Social Network Analysis and Mining (2018) 8:5
https://doi.org/10.1007/s13278-017-0482-x

CASE REPORT

User profiling of the Twitter Social Network during the impeachment of Brazilian President

Fabrício Olivetti de França[1] · Denise Hideko Goya[1] · Claudio Luís de Camargo Penteado[1]

Received: 2 May 2017 / Revised: 22 December 2017 / Accepted: 28 December 2017 / Published online: 11 January 2018

Abstract
The impeachment process that took place in Brazil in April, 2016, has generated a large amount of posts on the Social Networks. These posts came from ordinary people, journalists, traditional and independent media, politicians and supporters. The identification of the impact of this subject on each group of users can be an important analysis to verify the real interest of common Brazilian citizens on this matter. As such, we propose a way to segment the users into popular, activists and observers in order to filter out information and help us give a more detailed analysis of the event. The proposed segmentation may also help other studies related to the usage of Twitter during important events.

Keywords Social network analysis · User profiling · Analytics

1 Introduction

In April 2016, Brazil experienced the beginning of the impeachment of its recently reelected president, Dilma Rousseff (Jacobs 2016). After winning the 2014 presidential election by a small margin of votes, Dilma Rousseff confronted a wave of protests that began in 2015 (de Souza Carvalho et al. 2016a, b). These protests mobilized thousands of Brazilians disaffected with her government and with the Workers Party (PT) management, mainly by the use of social network sites (SNS), to disseminate information and demonstrations asking for her impeachment.

Allegations of corruption and problems in the national economy have undermined political capital of Dilma Rousseff. Recurrent news reports in the media revealed corruption scandals involving mainly politicians from her party and administration, and economic problems, like the increase in inflation and unemployment.

All these cases led to economic and political crises. These crises culminated in the opening of the impeachment process by the National Congress of Brazil (lower house) on April 17, 2016. With 376 (out of 511) votes of congressmen in favor of the Dilma Rousseff's impeachment, the Brazilian Congress started her impeachment. In the Brazilian Senate, Rousseff experienced another defeat that temporarily removed her of the presidency.

In social network sites, like Facebook, Twitter and WhatsApp, there had been intense use of these channels by groups pro and anti-impeachment trying to advocate for their political position. This virtual activism mobilized a million of users, transforming the SNSs into a space of political dispute in which each group aimed to construct a narrative in favor of their causes, through the dissemination, sharing and production of information about politics.

In order to understand the political use of the cyberspace, in this paper we aim to answer questions like: How to identify influential users? How they can be characterized? What is the importance of these users? For this purpose, we have analyzed the Twitter users behavior during the first voting of the impeachment of the Brazillian ex-president, Dilma Rousseff, in the Congress. This voting received great visibility with online transmission by TV, radio and discussions on the internet.

The paper is organized as follows: in Sect. 2 we give a brief description of the Twitter Social Network and we describe our methodology of user segmentation, proposed on this paper, in Sect. 3 we analyze the collected data obtained with the Twitter API showing some interesting

FAPESP-OSU-2015 MOBILITY GRANT 2015/50250-9.

✉ Fabrício Olivetti de França
folivetti@ufabc.edu.br

[1] Nuvem Research Strategic Unit, Universidade Federal do ABC, Santo André, SP, Brazil

facts regarding the event, finally in Sect. 4 we conclude this paper giving some highlights of the performed analysis.

2 Methodology

In this section, we will describe the data collection steps as well as the methodology used for the user profiling analysis.

2.1 The Twitter Social Network

The Twitter Social Network (Kwak et al. 2010) is a directed network with each node representing one user and the relationships modeled after a directional interaction between two users. Given two users, A and B, when we have a directed edge from A to B we can say that A follows B and that B is followed by A. In this Social Network there is no need for a follower to be followed back.

Each user can post a short message regarding any subject to be broadcast to his followers. The main interaction is the *retweet* in which a given user broadcast a message originally posted (or retweeted) by other users.

With this *retweeting* action, a message can be spread throughout the network even to those users that do not follow the original poster.

2.2 Data collection

We have used the Twitter Streaming API (dev.twitter.com) to collect the tweeting data from April 15 to April 19, 2016, corresponding to 2 days prior the voting and 2 days after. This API provides a continuous flow of tweets limited to 1% of the total amount of tweets being published globally at every time step (Makice 2009).

In order to narrow the results, the API allows us to set a filter regarding the contents and properties of the tweets, limiting the total number to be retrieved. Notice that, depending on the filter, this enable us to retrieve the entirety of the tweets pertaining to a given topic.

For example, if 1% of the tweets account for 1 million tweets and the applied filter retrieves 900 thousands tweets, then the API will provide the full set of requested tweets.

For this report, we have applied the content filter that retrieves only tweets containing one of the requested keywords. The retrieval process is case-insensitive and replaces most accented character to its non-accented equivalents.

The terms used for the filter in this work are described in Table 1, and they are grouped as terms commonly used by the group favorable to the impeachment (PRO), the group against the impeachment (CON) and the terms used by both groups. Notice that these terms were manually selected by observing the main active groups of each side.

