

# Large-Scale And Long-Term Characterization Of Political Communications On Social Media

Lucas Santos de Oliveira  
lsoliveira@uesb.edu.br  
State University of Southwestern of Bahia  
Jequié, Bahia, Brazil

Pedro Olmo Stancioli Vaz de Melo  
olmo@dcc.ufmg.br  
Federal University of Minas Gerais  
Belo Horizonte, Minas Gerais, Brazil

## Abstract

Social media play an important role in shaping political discourse, creating a public sphere that enables discussions, debates, and deliberations. Aware of this importance, politicians use social media for self-promotion and as a means of influencing people and votes. As an example of this assertion, in 2018, Brazilians democratically elected for president the far-right candidate Jair Bolsonaro. One of the most surprising feats of this outcome is that his party, PSL, had almost no television time. His victory was only possible because of his supporters' engagement and activism on social media platforms, such as Twitter, Facebook, and WhatsApp.

In this context, politicians need to decide how to communicate with their voters to build their reputations. While some politicians only share professional communications about their political agenda and activities, others prefer a more non-political and informal approach, sharing communications about the most varied subjects, such as religion, sports, and their families. Others, however, misuse platforms by spreading political messages that violate policies and circumvent electoral laws.

Aware of these problems, I propose the LOCPOC a methodology to characterize the communication of Brazilian politicians over years in terms of the amount of *political* and *non-political* messages they post. The methodology is robust to concept drifts over time, requiring few new labeled messages each year. From the classified messages, I was able to characterize the communication of politicians over time and identified new findings: (i) Brazilian congresspeople changed their communication behavior over time; (ii) concept drifts occurred during important events in Brazilian politics; (iii) the explosive rise of the right seen just before the 2018 elections; (iv) a broader and more evenly distributed right-wing participation than the left-wing, and, finally, (v) the increase of public engagement over time.

**Keywords:** political, social media, communication, characterization

---

In: IV Concurso de Teses e Dissertações (CTD 2022), Curitiba, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2022.

© 2022 SBC – Sociedade Brasileira de Computação.  
ISSN 2596-1683

## 1 Introduction

In *Candidate-centric Electoral Systems*, social media play an important role in shaping political scenario [2]. This is the case in Brazil, where Brazilians are considered some of the most enthusiastic users of social networks. Online platforms remain the main source of news within urban Brazil with massive content consumption and share [12]. In this scenario, the elections have been marked by the heavy usage of social media during the campaign [18], electoral laws violation [19] and the attempt to influence voters to change the outcome of the elections [12].

The peak of the social media influence during elections in Brazil occurred in 2018 when Brazilians democratically elected a candidate for president: Jair Bolsonaro [18]. His party, PSL, which in the 2014 elections obtained only 808,404 votes, was the party that received the highest number of votes in the 2018 House election, 11,640,033, electing 52 federal deputies. One of the most surprising and fascinating feats of this outcome is that all candidates affiliated with PSL, including Jair Bolsonaro, had almost no television time [4]. Before 2018, television time was considered a fundamental requirement for those aspiring to be the presidential chair [4]. Many analysts claim that his victory was only possible because of his supporters' engagement and activism on social media platforms, such as Twitter, Facebook, and WhatsApp [4, 18].

## 2 Problem Statement

Aware of the social media importance, politicians use social media to build their reputations and increase political engagement [7], though some are more successful in this than others [20]. Part of this success is related to the communication strategy of the politicians. While some politicians only share professional communications about their political agenda and activities, others prefer a more personal and informal approach, sharing communications about the most varied subjects, such as religion, sports, and their families [15]. Others, however, misuse social media platforms by spreading political messages that violate policies and circumvent electoral laws [19].

One way to investigate these problems and characterize the communication of these politicians is through the identification of *political* and *non-political* textual content posted on different social media [2, 15, 19]. Contents of the first

group explicitly describe political opinions and activities, helping the public to know what to expect from politicians when they occupy public offices, e.g., “Today I attended a meeting with Minister Ricardo Veléz at the Education Commission ...”. On the other hand, the contents of the second group are related to their private life or are not directly linked to their political activities and ideals, e.g., “Me and my daughter Maria Laura preparing our Christmas dinner...”

