their tweets this way.

The second approach finds the time interval between tweets generated by each account in this database. Accounts that tweeted at least twice consecutively within a time interval shorter than one second are identified as bots.

Table 4 summarizes the results of this method applied to the six cases analyzed:

TABLE 4.
DAPP Methodology Results

		DAPP TOTAL						DAPP SAMPLE					
		Verified		Non-Verified		Identified Bots			Verified		Non-Verified		Identified Bots
	Total of Accounts	total(	%)	total(	%)	total(	%)	Sample	total(	%)	total(	%)	total(	%)
2014 Elections - 1st R.	320.091	320.091	100%	0	0%	4.863	1,52%	663	663	100%	0	0%	19	2,87%
2014 Elections - 2nd R.	286.452	286.452	100%	0	0%	6.006	2,10%	663	663	100%	0	0%	19	2,87%
Pro-Impeachment	383.469	383.469	100%	0	0%	5.822	1,52%	662	662	100%	0	0%	11 1	,66%
São Paulo Elections	251.423	251.423	100%	0	0%	1.386	0,55%	651	651	100%	0	0%	33	5,07%
General Strike	383.469	383.469	100%	0	0%	5.184	1,35%	663	663	100%	0	0%	6	0,90%
Labour Reform	76.614	76.614	100%	0	0%	724	0,94%	659	659	100%	0	0%	8	1,21%

Source: elaborated by FGV DAPP

2 Available at: http://dapp.fgv.br/wp-content/uploads/2017/08/Robos-redes-sociais-politica-fgv-dapp.pdf
3 Available at: http://support.gnip.com/sources/twitter/data_format.html

FGV DAPP restates that the process of identifying and checking evidence of bot activity should not obey solely volumetric parameters for reasons of methodological precision. That is, the simple identification of a remarkable increase in mentions to a particular term, keyword, or phrase does not necessarily indicate engagement of bot profiles, and rarely this is the most relevant metric to investigate automated posts about any subject in any period analyzed.

This is due to the nature of the public debate on social networks, which follows themes, actors and points of interest of civil society according to the changes and developments of the debate, which is always variable and susceptible to the relevance of what is immediate, current, and newly ascendant.

The debate on a particular theme may cover different aspects, and not all of them are necessarily relevant to the study in question. In this sense, we analyzed the overall debate to build a linguistic approach considering only the semantic field used for the debate in question (RUEDIGER, 2017b).

The identification of bot activity must be performed after database collection (verified and filtered). Therefore, information provided by FGV DAPP comply with these search criteria; it is not causally related only to the possible increase of mentions to the subject in any period of debate.

CASE STUDIES

- Based on the refined methodology of detection and analysis of automated profiles, FGV DAPP has examined high-profile events in the Brazilian public debate or of strategic interest, such as the electoral processes of Brazil's neighboring countries.

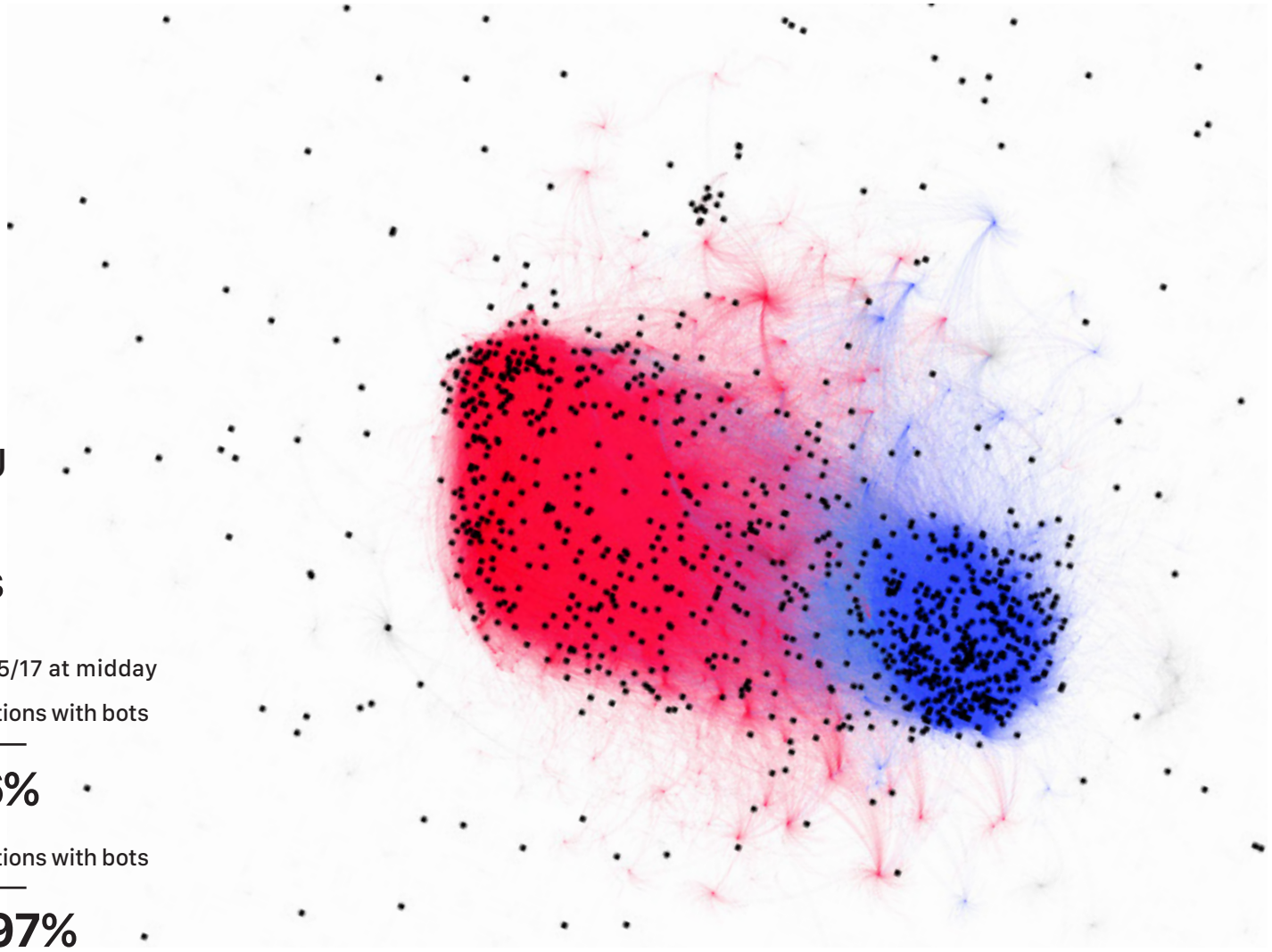
- Since the publication of the first study on the subject, Bots, social networks, and politics in Brazil, in August 2017, we identified a significant presence of bots interfering illegitimately with the public debate at least four times: (1) the controversy involving the Queermuseu exhibition, in Porto Alegre; (2) the pre-election debate in Paraguay prior to the 2018 elections in April; (3) the discussions on eight Brazilian political actors in a pre-election context in the late 2017; and (4) the trial of former President Lula in the Federal Regional Court of the 4th Region (TRF-4), in the end of January 2018.