Table 1 Terms used for the data collection on Twitter

PRO	CON	Both
ImpeachmentJá, RenunciaDilma, ForaDilma, ForaLula, ForaPT, DilmaSai, SaiDilma, PTNuncaMais, LulaNuncaMais, jesuijararaca, somostodosmoro, ACasaTaCaindo, DiaHistorico, OpLavaJato, DeixaAPFTrabalhar, FimdaEraLula, Aletheia, OpAletheia, Lulanapapuda, LulaPreso, EleNãoSabeDeNada, CalaBocaVcVotouNoPT	Golpe, NãoVaiTerGolpe, FicaDilma, DilmaFica, MidiaGolpista, OcupaRedeEsgoto, Aeciomaiscitadoque, DilmaMudaMais, ParaTiGlobo, IstoÉGolpe, LulaEstamosComVoce, LulaEuConfio, SomosTodosLula, PovoComLula, Lula2018, LulaPresoPolítico, FechadoComOLula, MoroExonerado, VemPraRua13Mar, VemPraRua, 13MarEuVou, 13MarVemPraRua, 13Mar, VemPraDemocracia	Impeachment, OBrasilNãoÉParaAmadores, Polícia Federal, Delcídio, Congonhas, Condução coercitiva, LulaMinistro, QuedaDoPlanalto, Catta Preta, MortadelaDay

Between these dates, we have collected 2, 372, 914 tweets from 503, 181 different users containing one of these keywords.

The collected data were then processed into a tabular format containing the user id, the tweet text, the hashtags, and the id of the original tweet, if this is a retweet. The retweet information was extracted from the corresponding fields of the data collected through the API.

2.3 User profiling

The popularity of a given discussion may be assessed by inspecting the number of users talking about the topic and the amount of content published by most users. If the data set contains just a few unique users, it means that the topic at hand is not of the overall public interest.

On the other hand, if the data set contains many unique users, it should be checked whether most users are engaged on the discussion or merely observing. This situation depicts a widespread interest on the topic but, also, a lack of opinion. The majority of the users are just observing and absorbing information.

Finally, when we have many active users publishing a considerable amount of information, we can say that the topic is of interest to the general public and that they feel certain of their own opinion.

In order to profile the users involved in the discussion of the impeachment, and verify the popularity and general engagement on this subject, we have segmented them into three major classes:

– *Popular* those users with a large quantity of retweets, whose opinions are shared among many users.
– *Activists* those users with a higher than normal quantity of tweets, whom actively emits opinions.
– *Observers* most users, with low rate of tweets and retweets, that have brief participation on the discussion.

In order to classify the users into these categories, we will exploit the fact that we can often observe a heavy-tailed distribution when measuring different quantities on Social Networks (Barabási and Albert 1999). The heavy-tailed distribution depicts the situation where a few individuals account for most quantities of the measured item. For example, few people account for most of the world wealth, few words of a given language are used much more often than most words, a minor percentage of a Social Network user is followed by many more users than most people.

Given a measurable random variable X, the heavy-tailed distribution is characterized as:

$$\lim_{x\to\infty} e^{\lambda x} Pr[X > x] = \infty \text{ for all } \lambda > 0. \tag{1}$$

One famous heavy-tailed is the power-law distribution, with a Pr characterized by:

$$Pr(x) = C \cdot x^{-\alpha}, \tag{2}$$

where x is the quantity being measured, α is the exponent and C is the amplitude of the distribution, $Pr(x)$ gives the expected number of users with x quantity of the measured content.

The different heavy-tailed distributions differ chiefly on the tail part, as such, the power-law distribution can be satisfactorily fitted on the head of the distribution.

As such, we define that a user belongs to a given category if its behavior reside in the head of the power-law distribution. The boundary between the head and the tail of the curve is given by (Celma 2010):

$$X_{\text{bound}} = N_{50}^{4/3}, \tag{3}$$

where N_{50} is the median of the distribution given by (Newman 2005):

$$N_{50} = x_{\text{m}} \cdot 2^{1/(\alpha-1)}, \tag{4}$$

where x_{m} is the minimum value that the power-law holds true and α is the exponent of the distribution.

Given these definitions, the popular users will be those in the head of the distributions of received retweets; the activists will be those in the head of the distribution of created tweets; and the observers will be those users that do not belong to the neither the popular nor the activists categories.

In order to verify whether the measured distributions follow a heavy-tailed distribution and in order to extract the head part of the distributions, we have used the *Powerlaw* library written in Python (Clauset et al. 2009; Klaus et al. 2011). This library uses a combination of maximum-likelihood fitting methods with goodness-of-fit tests based on the Kolmogorov–Smirnov statistic and likelihood ratios, as described in Clauset et al. (2009).

All distributions were fitted as a power-law and compared against the truncated power-law, lognormal, positive lognormal, exponential and stretched exponential distributions, in order to verify the most likely distribution.

In this work, we will focus on the analysis of the popular and activists, since the observers are too numerous to render a detailed report.

2.4 Sentiment analysis

An interesting analysis is to estimate the sentiment of the tweets as in favor or against the impeachment. In order to automatically classify the tweets, we have tagged two sets of hashtags that clearly represented a tweet in favor (pro) or against (con) the impeachment. We initially labeled as neutral some hashtags from news media, but we have decided to disregard this category since most news reported during these days could be classified in one of the other two categories.