Both messages above are real and were very popular on Twitter during the 2018 elections, showing that the public might like these two types of communication. Therefore, understanding how politicians use these two types of communication leads us to the following research questions:

- (R0):** Can we say that, in general, the public appreciates and react to *political* and *non-political* content equally?
- (R1):** Are there differences in behavior depending on the electoral success?
- (R2):** Do politicians and their followers behave differently depending on the ideological spectrum?
- (R3):** Do politicians change their communication behavior on social media based on the public response?

I hypothesize that the answers to these questions can provide valuable insights for understanding the Brazilian political scenario and the communication of politicians over the years. If I am able to automatically group these posts, I could characterize all politicians that have (and use) social media by the amount of *non-political* and *political* content in their communications. With that, tools could be easily provided for the public to, for instance, identify the political agenda of politicians, identify politicians who are not transparent about their political views, and identify candidates that misuse social media platforms.

Although many studies have analyzed the content of social media messages posted by politicians [10], the problem of identifying *non-political* and *political* content, on a large scale, is not trivial. In short, the proposed solutions are very difficult to generalize or to keep accurate over a long period of time. Existing efforts ignore the content of the post [1], focus on manual inspection of small sets of messages [17], or propose aggregate functions (e.g. count, frequency, etc.) based on specific keywords or hashtags (e.g. “abortion”) to quantify how much politicians are dedicating their communications to a specific topic [8]. Moreover, even works that perform a classification task do not characterize the communication of politicians over time and barely consider the nature of these contents, i.e., whether they are communicating about *political* or *non-political* topics [6].

Furthermore, when the topics covered in these posts change significantly, due to noise disturbance or long-term evolution, for instance, a concept drift is characterized. Under this circumstance, a classifier responsible for identifying political contents loses its generalization capability, failing to capture nuances in the texts from different periods [11]. In this

case, it is necessary to adopt mechanisms for the classifier to understand this new context.

In my thesis, I proposed the LOCPOC (*L*ong-term *C*lassification of *P*olitical *C*ommunications), a computational methodology able to characterize the communication behavior of politicians by processing and classifying large-scale social media messages that span for several years. Moreover, the classifier is robust to concept drifts, where topic changes are identified over the years by an unsupervised drift detection method [3] that applies an Active Learning [22] technique in which the true labels of the messages are required only after drift detection. Finally, I used a transfer learning approach [9] to retrain the classifier after drift detection that maintains a high sensitivity to the long-term data changes. From the classified messages, I was able to characterize the communication of politicians over time. I could identify, for instance, if they change their communication behavior over time, e.g. around elections or when they take over or leave public office.

### 3 Contributions

To demonstrate the usefulness of my proposed methodology, I applied it in two case studies related to Brazilian political communications.

**(C1).** Objective: To Analyze whether the nature of politicians’ communications changes around the elections and whether this behavior is related to electoral success. This objective answers the research questions **R0** and **R1**, and partially **R3**. Publications related to this objective are [13, 15, 16].

For this task, I analyze a collection of tweets posted by Brazilian politicians from one year before the 2014 Elections to one year after. I collected all tweets posted by **all** 692 congresspeople who were active on Twitter and worked in the Brazilian parliament from October 2013 to October 2015. The congresspeople were labeled as (1) *newcomer* (*NC*) if they were elected in 2014 but were not in Congress in the previous political term; (2) *reelected* (*RE*) if they were elected in 2014 and were also in Congress in the previous political cycle; and (3) *loser* (*LS*) if they were in Congress in the previous political term, tried to re-elect and were not able to be elected in 2014. This labeling is useful to separate politicians according to their success in the 2014 Elections and according to their position as a congressperson in the year before and/or after the elections. After classifying all messages, I numerically characterize the politicians by the proportion of *political* communications they post, which I call the *Political Communication Index* (*PCI*). With that in hand, I identify if they change their communication behavior over time, e.g. around elections or when they take over or leave public office. Moreover, my methodology allows us to identify which communication behavior evokes, in general, more (and less) engagement with the public on Twitter, both

in terms of social media popularity and in terms of votes in elections.