2.1 The Queermuseu exhibition

The survey by FGV DAPP on the cancellation of the exhibition "Queermuseu - cartographies of difference in the Brazilian art" (Queermuseu - cartografias da diferença na arte da brasileira), in Porto Alegre, had 778,000 posts on Twitter from 12 AM on September 8 to 12 PM on September 15, 2017. The map of interactions of this database identifies two opposed groups in the debate: blue, those who were against the exhibition; red, those who advocated the permanence of the exhibition.

Profiles acting automatically in the debate were identified according to the platform used to post tweets or the time pattern of tweets from the same account. About 12.97% of the blue cluster interactions and 7.16% of the red cluster interactions indicate bot activity. In total, 8.69% of interactions were identified as coming from bots in the general discussion about the exhibition.

The graph also shows a more diffuse red group, which reveals a wider range of discussion and diversity of arguments. The blue group, quite concentrated, shows more unit in their arguments. Quantitatively, however, they are nearly equivalent in number of accounts and interactions.



THE QUEERMUSEU EXHIBITION

777.930 Tweets

Period of data collection:
09/08/17 at 00:01 a.m. to 09/15/17 at midday

Pro-exhibition	Interactions with bots
50,48%	**7,16%**

Against the exhibition	Interactions with bots
35,66%	**12,97%**

Source: elaborated by FGV DAPP

2.2 The pre-electoral debate in Paraguay

On December 17, 2017, the Paraguayans went to the polls to begin the process to the 2018 elections in April. Simultaneous party primaries defined the candidate of each party who would run the presidential election. These internal disputes, especially in the Colorado Party (Asociación Nacional Republicana – ANR) and in the Authentic Radical Liberal Party (Partido Liberal Radical Auténtico – PLRA), were already anticipating the latent concern in the 2018 presidential election cycle in Latin America: illegitimate interferences with the public debate on the web, especially the influence of bots.

FGV DAPP mapped the debate on the Paraguayan party primaries that took place on Twitter between November 1 and 30, 2017 to identify bots in this discussion. About 270,000 tweets related to the debate were analyzed, with the occasional presence of bots especially in the debate between supporters of the Colorado Party and president Horacio Cartes at the time.

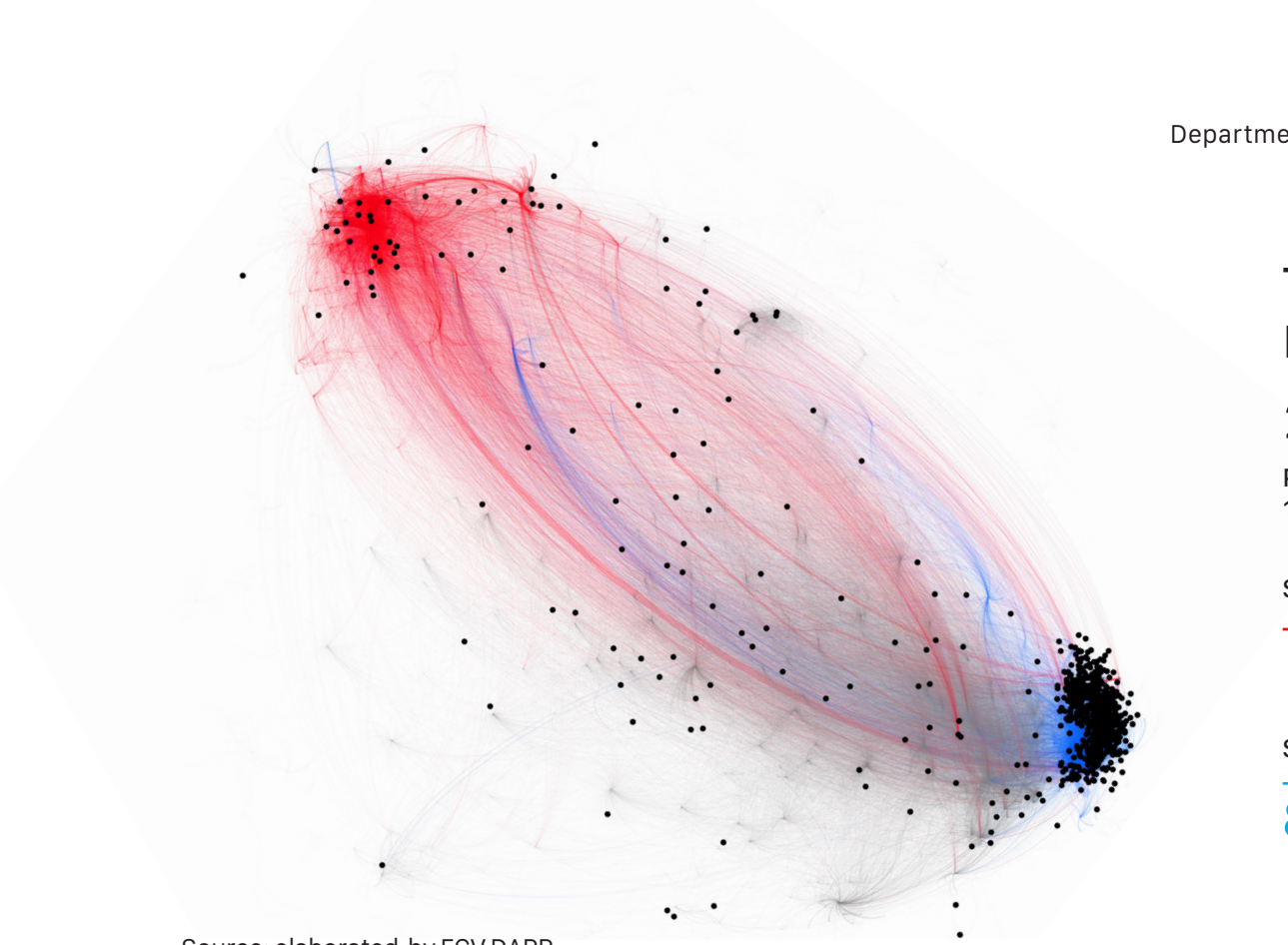
With the withdrawal of Cartes from the run for re-election after a failed attempt to implement a constitutional amendment to allow his reinstatement, the Colorado Party led its members to choose between a continuity candidate (former Minister Santiago Peña) and Mario Abdo Benitez, in addition to other currents within the party. The following graph represents the debate on the party primaries and shows a clear polarization between the groups supporting Peña (in red, representing 16% of the debate) and Abdo (in blue, with 9% of the discussions).

Other accounts analyzed focus on media profiles, with about 70% of the total. Among the liberals, the decision of the head of the ticket was between the president of the party ticket, Efraín Alegre, and other fellow party members. On the networks, the debate about them does not have a representative concentration of profiles (clusters).