The labeled samples were divided into training and validation data sets (80–20%). The training data set was used to train the classification model and the validation was used to evaluate the model parameters. The classification model used in this work was the Extreme Gradient Boosting Tree (xgboost) (Chen and He 2015). The best performance was obtained with 100 estimators, a max depth of 3 and learning rate set as 1. With these parameters, we have obtained 99.83% of accuracy on the validation data. Comparing the classification model against the pre-labeled data we found that 0.09% of tweets in favor of the impeachment was classified as contrary and 0.5% of tweets against the impeachment was classified as in favor.

The labeled data are publicly available at the project site [1].

3 Data analysis

In this section, we will first analyze the main users of each category following our methodology and their role during the impeachment process. Following we will discuss the results altogether in order to make a relation between the different groups.

3.1 Popular users

The popular users were measured by means of the power-law distribution of the number of retweets each user received during the data collection.

Even though any user can be actively engaged into the discussion, it is also important that their message is absorbed and shared by other users. The message posted by any given user can be seen by those users that follow him. Notice that, determined by the Twitter internal ranking algorithm, a tweet from user A may or may not be shown in the timeline of his follower, user B, and, also, it may or may not be highlighted.

Additionally, any user who sees this tweet may share it to his followers by the retweet action (RT). The more a tweet is retweeted, more visibility it has.

So, the popularity or influence a given user has can be measured by how many users follow him and how many retweets his tweets have.

As we can see in Fig. 1, the head of the distribution of retweets can be fitted by power-law distribution. The

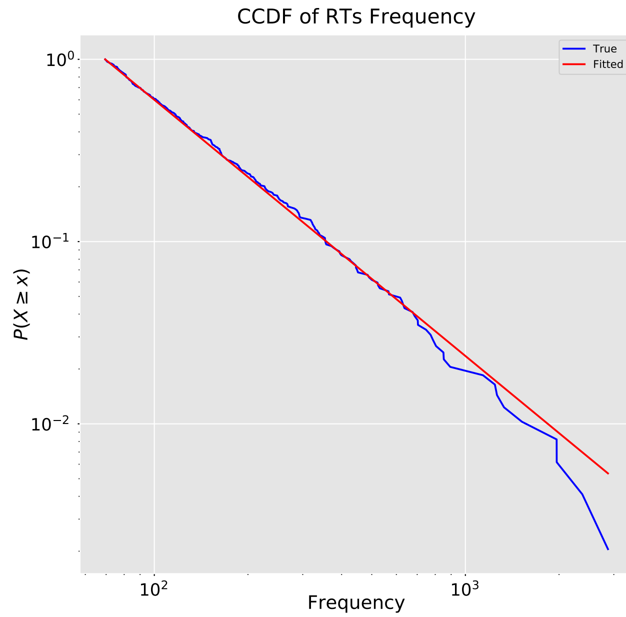

Fig. 1 The observed and fitted complementary cumulative distribution functions of retweets

best distribution found was the truncated power-law with $\alpha = -1.77$ but with a statistic significance only against the exponential distributions. Notice that the data set contains a total of 70,272 unique users with at least one retweet.

By following Eq. 2, we have found that the 36 most retweeted users can be classified as belonging to the popular users. These users are highlighted in Fig. 2.

Regarding the top 36 retweeted users, we can see that the most retweeted user was Estadao, that is the official Twitter account for the newspaper O Estado de S. Paulo. This is not surprising since the news media kept a live reporting of recent news regarding the impeachment. Most interested people on the subject, in favor or against of the deposition of the president, retweeted that content to inform their contacts and emit opinion about the news.

Other companies related to the press or online media appeared on the top list. There are two newspaper: JornalOGlobo, from Globo Group (the most influent mass media group in Brazil), and Folha, another well-known newspaper media. G1 (from Globo Group), UOLNoticias (from Folha Group) and BBCBrasil are web portals, with strong audience in news pages. RevistaEpoca and Veja are two Brazilian weekly magazines with national broad circulation. MidiaNinja, j_livres (free journalists), and LupaNews are independent news source.

BlogdoPim is an account belonging to a columnist of Veja magazine, a weekly magazine of general circulation in Brazil; he talks about politics, biased in favor of impeachment. BlogdoNoblat belongs to a columnist of O Globo; the journalist disseminates links and news from the journal