(C2). Objective: To Analyze if there are differences in the communication strategies used by the *right*, *left*, and *center* politicians over time. Also, try to understand the growth of the right-wing in recent years. This objective answers the research questions **R0**, **R2** and **R3**. Publication related to this objective is [14].

For this task, I collected more than 3M tweets from the 914 politicians with a valid Twitter account and who occupied a public office in Congress for at least one year from October 2013 to October 2019. These congresspeople were labeled as *right* (*R*), *left* (*L*), and *center* (*C*) based on their party ideology. By separating the deputies' tweets according to these labels, I were able to analyze the communication behavior of politicians over 6 years. I reveal striking and even abrupt changes in the behavior of politicians and the public. While before 2016 most of the communications posted by politicians were *non-political* after 2016 *political* communications started to heavily dominate. Also, although total left-wing politicians post more on social media, the participation of right-wing politicians is broader, more evenly distributed, and increased significantly after the 2018 elections. The overall engagement of the public with politicians grew constantly since 2013, especially with *political* communications. Differences in engagement with respect to ideology and content increased greatly after 2015, which is possibly catalyzing the increase of polarization in Brazil [18, 21]. Politicians from the center became much less popular than those from the left and right. Also, after 2017 the public engagement with right-wing communications became much higher than with those from the left, even considering the usually less popular *non-political* messages. In summary, my results corroborate the hypothesis that social media was fundamental for the recent rise of the *right* in Brazil and its success in the 2018 elections.

My thesis extends the findings of previous research in four-fold contributions, which are described as follows:

- A computational methodology for identifying *political* and *non-political* textual content that is robust to concept drifts and that can be applied in different languages.
- Two parsimonious characterizations of Brazilian congresspeople' communications that comprise the 2014 and 2018 elections. I analyze the nature of these communications and how they may be linked to concepts that characterize Brazilian political relations. To the best of my knowledge, I am the first to analyze such a large number of messages (3, 337, 744 tweets) from a significant number of deputies (914) in such a long period (6 years span).
- To the best of my knowledge, I am the first that shows quantitatively and qualitatively at a large scale that

*non-political* discourse arises and dominates the campaigns during the 2014 elections. Surprisingly, in spite of that, I showed that *political* messages are far more popular among the public. Also, I am the first to characterize and demonstrate quantitatively the rise of the right in Brazil in this decade from social media data.

- Finally, I provide two datasets for the research community: one containing more than 3M tweets from 914 Brazilian politicians; the other containing a set of 3, 116 tweets labeled as *political* and *non-political* comprising the years from 2013 to 2019;

## 4 Discussion

The classification of short texts is not a trivial task due to the amount of noise in the communication and the limited number of characters that make it difficult to infer the context of the message. There are several attempts in the literature to classify this type of communication, however, they are not effective or cannot be applied on a large scale. Therefore, my study is an important step in social computing literature, providing a supervised machine learning classifier that labels all textual messages from different social media platforms as *political* or *non-political*. Moreover, the classifier is robust to concept drifts, where topic changes are identified over the years by an unsupervised drift detection. From that, I parsimoniously characterized the social media communications posted by Brazilian parliamentarians over time. In other words, the *political* message classifier enabled us to address two case studies with different natures and characteristics.

More specifically, it enabled us to analyze how politicians present themselves in the digital environment and how the public reacts to them. I investigated the behavior of politicians over time, i.e., whether they change their communication and if there is a typical temporal pattern that is chosen by the majority of politicians. I noticed that politicians changed their communication behavior over time, especially during important political events. Moreover, I showed that *political* and *non-political* tweets become significantly more popular over time, especially the *political* ones, which are much more preferred than the *non-political*. Finally, I show that the *right* strongly increased its participation over time, which was reflected in the votes and the number of elected politicians.