The presence of bots

Some automated accounts with the purpose of promoting candidates were found in the analysis. Bots had electoral purposes and worked to increase engagement and a possible spread of tweets in favor of a candidate or a wing of the party.

The accounts highlighted in **black** were classified as bots due to the consecutive production of at least two tweets within an interval shorter than one second and the use of platforms that automate account management (DAPP Methodology). We identified 4,977 bots promoting directly and indirectly 14% of the total interactions on the network analyzed; 13.35% of interactions in the red group, and 49.87% in the blue group.



Source: elaborated by FGV DAPP.

THE PRE-ELECTORAL DEBATE IN PARAGUAY

276.097 Tweets

Period of data collection:
11/01/17 at 00:01 a.m. to 11/30/17 at 11:59 p.m.

Support for Santiago Peña

15,90%

Interactions with bots

13,35%

Support for Mario Abdo Benitez

8,85%

Interactions with bots

49,87%

The map of interactions shows the support group to Peña greatly detached. Despite the polarization, Abdo's support group is strongly connected with the gray group, which comprises the local press and profiles not aligned to the dispute, representing about 70% of the debate.

This detachment appears to have been partially caused by different tactics of using bots. While in the red group automated accounts retweet almost exclusively Peña's account, there is a distribution of automated retweets in the blue cluster among different accounts. The most shared content in this group were publications of accounts with 8 to 100 followers, created in September 2017, but they reached about 300 retweets each.

2.3 The trial of former president Lula

The former president Lula's trial on January 24 by the Federal Regional Court of the 4th Region (TRF-4) generated 1.21 million mentions on Twitter in Brazil in only 24 hours. The map of interactions shows this debate was divided into three main groups: blue, in favor of the conviction of former president, with about 35.4% of accounts in the debate; red, representing 44.1% of accounts, formed mostly by actors against Lula's sentence; green, representing 15.3% of accounts, comprising mostly humorous profiles and tweets.

FGV DAPP identified 8% of bots in the red group and 6% in the blue group. In both cores of automated posts the predominant retweets were posted by the main influencers of each core – not original tweets made by bots – and vehicles whose production is openly party supporter, but not necessarily committed to the dissemination of fake news and information.

In the red cluster, we collected 37.201 tweets from profiles whose behavior presents automation evidence. The most shared posts within the cluster are heterogeneous regarding content and authors (non-bot), including official accounts of former president Lula (@LulapeloBrasil), PT parliamentarians and influencers aligned to the left wing that participate in most of political debates on Twitter.

For carrying forward comments and links from "real" influencers, the content of retweets is largely similar to that observed in the general debate on the trial. All most influential tweets shared by bots are news and expressed opinions, without an (relevant) automated account responsible for initiating the spread of content, but replicating it.

THE TRIAL OF FORMER PRESIDENT LULA

1.298.246 Tweets

Period of data collection:
01/24/2018 at 00:01 a.m. to 01/25/2018 at 00:01 a.m.

Against the conviction

42,12%

Interactions with bots

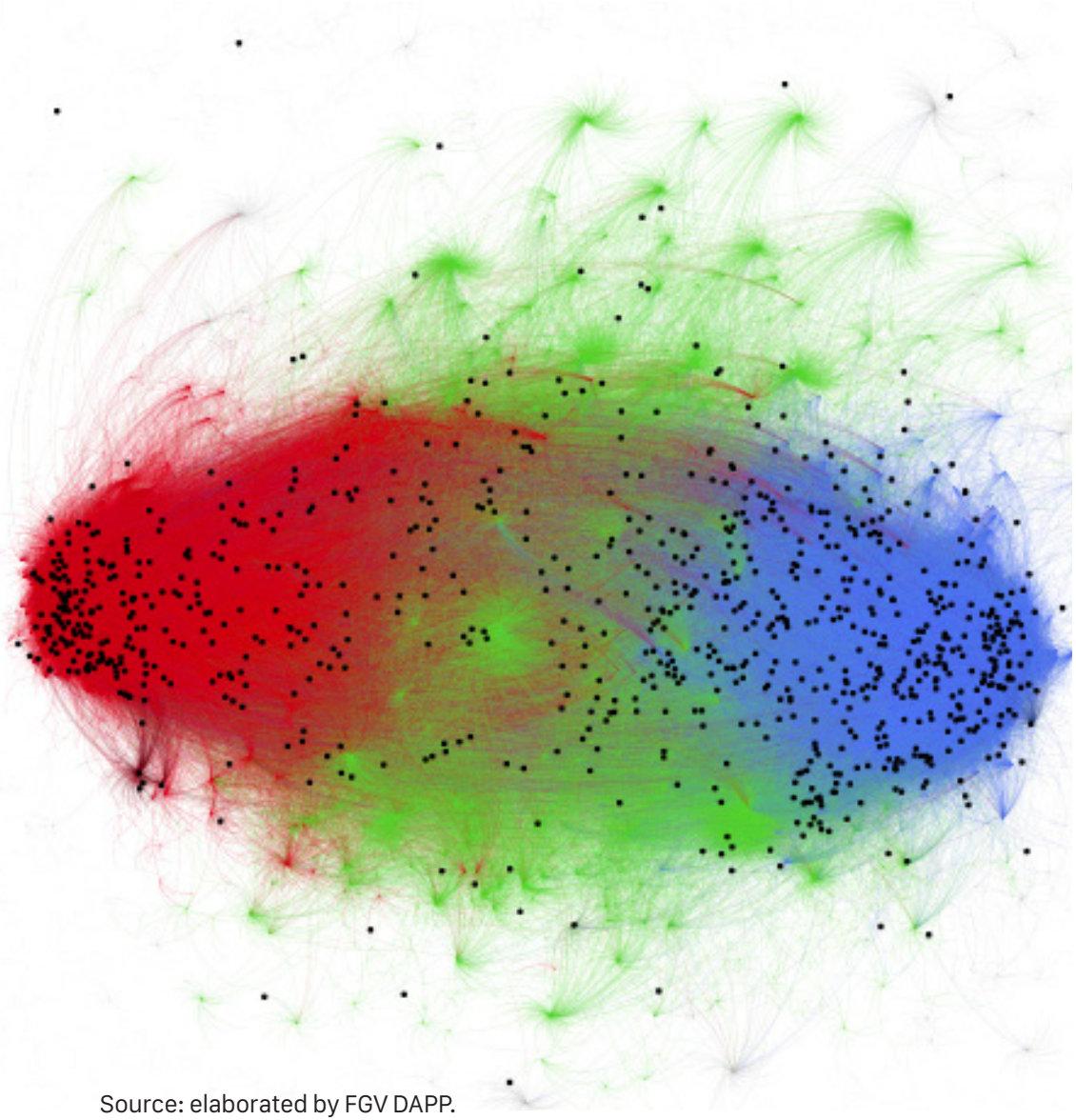
8,02%

For the conviction

35,38%

Interactions with bots

6,13%



Source: elaborated by FGV DAPP.