[1] http://professor.ufabc.edu.br/~folivetti/eleicoes/impeachmentSent.csv.zip

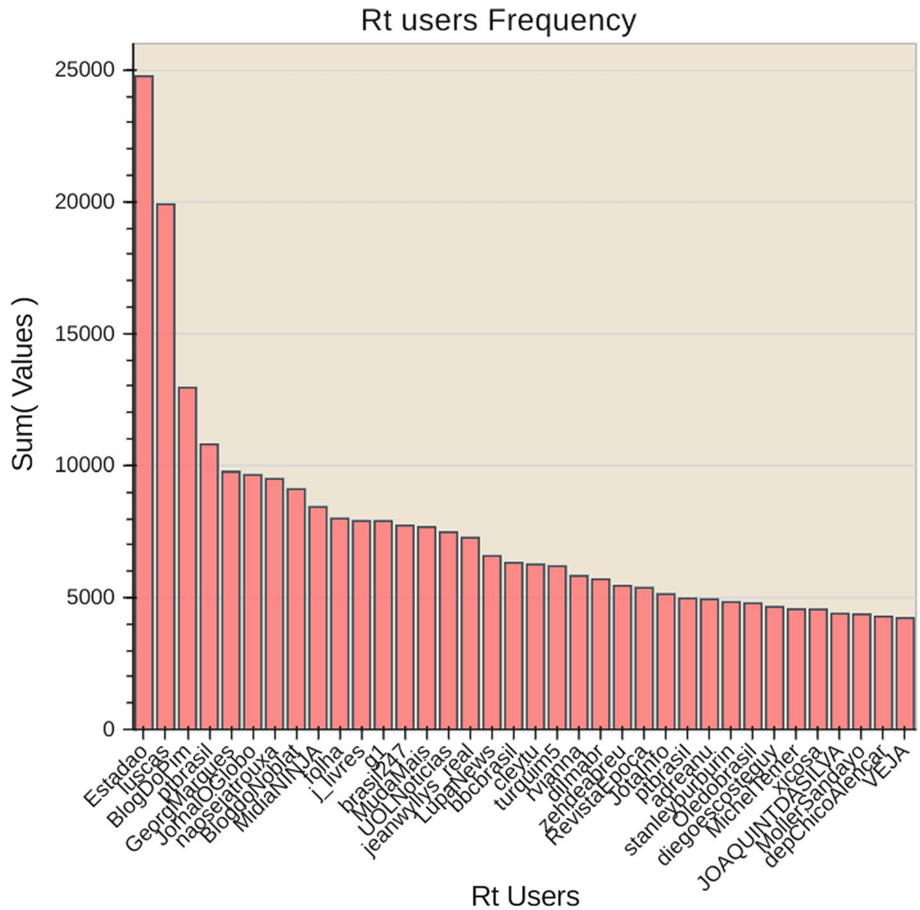

Fig. 2 Distribution of retweets for the 36 popular users observed on the data set

or other news about politics. Diegoescosteguy is journalist from Época Magazine

GeorgMarques is a journalist who also works in public relations, who covers news about the Brazilian Congress and politics. Brasil247 is a news portal, specialized in short news to be readed in small screens. JotaInfo delivers legal and judicial news.

PTbrasil is the Worker Party's account, the Dilma Rousseff party, there is also a PTbrasiI account in later positions that supports the original account of the party. Dilmabr is the official account of Dilma Rousseff. Jeanwyllys_real and depChicoAlencar are deputies, both against the impeachment. MudaMais, Turquim5, rvianna, zehdeabreu, Stanleyburburin, and xicosa are supporters or contrary to the impeachment.

MichelTemer is the official account of the Brazillian vice-president Michel Temer.

Apparently, Luscas is a teenager, very popular on social media, who tells jokes and talks about different topics. Similarly, naosejatrouxa is another user account that does not deal with politics; addresses issues from the adolescent universe and has more than 2 million followers on Twitter. Cleytu, Adreanu are other similar profiles. JoaquinTdaSilva is a retired person who also does not talk seriously about politics. OledoBrasil is a humor profile that creates fake news about sports. MollerSandayo is a deactivated account and cannot be currently analyzed.

In Table 2, we can see a summary of the popular profiles categorized by their type and opinion. Notice though that this categorization is based only on the tweets made by those accounts during this period. These opinions might have changed by now or have not been sincere at that moment.

3.2 Active users

A heavy-tailed distribution for the tweets generated by each user means that a small percentage of the users are more engaged in spreading the information than most users.

The distribution of the number of tweets created by each of the 503,181 unique users follows a truncated power-law (Fig. 3) with an exponent $\alpha = -2.28$ and, again, with a significant difference only when compared against the exponential distributions.

We can see from these figures that the impeachment voting engaged many users into discussing the political event, even among the top 1000 users we can observe that they posted more than 500 messages during the collection period.

Table 2 Categorization of the popular profiles

Profile	Type	Opinion
Estadao	News media	Neutral
luscas	Popular	Neutral
BlogDoPim	Journalist	Pro
ptbrasil	Political party	Con
GeorgMarques	Journalist	Con
JornalOGlobo	News media	Neutral
naosejatrouxa	Entertainment	Neutral
BlogDoNoblat	Journalist	Pro
MidiaNINJA	News media	Neutral
folha	News media	Neutral
j_livres	News media	Neutral
g1	News media	Neutral
brasil247	News media	Neutral
MudaMais	Political movement	Con
UOLNoticias	News media	Neutral
jearwyllys_real	Politician	Con
LupaNews	News media	Neutral
bbcbrasil	News media	Neutral
cleytu	Popular	Neutral
turquim5	supporter	Con
rvianna	Journalist	Con
dilmabr	Politician	Con
zehdeabreu	Actor	Con
RevistaEpoca	News media	Neutral
JotaInfo	News media	Neutral
andreanu	Unknown	Undetermined
stanleyburburin	supporter	Con
Oledobrasil	Humor	Neutral
diegoescosteguy	Journalist	Neutral
MichelTemer	Politician	Con
xicosa	Journalist	Con
JOAQUINTDASILVA	Deactivated	Undetermined
MollerSandayo	Deactivated	Undetermined
depChicoAlencar	Politician	Con
VEJA	News media	Neutral

The *type* is related to whether it is a person, celebrity or news media, the *opinion* relates to whether they are in favor of the impeachment (PRO), against it (CON), or neutral. Notice that it is classified as Pro or Con only the users who explicitly positioned themselves

Through our methodology, we have found, though, that the first 28 users were much more engaged than the remainders.