My thesis opens a wide range of possibilities for new applications and research. For example, I used my proposed CNN architecture to investigate the use of *political* Ads on Facebook during the 2018 election campaign [19]. I was able to identify many ads with *political* content that are not part of the Facebook Ad Library for *political* ads. The Brazilian election legislation stipulates that ads during the electoral period need to be labeled with the Propaganda Electoral tag to display the national identification number (CPF for individuals and CNPJ for institutions) of the advertiser. A small fraction of advertisers in their dataset have the right

disclaimer stipulated in the Brazilian election law but they did not declare their ads as *political* to Facebook, hence, they do not appear in the Ad Library.

Another possibility is the creation of new tools. Twitter recently announced that it will no longer allow *political* ads on its website [5]. I hope my findings and all the real-world experience of deploying a real system during the 2018 Brazilian elections will inform debates around public policies that regulate political advertising on the Internet. If a system like this is implemented on a widespread scale, political campaigns might adopt adversarial strategies that change their marketing strategies in order to exploit the false-negative rate. To assist in this task, I developed a tool that uses my classifier and enables the automatic labeling of *political* and *non-political* messages. The tool receives as input a message or a file with several messages and returns the label and the confidence of the classifier for each message. Although there are some efforts to make the elections in Brazil transparent, having multiple independent systems would make the monitoring of *political* messages more robust to attackers. I hope my effort will inspire other initiatives around the world. My work not only highlights the importance of independent auditing platforms for *political* ads but also provides all necessary framework to make it feasible as my code is open source<sup>1</sup>.