Among the links most shared by bots, there is low participation of the traditional press, with the exception of an article published in the newspaper Folha de São Paulo by one of Lula's lawyers, Cristiano Zanin Martins, shared 39 times. Among the ten most shared links in the red cluster, four are notes from the official PT website and three are from the opposition blog Brazil 247.

Among the automated profiles opposed to Lula (49,433 tweets), there is a greater dispersion of retweets by actors belonging to different political groups, but gathered in the same cluster due to their common rejection to PT and to the former president. Not all profiles manifest their express defense of some electoral actor, and in many debates they interact in different clusters.

Regarding the largest sharing of links, posts from the webpage O antagonista predominate, with six links to the webpage's posts among the 20 most shared links. There are also links to news webpages of the traditional press, such as O Estado de São Paulo and G1. Curiously, the most shared link (62) is from 2010 and comes from the account of the former president Dilma Rousseff, in which she manifests herself in favor of the Clean Record Law – a law that makes Lula, once convicted, ineligible for the 2018 elections.

Among the 20 links most shared by bots in the blue group, two are websites with characteristics of false information dissemination. One is from the webpage Imprensa viva, which simulates the journalistic language, but without canonical attributes of news production (source checking, data verification, etc.). The link shared describes the supposed identification of a pro-Lula protester accused of stealing a mobile from a reporter from the TV show Fantástico. The other link is a video from the webpage News atual, unavailable online, in which the actress Lucinha Lins aledgedly criticizes the musicians Chico Buarque and Caetano Ve-

loso and the actor Wagner Moura, who declared support for Lula.

Therefore, when compared to the general debate on the trial, the engagement motivated by automated publications does not differ from the content manifested by groups in favor and against Lula. This is so that the most retweeted tweet by bots in the opposition group to the former president is the same post with most engagement among Twitter posts on the subject: @jqteixeira quips an image of "a left-wing militant" as a History teacher who lives with his mother and delegitimize the evidence presented by the judges.

Bots operate primarily in disseminating influencers already established in each cluster and links to posts of vehicles aligned to each group. Notable exceptions are the posts about an alleged "mobile thief" and the video by Lucinha Lins, whose veracity is not subject to verification by regular mechanisms of professional journalism.

The use of automated accounts to perpetuate dubious information is one of the main threats to the integrity of the public debate on the web, and it was present, although not prominently, in the discussion on Lula's trial.

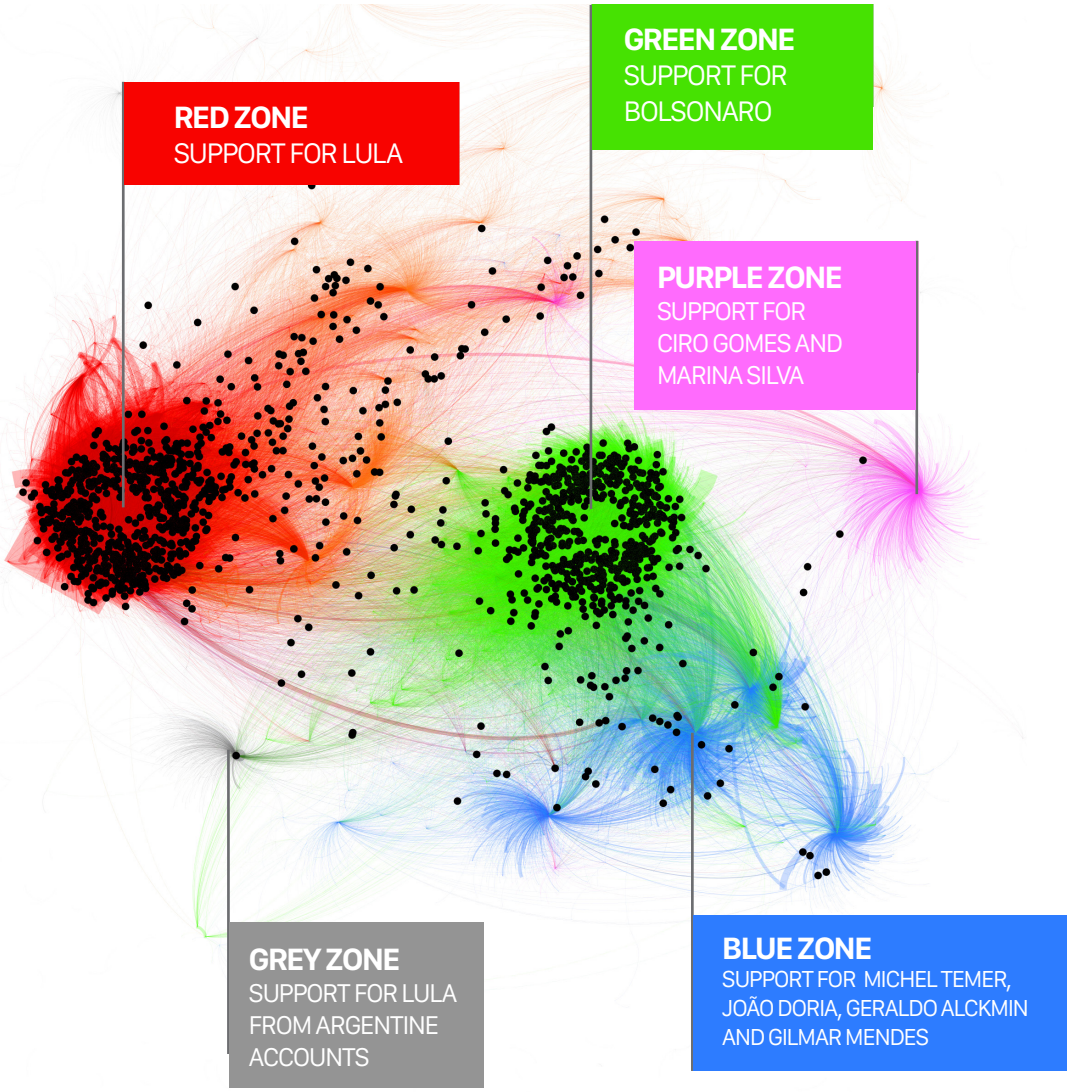
2.4 Political Actors

In this case study, we experimentally included a third criterion for bot identification in addition to the two ones used in the other cases. To that end, we collected the debates that happened on Twitter over a period of one month (November 18 to December 17, 2017) about eight political actors: Lula, Jair Bolsonaro, Ciro Gomes, Marina Silva, Michel Temer, João Doria, Geraldo Alckmin and Gilmar Mendes. After that, we applied the DAPP Methodology with the addition of the third criterion:

1. **First criterion: we verified the use of tweet-generating platforms as described in the study "Bots, Social Networks and Politics in Brazil" (RUEDIGER, 2017a);**

2. **Second criterion: we observed the non-human time of the publications;**

3. **Third criterion: we verified the accounts that produced exactly the same tweets using the so-called correlation method through identical messages, a strong indication of account automation that also allows the verification of the botnets operating together.**



Source: elaborated by FGV DAPP.