In Fig. 4, we can see these 28 users. Briefly looking closely, we can see that the top user, gumartinslive, tweeted about 30% more than the remaining top 10. By further inspecting this account, we can see that he is mainly a news propagator, regardless of the subject. He is not a political activist user and shows an unusual behavior, with an average of over 1500 posts per day; sometimes he replicates the same messages many times. Another news propagator are Laineandreier and CleuzaBapti; they copy or retweet messages about many subjects.

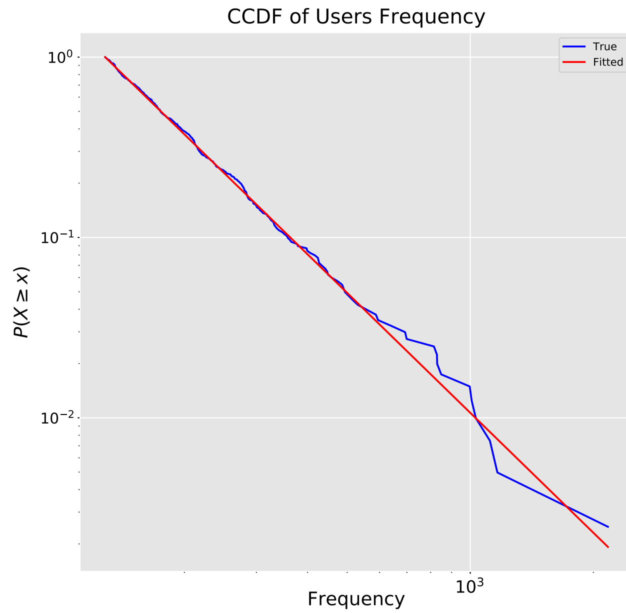

Fig. 3 The observed and fitted complementary cumulative distribution functions of tweets

Dionianjos, Leucio02, Leleabreuv, RonaldoRuffo, oConsciente, woodstock_59, and midiacrucis are supporters of the PT; they intensively tweet news and comments against impeachment or in favor of Dilma. Other users usually retweet many posts anti-impeachment; they don't create many new posts, but replicate the PT supporters opinion: moemasbc57, lacerdagalo, tadeu_alves, araujosergio, CartasMarottas, LisandraBarros1, mariaap94213193, claudiocalente, and AurelioPerna.

LuxAjuda, beijopai, Helenaguimara, benevides_cris, and SakaSakamori are impeachment supporters; they tweet news and comments against the PT or in favor of Dilmas impeachment. Similarly, br45ilnocorrupts posts talk about corruption in Brazil, but they only cite news involving PT members, so they are pro impeachment. Beamaral84 retweets news and opinions anti-Dilma and frequently invites people to participate in demonstrations pro impeachment.

_carolina__m and Troca_de_canal were two very active users that seem to have their accounts deactivated.

3.3 Activity versus popularity

In the previous sections, we have verified that there are many active users and many popular users. One open question, from these analysis, is whether the most active users intersect with the most popular. In other words, *does a high activity implies a high popularity on this data set?*