## References

- [1] Marcelo Santos Amaral and José Antônio Gomes de Pinho. 2016. Tuitando por Votos: Congressistas Brasileiros e o Uso do Twitter nas Eleições de 2014. In *Proceedings of the XL Encontro da Anpad*. Anpad, 1–19.
- [2] Josemar Alves Caetano, Jussara Almeida, and Humberto T. Marques-Neto. 2018. Characterizing politically engaged users' behavior during the 2016 us presidential campaign. *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018* (2018), 523–530. <https://doi.org/10.1109/ASONAM.2018.8508459>
- [3] Albert Franca Josua Costa, Regis Antonio Saraiva Albuquerque, and Eulanda Miranda Dos Santos. 2018. A Drift Detection Method Based on Active Learning. *Proceedings of the International Joint Conference on Neural Networks 2018-July* (2018). <https://doi.org/10.1109/IJCNN.2018.8489364>
- [4] Council on Foreign Relations. 2018. WhatsApp's Influence in the Brazilian Election and How It Helped Jair Bolsonaro Win. <https://www.cfr.org/blog/whatsapps-influence-brazilian-election-and-how-it-helped-jair-bolsonaro-win>
- [5] Lauren Feiner. 2019. Twitter Bans political ads – CNBC. <https://www.cnn.com/2019/10/30/twitter-bans-political-ads-after-facebook-refused-to-do-so.html>
- [6] Lei Gao, Alexis Kuppessmith, and Ruihong Huang. 2017. Recognizing Explicit and Implicit Hate Speech Using a Weakly Supervised Two-path Bootstrapping Approach. arXiv:1710.07394
- [7] Will J. Grant, Brenda Moon, and Janie Busby Grant. 2010. Digital dialogue? australian politicians' use of the social network tool twitter. *Australian Journal of Political Science* 45, 4 (2010), 579–604. <https://doi.org/10.1080/10361146.2010.517176>
- [8] Didier Grimaldi. 2019. Can we analyse political discourse using Twitter? Evidence from Spanish 2019 presidential election. *Social Network Analysis and Mining* 9, 1 (2019), 49.
- [9] Yu He, Jianxin Li, Yangqiu Song, Mutian He, and Hao Peng. 2018. Time-evolving text classification with deep neural networks. *IJ-CAI International Joint Conference on Artificial Intelligence 2018-July* (2018), 2241–2247. <http://www.cse.ust.hk/~yqsong/papers/2018-IJCAI-EvolvingTextCategorization.pdf>
- [10] Andreas Jungherr. 2016. Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics* 13, 1 (jan 2016), 72–91. <https://doi.org/10.1080/19331681.2015.1132401>
- [11] Bartosz Krawczyk, Leandro L. Minku, João Gama, Jerzy Stefanowski, and Michał Woźniak. 2017. Ensemble learning for data stream analysis: A survey. *Information Fusion* 37 (2017), 132–156. <https://doi.org/10.1016/j.inffus.2017.02.004>
- [12] Caio Machado, Beatriz Kira, Gustavo Hirsch, Nahema Marchal, Bence Kollanyi, Philip N Howard, Thomas Lederer, and Vlad Barash. 2018. News and political information consumption in Brazil: Mapping the first round of the 2018 Brazilian presidential election on Twitter. *Computational Propaganda Project* (2018).
- [13] Lucas S. Oliveira, Pedro Vaz de Melo, Marcelo Amaral, and José Antônio Pinho. 2018. When Politicians Talk About Politics: Identifying Political Tweets of Brazilian Congressmen. *International AAAI Conference on Web and Social Media* (2018).
- [14] Lucas Santos de Oliveira, Marcelo Santos Amaral, and Pedro OS Vaz-de Melo. 2021. Long-term Characterization of Political Communications on Social Media. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. 95–102.
- [15] Lucas Santos de Oliveira, Pedro O. S. Vaz-de Melo, Marcelo S. Amaral, and José Antônio G. Pinho. 2020. Do Politicians Talk about Politics? Assessing Online Communication Patterns of Brazilian Politicians. *ACM Transactions on Social Computing* 3, 4, Article 19 (Sept. 2020), 28 pages. <https://doi.org/10.1145/3412326>
- [16] Lucas Santos de Oliveira and Pedro O.S. Vaz Melo. 2017. How to Find the Relevant Words Politicians Use in Twitter?. In *Proceedings of the 23rd Brazilian Symposium on Multimedia and the Web - WebMedia '17*. ACM Press, New York, New York, USA, 465–468. <https://doi.org/10.1145/3126858.3131590>
- [17] Joyojeet Pal. 2015. Banalities turned viral: Narendra modi and the political tweet. *Television and New Media* 16, 4 (2015), 378–387. <https://doi.org/10.1177/1527476415573956>
- [18] Raquel Recuero, Felipe Bonow Soares, and Anatoliy Gruzd. 2020. Hyperpartisanship, Disinformation and Political Conversations on Twitter: The Brazilian Presidential Election of 2018. *Ted Rogers School of Management, 3 Social Media Lab* 1, 2 (2020), 569–578.
- [19] Márcio Silva, Lucas Santos de Oliveira, Athanasios Andreou, Pedro Olmo Vaz de Melo, Oana Goga, and Fabricio Benevenuto. 2020. Facebook Ads Monitor: An Independent Auditing System for Political Ads on Facebook. In *Proceedings of The Web Conference 2020*. ACM, New York, NY, USA, 224–234. <https://doi.org/10.1145/3366423.3380109>
- [20] Andranik Tumasjan, Timm Oliver Sprenger, Philipp G Sandner, and Isabell M Welpe. 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM* 10 (2010), 178–185.
- [21] Ingmar Weber, Venkata R Kiran Garimella, and Alaa Batayneh. 2013. Secular vs. islamist polarization in egypt on twitter. In *Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining*. 290–297.
- [22] Indre Zliobaite, Albert Bifet, Bernhard Pfahringer, and Geoffrey Holmes. 2014. Active learning with drifting streaming data. *IEEE Transactions on Neural Networks and Learning Systems* 25, 1 (2014), 27–39. <https://doi.org/10.1109/TNNLS.2012.2236570>

<sup>1</sup><https://github.com/lucasant10/>