MAP OF INTERACTIONS RELATED TO THE POLITICAL ACTORS

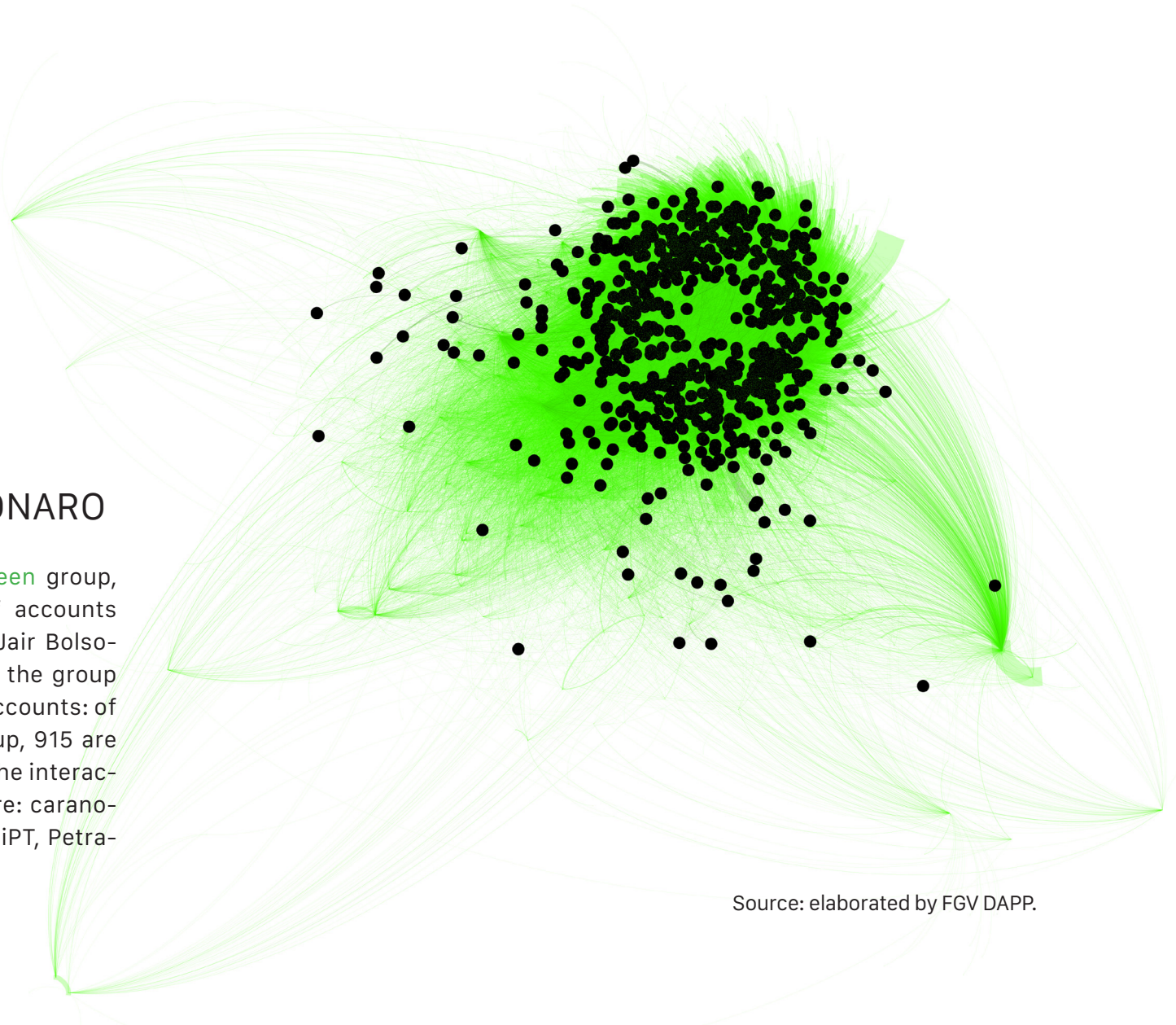
731.844 Tweets

Period of data collection:
11/18 at 00:01 a.m. to 12/17 at 11:59 p.m.

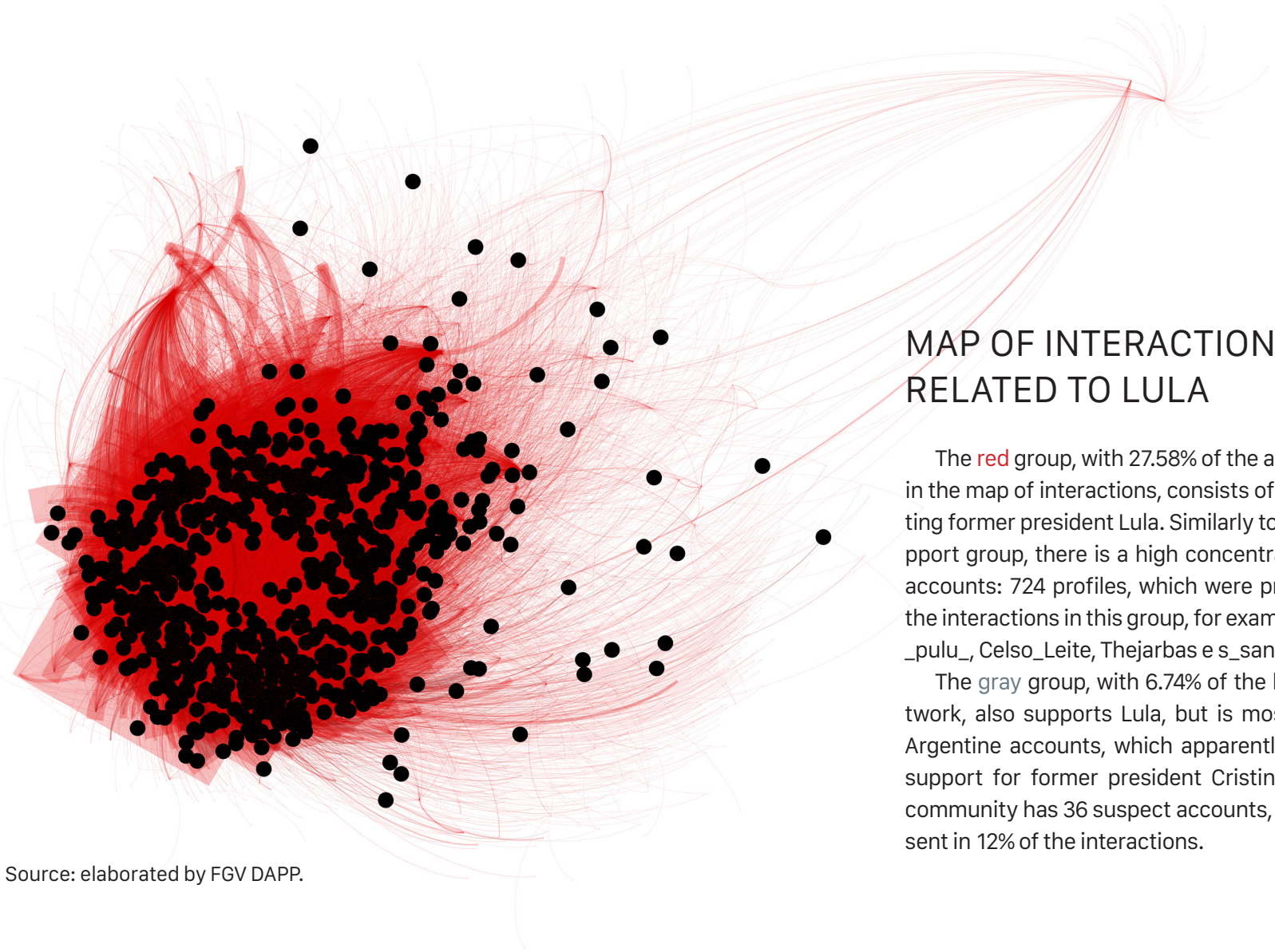
We then selected the tweets created by accounts that generated at least five tweets during the analyzed period and verified the percentage of correlation between each potential pair of accounts (the proportion of identical messages made by the two accounts). The minimum of five tweets per account was used in order to avoid false positives, since the possibility of correlation is inversely proportional to the volume of tweets. After this section, we analyzed 731,844 tweets (81.85% of the total collected).