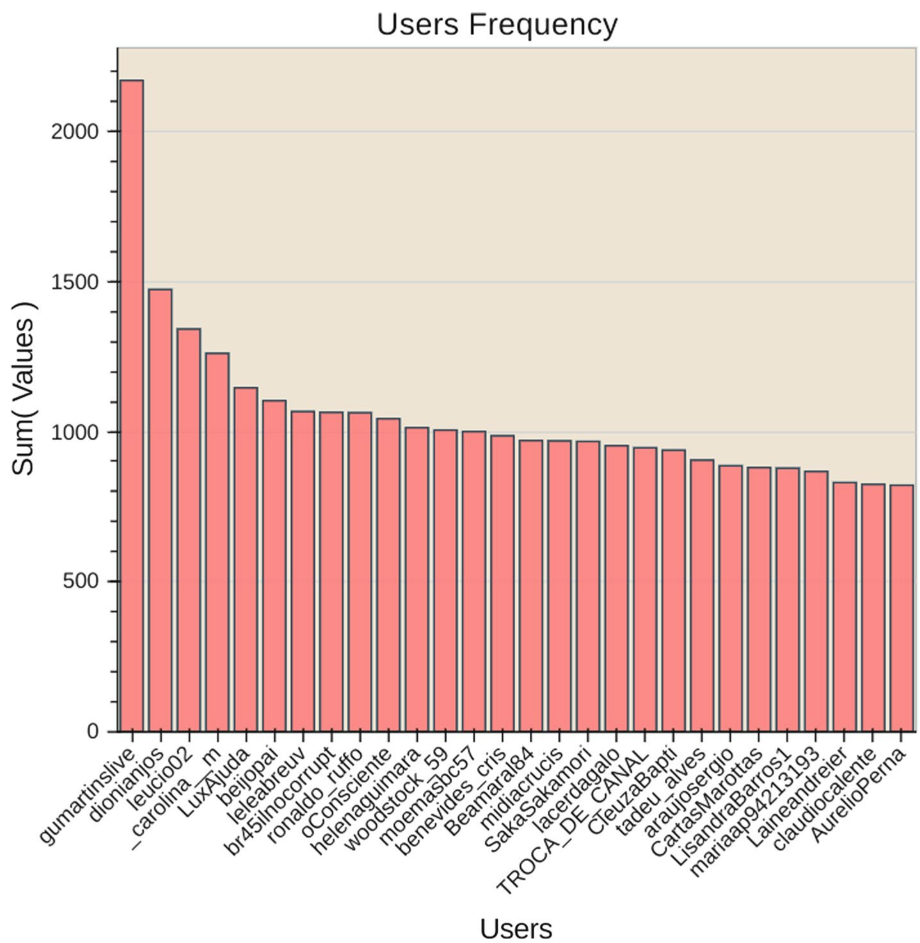

Fig. 4 Distribution of tweets by the top 28 users observed on the data set

To answer this question, in Fig. 5 we plotted the number of tweets and number of retweets for every user in the data set. As we can see, there is no measurable correlation between these two variables, in fact the users that were more retweeted are located in the area with less tweets and just three of the top 100 users with more tweets appear in the top 100 RT users (Fig. 6). The users were: GeorgMarques, a journalist against the impeachment, br45ilnocorrupt, a supporter group in favor of the impeachment, oConsciente, a supporter group against the impeachment.

3.4 Influence

In order to identify the influence of popular and activists over the observers, we have observed the role of the three categories of users in messages flow.

Over the total of collected posts during the event, 60.1% are retweets, while 39.9% are original posts. Near 2/3 of users (67.9%) retweeted at least once and 14.0% of users were retweeted. These numbers indicate that the retweet relationship is relevant to the influence analysis.

In Sect. 3.3, we have shown evidence that activists are essentially unpopular, comparing to the most retweeted in this event. The top 28 activists together received less than 1% of the total of retweets. In addition, while activists individually produce significant amounts of messages, they are not responsible for the large mass of posts. Considering the total of posts (new tweets + RT), all the top 28 activists produced only 1.2% of the messages. The top 100 actives were responsible for 3.2% of posts. The most retweeted users posted very few tweets when compared to the activists (Fig. 5). Therefore, we can see that the large mass of posts was generated by the observers.

An interesting observation is that activists not only tweeted but also retweeted a lot. The most frequent retweeters coincide, in part, with the frequent tweeters: 64.3% of the top 28 activists are among the top 1000 retweeters.

The activists played a fundamental role for ideas dissemination of popular users. About 91.3% of all retweets by activists were from the top 36 popular (Fig. 7). Thus, it can be concluded that the activists, although they were retweeted less, shared and spread messages from the most popular, helping in the dissemination of their opinion or their narratives of the facts.

Most of the retweets that circulated in the period were sent by observers, totaling 99.7%. A portion of 18.9% of

Table 3 Categorization of the active profiles

Profile	Type	Opinion
gumartinslive	Popular	Neutral
dionianjos	Popular	Con
leucio02	Popular	Con
_carolina__m	Popular	Con
LuxAjuda	Popular	Pro
beijopai	Suspended	Pro
leleabreuv	Popular	Con
br45silnocorrupt	supporters group	Pro
ronaldo_ruffo	Popular	Con
oConsciente	supporters group	Con
helenaguimara	Popular	Pro
woodstock_59	Popular	Con
moemasbc57	supporter	Con
benevides_cris	supporter	Pro
Beamaral84	Popular	Pro
midiacrucis	Suspended	Con
SakaSakamori	Journalist	Con
lacerdagalo	Blocked	Con
TROCA_DE_CANAL	Deactivated	Neutral
CleuzaBapti	Suspended	Pro
tadeu_alves	Popular	Con
araujosergio	Blocked	Con
CartasMarottas	Journalist	Con
LisandraBarros1	Popular	Con
mariaap94213193	Popular	Con
Laineandreier	Popular	Con
claudiocalente	Popular	Con
AurelioPerna	Popular	Con

The *type* is related to whether it is a popular person, supporter or journalist, also the current situation of the account (suspended, blocked, etc.) is reported. The user is classified as supporter if they exclusively tweet about politics regarding his party of choice. The *opinion* relates to whether they are in favor of the impeachment (PRO), against it (CON), or neutral. Notice that it is classified as Pro or Con only the users who explicitly positioned themselves

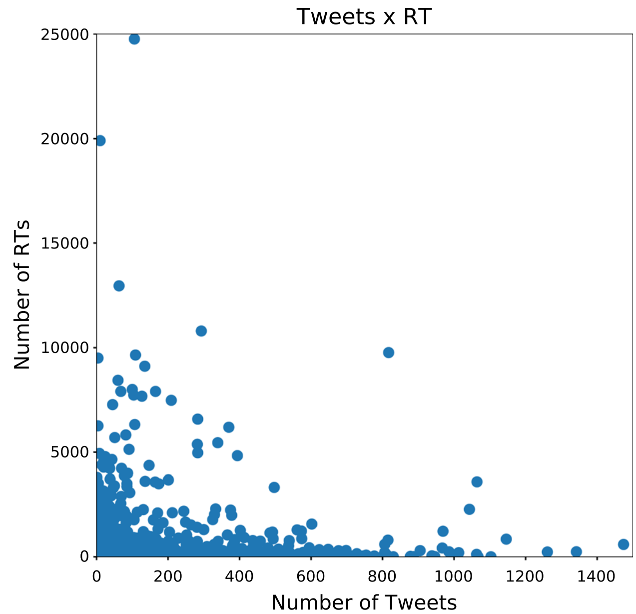

Fig. 5 Relationship between number of retweets and activity

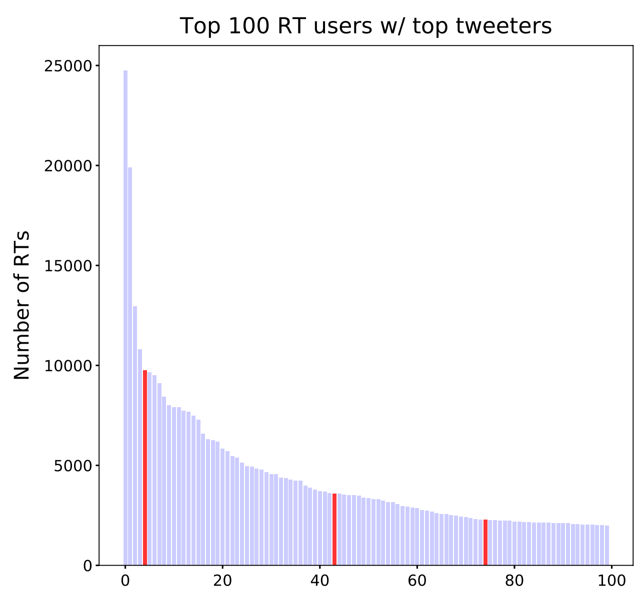

Fig. 6 The top 100 popular users with active users highlighted in red

retweets by observers was from the 36 popular users (Fig. 7). This proportion suggests that popular users produced relevant content for observers and activists.

Another significant finding is that 11.5% of the collected messages are due to the 36 most popular (directly or indirectly, by RT), indicating that these users played an important role in the discussion.

To complement the influence study, we analyzed the sentiment of each tweet aiming to identify any trend or whether the general sentiment of popular or activists was propagated to observers (Table 3).

From the total of collected tweets, 75.3% were classified with negative sentiment (pro impeachment), while 24.7% were classified with positive sentiment (con impeachment). Within the set of retweets, 73.85% were negative and 26.15% were positive.

The proportion of retweets classified as pro or against the impeachment is shown in Table 4. Users who received at least one retweet were ranked based on their popularity (total of retweets received), and each row in Table 4 shows an interval of that ranking. It is possible to observe that the proportion of RT pro and against impeachment is stable in all the ranges of popularity.

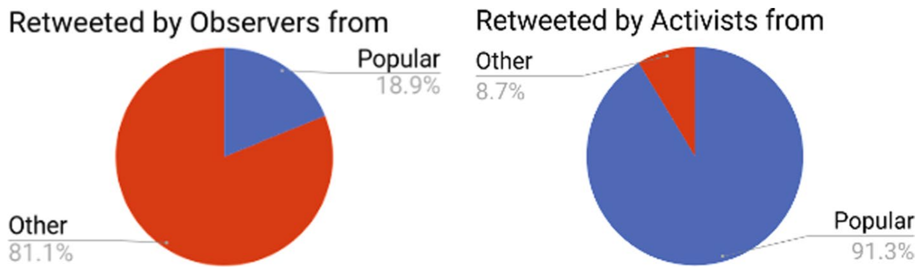

Fig. 7 Observers and activists retweet from the top 36 popular users

Table 4 Retweets from users ranked by popularity, classified as pro and con impeachment

Popularity range (retweeted from)	Total of RT (%)	Sentiment (%)	
		Pro	Con
Top 36	19.4	73.3	26.7
37–100	12.0	72.0	28.0
101–1000	35.3	73.7	26.3
1001–10,000	24.3	74.2	25.8
10,001–72,271	9.0	77.4	22.6

For Table 5, users were ranked according to their activity (total of tweets and retweets). In Table 5, it can be seen that retweets pro impeachment increase in proportion as they were retweeted by less active users. On the other hand, retweets against impeachment decrease as they were done by less active users. For original posts (non-retweets), the pro and con ratio is more similar to that of Table 4. These results point to the fact that anti-impeachment activists retweeted in large numbers (almost balancing the quantities of retweets pro and con made by activists). But they didn't create sufficient original posts to keep the same result of general sentiment in non-RT posts. It also suggests that observers in general may have absorbed fewer tweets contrary to the impeachment. Along with the results from Table 4, we can conclude that posts from observers and popular were predominantly pro impeachment.