MAP OF INTERACTIONS RELATED TO JAIR BOLSONARO

The main characteristic of the green group, with 51.96% of the total amount of accounts analyzed, is the support for deputy Jair Bolsonaro. As demonstrated by the graph, the group has a high concentration of suspect accounts: of the 101 thousand profiles in the group, 915 are suspects and were present in 33% of the interactions in the group. Some examples are: caranovanocongr, MELBOURNE6712, Ary_AntiPT, PetralhaKiller e ANTIPETISTA1.



Source: elaborated by FGV DAPP.

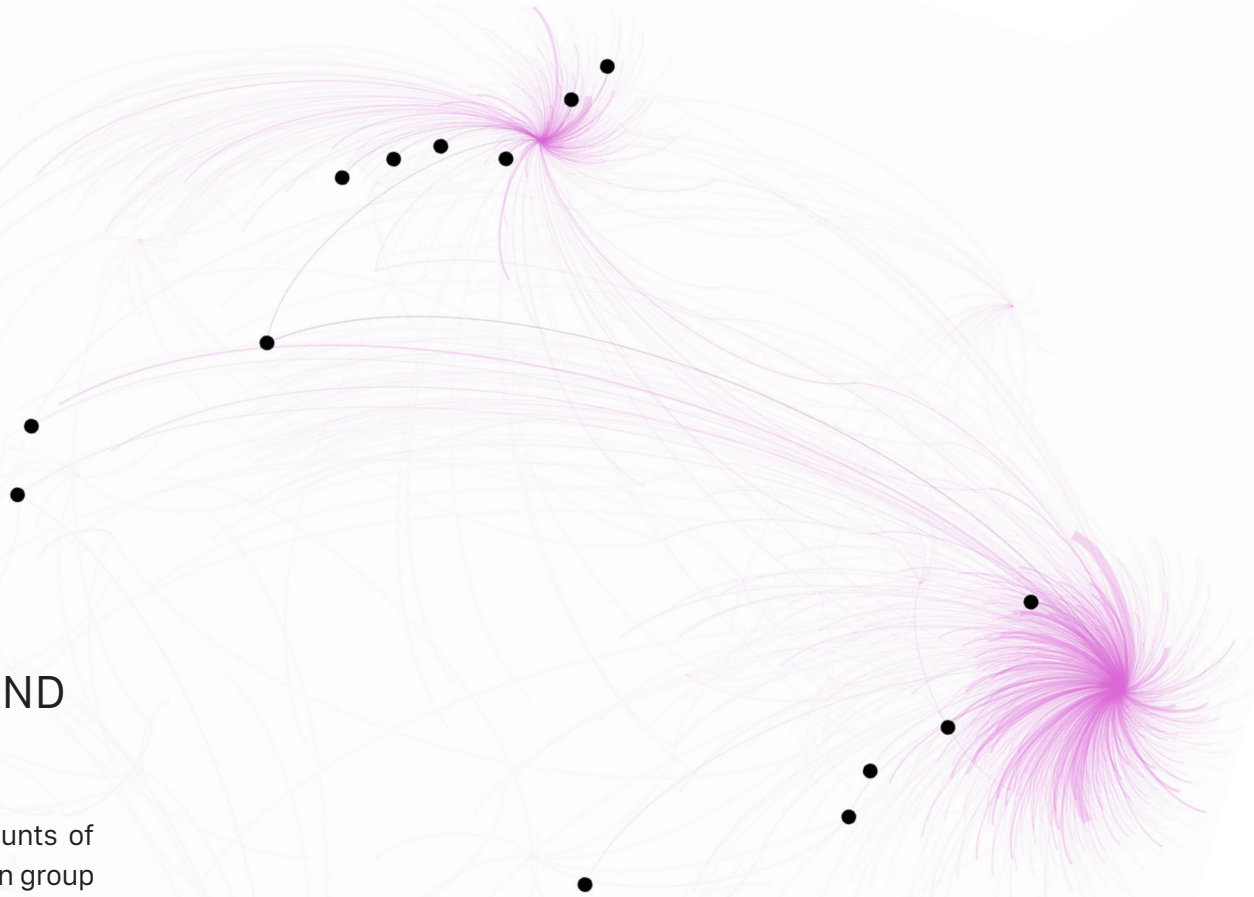


Source: elaborated by FGV DAPP.

MAP OF INTERACTIONS RELATED TO LULA

The red group, with 27.58% of the accounts present in the map of interactions, consists of profiles supporting former president Lula. Similarly to Bolsonaro's support group, there is a high concentration of suspect accounts: 724 profiles, which were present in 27% of the interactions in this group, for example: VieiMichele, _pulu_, Celso_Leite, Thejarbas e s_santos123.