In order to detect possible change of sentiment and if this may be related to the action of popular or activists, we have calculated the general sentiment for each user, which is the sum of sentiment of all his or her tweets. The majority of the users, or 80.90%, was pro impeachment, while 13.75% was against. The low percentage of users who obtained general sentiment equal to zero, 5.35%, points to a scenario in which the majority of participants in Twitter already had a determined position on the days of collection.

To verify if there was a change of position, we analyzed in more detail the users with general sentiment between − 1 and 1, for which it was possible to identify a positive or negative sentiment in two different days. We call these users as undefined and only 3.9% of users were classified as such. 55.0% of these undefined users presented a peak of positive sentiment in the first few days and subsequently evolved to a peak of negative postings on another day; 39.4% of these evolved into positive posts; 5.6% remained indeterminate: they oscillated between positive and negative throughout the days.

Undefined users retweeted 1.6% of total collected RT. They retweeted popular and activists in a proportion analogous to what observers in general did (19.5% were RT from the 36 most popular users and 0.17% from the top 28 activists). The sentiment of these retweets was greater for pro impeachment relative to both popular and activists (55.9 and 62.5%, respectively). Most of RT from popular, or 59.9%, were retweeted by undefined users that evolved to negative (pro), while 32.3% were retweeted by those that evolved to positive (con).

These sentiment change results do not necessarily indicate shifts in opinion, since the analysis is performed in data collected during a small time window. It is also not possible to say that the change of sentiment was caused by the action of popular, activists or other users, since the general behavior of undefined users was similar to that of users in general.

Table 5 Retweets and original posts made by users ranked by activity, classified as pro and con impeachment

Activity range (retweeted by)	Retweets (%)			Original posts (%)		
	Total	Pro	Con	Total	Pro	Con
Top 28 active	1.2	54.9	45.1	1.3	72.6	27.4
29–100	2.6	54.0	46.0	1.0	62.9	37.1
101–1000	13.0	61.0	39.0	7.2	73.8	26.2
1001–10,000	24.6	68.4	31.6	24.1	77.3	22.7
10,001–100,000	33.0	78.2	21.8	37.7	76.7	23.3
100,001–503,181	25.5	83.0	17.0	28.7	80.7	19.3

4 Final remarks

This work made a brief study of the Twitter users during the impeachment voting event that took place in Brazil in April, 2016. For this purpose, we have proposed a way to classify the users as populars, activists and observers, and then we have analyzed the first two sets in more detail. We have identified who retweeted messages from these groups and which sentiment was associated to these retweets.

From the analysis, it is interesting to notice that regarding the popular users, the majority belonged to the media. Even thought this shows that most people still listen to the traditional media, they are also starting to follow independent media that shares different point of views. Also it was identified that popular users were more influential than the active users, but the latter helped the former to disseminate their ideas.

Another interesting fact is that the political parties, from both sides, have a significant amount of *supporters* that help to spread their own ideology online. This communication *ecosystem* made possible a democratic political discussion with equal opportunity for both sides. Observers retweeted messages from popular from both sides in significant amount; thus, it is possible to conclude that their messages matters.

Finally, we have observed that the popular and activists set are almost disjoint, meaning that an overload of information does not make a given user popular, and that people still look for authorities, in general, in order to become informed.

The analysis was carried out on political content, in a scenario of strong polarization of users (de Souza Carvalho et al. (2016a)). The proposed segmentation led to useful results within this context, but in other scenarios it is recommended that further testing be performed.

References

Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512

Celma Ò (2010) The long tail in recommender systems. In: Music recommendation and discovery, Springer, Berlin, pp 87–107. http://www.springer.com/br/book/9783642132865?wt_mc=ThirdParty.SpringerLink.3.EPR653.About_eBook#otherversion=9783642132872

Chen T, He T (2015) Xgboost: extreme gradient boosting. R package version 0.4-2

Clauset A, Shalizi CR, Newman ME (2009) Power-law distributions in empirical data. SIAM Rev 51(4):661–703

de Souza Carvalho C, de França FO, Goya DH, de Camargo Penteado CL (2016a) Brazilians divided: political protests as told by twitter. In: Hutchison D, Kanade T, Kittler J et al (eds) Transactions on large-scale data-and knowledge-centered systems, vol 27. Springer, Berlin, pp 1–18

de Souza Carvalho C, de França, FO, Goya DH, de Camargo Penteado CL (2016b) The people have spoken: conflicting Brazilian protests on twitter. In: 2016 49th Hawaii international conference on IEEE system sciences (HICSS), pp 1986–1995

Jacobs A (2016) Brazil's lower house of congress votes for impeachment of Dilma Rousseff. The New York Times, p 17

Klaus A, Yu S, Plenz D (2011) Statistical analyses support power law distributions found in neuronal avalanches. PloS one 6(5):e19779

Kwak H, Lee C, Park H, Moon S (2010) What is twitter, a social network or a news media? In: Proceedings of the 19th ACM international conference on World Wide Web, pp 591–600

Makice K (2009) Twitter API: up and running: learn how to build applications with the Twitter API. O'Reilly Media Inc., Sebastopol

Newman ME (2005) Power laws, pareto distributions and zipf's law. Contemp Phys 46(5):323–351