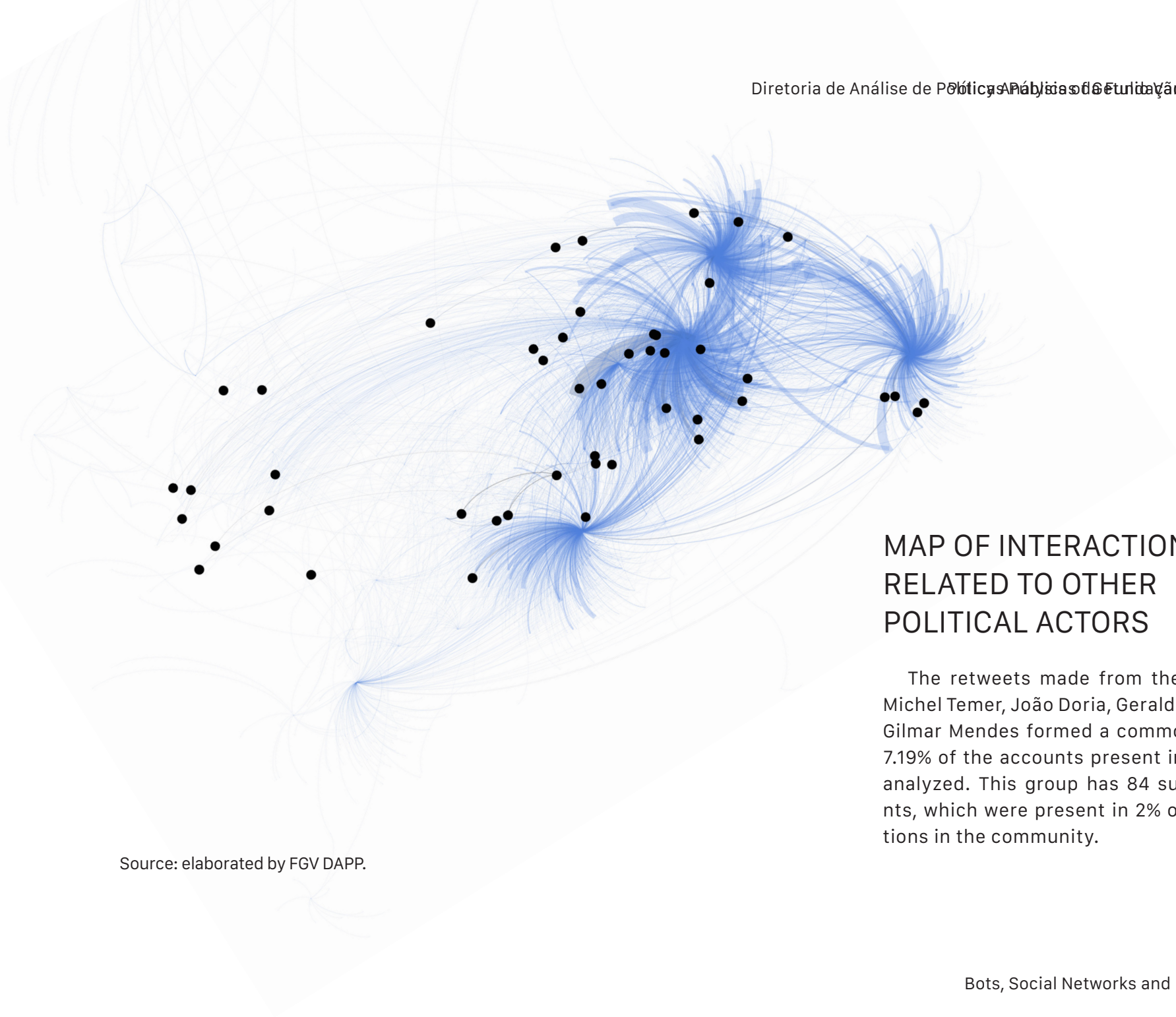
The gray group, with 6.74% of the knots in this network, also supports Lula, but is mostly made up of Argentine accounts, which apparently maintain their support for former president Cristina Kirchner. This community has 36 suspect accounts, which were present in 12% of the interactions.



MAP OF INTERACTIONS RELATED TO CIRO GOMES AND MARINA SILVA

The retweets made from the official accounts of Ciro Gomes and Marina Silva formed a common group for both of them. Composed of 3.2% of the accounts in the network, the group had 26 suspect accounts, which were present in 3.5% of the interactions.

Source: elaborated by FGV DAPP.



Source: elaborated by FGV DAPP.

MAP OF INTERACTIONS RELATED TO OTHER POLITICAL ACTORS

The retweets made from the accounts of Michel Temer, João Doria, Geraldo Alckmin and Gilmar Mendes formed a common group with 7.19% of the accounts present in the network analyzed. This group has 84 suspect accounts, which were present in 2% of the interactions in the community.


3

IMPLICATIONS FOR THE POLITICAL DEBATE IN THE 2018 ELECTIONS

> Due to its power to manipulate the public debate, the presence of bots should be identified, especially in times of great political importance.

IMPLICATIONS FOR THE POLITICAL DEBATE IN THE 2018 ELECTIONS

Due to its power to manipulate the public debate, the presence of bots should be identified, especially in times of great political importance. Only by analyzing bots we can distinguish which situations are real or forged in the debates on social networks, which is extremely important in an intensely polarized political context as the current one.

However, besides the identification of bots, responsible bodies must ensure the fairness of the electoral process by recognizing this type of mechanism and developing measures for full transparency of its use by political actors. Not every bot aims to tamper the debate; therefore, its use does not need to be discouraged, since it complies with parameters that favor its accountability.

It is a challenge to the Superior Electoral Court to observe the virtual world as a scenario that allows the amplification of political strategies of defamation, manipulation, and even campaign. An example of this difficulty of the Electoral Justice was clear during the trial of representations by the Electoral Attorney General against Lula and Jair Bolsonaro in the TSE.

On December 5, the Court understood that there was no early campaign in the dissemination of videos on YouTube by the two political actors regarding the 2018

elections. The trial was considered a barometer of what would happen in the coming months, but it showed that there is not an understanding in the Court on how early campaign happens on the networks. Even members of the Federal Supreme Court that are part of the TSE expressed disagreement. The votes were 4 to 3 in the case of Lula, and 5 to 2 in the case of Bolsonaro.

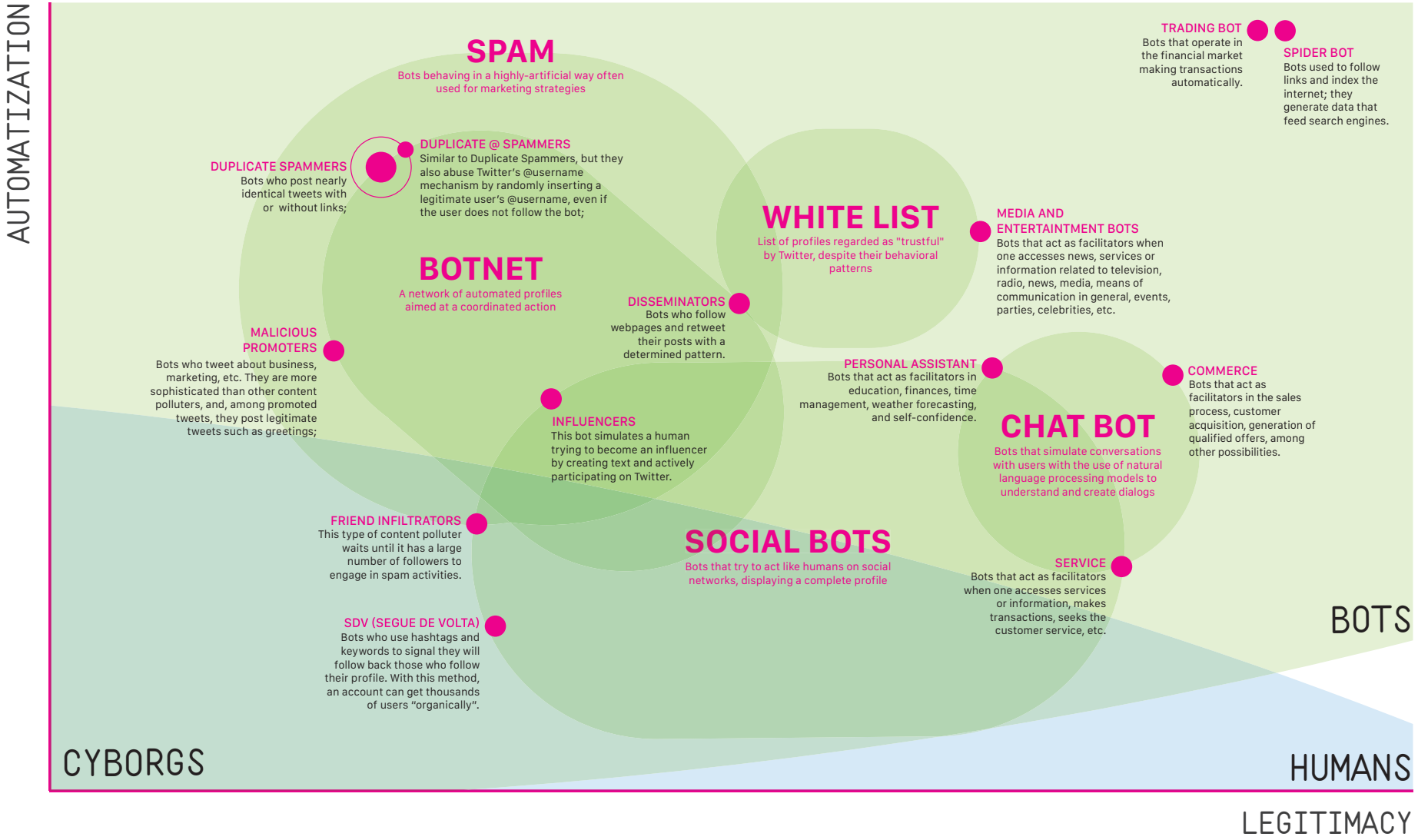
According to the law, the campaign could only start on August 15, 2018 under penalty of a fine of R$ 5,000.00 to R$ 25,000.00 for anyone who violates this restriction. However, according to the Law on Elections (Law No 9504/1997), a campaign only happens when candidates asks for votes. Therefore, mentions to an application, exaltation of a pre-candidate's qualities, participation in events and interviews do not qualify as early campaign.

Aware that the law restrains pre-candidates from asking for votes, Lula and Bolsonaro did not explicitly do so on the networks. However, in addition to videos, both presidential hopefuls intensely published their travel schedules across the country on their social networks and posted regularly about their electoral intents.

Also, there was a time gap of about six months between the release of videos on the network, the attorney's request for representation, and the Court's sentence. This was a crucial time, as, in case the claims were granted, there would have been six months of irregular campaign.

The difficulty of identifying the authorship of publications also appeared as a defense weapon by the presidential hopefuls. Although this factor did not lead to the denial of claims by the Court — but the fact that the presidential hopefuls did not ask for votes explicitly —, Lula's lawyer said that the video was made and posted by supporters and without the former president's knowledge. This defense strategy can be easily repeated — either true or not —, since it can at least delay a decision.

Bot Glossary



References

ABOKHODAIR, N.; YOO, D.; e McDONALD, D. W. Dissecting a social botnet: Growth, content and influence in twitter. In: 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, Vancouver, British Columbia, Canada, 2015, 839–851.

BUSSAB, W.O.; MORETTIN, P. A.. Estatística Básica: 9. ed. São Paulo: Editora Saraiva, 2017.

CHAVOSHI, N.; HAMOONI, H.; MUEEN, A. Identifying Correlated Bots in Twitter. International Conference on Social Informatics, 8, 2016a. Anais... Bellevue: Springer, Cham, 2016a, 14-21.

CHAVOSHI, N.; HAMOONI, H.; MUEEN, A. DeBot: Twitter Bot Detection via Warped Correlation. In: International Conference on Data Mining, 16, 2016b. Anais... Barcelona: IEEE, 2016b, 817-822.

CHU, Z.; GIANVECCHIO, S.; WANG, H.; e JAJODIA, S. Who is tweeting on twitter: human, bot, or cyborg? In: Computer Security Applications Conference, 26, 2010. Anais... Austin: Applied Computer Security Associates, 2010, 21–30.

CHU, Z.; GIANVECCHIO, S.; WANG, H.; JAJODIA, S. Detecting automation of Twitter accounts: Are you a human, bot, or cyborg?. IEEE Transactions on Dependable and Secure Computing, v. 9, n. 6, p. 811-824, 2012.

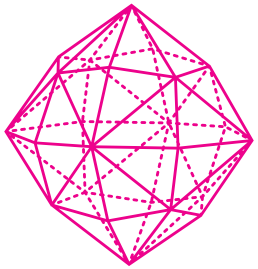
CRESCI, S.; Di PIETRO, R.; PETROCCHI, M.; SPOGNARDI, A.; TESCONI, M. The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In: International Conference on World Wide Web Companion, 26, 2017. Anais... Perth: International World Wide Web Conference Committee, 2017.

LEE, K.; EOFF, B. D; CAVERLEE, J. Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter. In: International Conference on Weblogs and Social Media, 5, 2011. Anais... Barcelona: Association for the Advancement of Artificial Intelligence, 2011, 17-21.

RUEDIGER, M.A. (coord.) Robôs, Redes Sociais e Política no Brasil: estudo sobre interferências ilegítimas no debate público na web, riscos à democracia e processo eleitoral de 2018. Rio de Janeiro: FGV DAPP, 2017a.http://dapp.fgv.br/robos-redes-sociais-e-politica-estudo-da-fgvdapp-aponta-interferencias-ilegitimas-no-debate-publico-na-web/, acesso em 13/03/2018.

RUEDIGER, M. A. (coord.) Nem tão #simples assim: o desafio de monitorar políticas públicas nas redes sociais. Caderno de Referência de Metodologia 1. 2ª edição. Rio de Janeiro: FGV DAPP, 2017b. http://dapp.fgv.br/publicacao/nem-tao-simples-assim-o-desafio-de-monitorar-politicas-publicas-nas-redes-sociais/, acesso em 20/04/2018.

VAROL, O.; FERRARA, E.; DAVIS, C.; MENCZER, F.; FLAMMINI, A. Online Human-Bot Interactions: Detection, Estimation, and Characterization. In: 11th International AAAI Conference on Web and Social Media, Montreal, Quebec, Canada, 2017.



FGV DAPP

INNOVATION FOR PUBLIC POLICIES

FGV.DAPP FGVDAPP FGVDAPP

dapp.fgv.br | dapp@fgv.br

+55 21 3799.4300