

# Google it: Brazil's 2018 election

Teógenes Moura, Alexandre Gomes, Max Stabile, Alexandre Arns Gonzales,  
Alana Karoline Fontenelle Valente, Marisa von Bülow, Lorena Vilarins, Beatriz Delgado  
Val Franco<sup>1</sup>

Political Science Institute  
University of Brasília, Brazil

**Pre-print version**

**Keywords:** search engine result page; Brazilian elections; digital activism; algorithms; Google

## **Abstract**

The discussion around the effects of digital technology on democracy has gained the spotlight in recent years. In this article, we analyze an understudied aspect of these effects: the discussion around the role of search engines in electoral contexts. Our study focuses on the content of search engine results page (SERP) by Google during the Brazilian 2018 Presidential election. It is based on an innovative research strategy. We created avatar email accounts intending to represent, as accurately as possible, the digital behavior of a variety of voters. We repeatedly collected the results of these logged accounts during the election period, based on a common list of search terms. This resulted in a database of more than 235 thousand URL records and of over two million words. These results were analyzed according to the distribution of sources of the URLs shown by Google. Furthermore, we used Word2Vec to analyze the contents of the texts of each link in the results. The analysis

---

<sup>1</sup> This article is the result of an ongoing collective research project that analyzes the role played by Google in key political events. The authors thank the support of the INCT - Instituto da Democracia e da Democratização da Comunicação, funded by FAPMiG and CAPES-Brazil, as well as of the other members of the research group Resocie (Repensando as Relações Estado-Sociedade - Rethinking State-Society Relations) at the Political Science Institute, University of Brasília, Brazil. We also thank Professor Claudia Melo, from the Computer Science department, and her Software Engineering class, for her support in the realization of this research effort, including the students involved in developing the code for collecting data from Google: Lincoln Barbosa, Luís Felipe Braga, Arthur Couto, Leonardo Moraes, Guido Oliveira, Ricardo Rachaus e Gabriel Taumaturgo. Their work can be accessed in: <<https://github.com/unb-cic-esw/Observatorio-google>> , last accessed on February 14, 2020.

confirms that violent rhetoric characterized the electoral debates. It also shows that search results were concentrated on a small number of online mainstream media sources, while social media links and alternative media sources played smaller roles.

## Introduction

This article sheds light on an understudied topic in political science: the role played by internet search engines as information providers in electoral contexts. One of the critical aspects of the digitalization of political life is the change in the distribution of information. In spite of the potential impacts of such change, few scholars have strived to better understand how search engines work and their consequences in political life, especially in electoral contexts (but see Pan *et. al.*, 2007; Epstein and Robertson, 2015; Epstein, 2015; Robertson, 2016, 2018).

Our study focuses on search engine results page (SERP) by Google during a key event: the Brazilian 2018 Presidential election<sup>2</sup>. Google was the most accessed search engine during the election and is the most accessed site in Brazil overall (Alexa, 2019). According to a ranking of the 50 most accessed sites in the country, the URLs <google.com> and <google.com.br> hold first and third positions<sup>3</sup>. The second search engine cited, <yahoo.com>, appears only in the eleventh position (Idem).

The 2018 election was a disruptive process in Brazilian politics. It ushered to the Presidency Jair Bolsonaro, an extreme-right politician who was an underdog candidate until a few months before the vote. It interrupted a cycle of twenty-four years of electoral dispute between the Workers' Party (PT – Partido dos Trabalhadores) and the Brazilian Social Democracy Party (PSDB – Partido da Social Democracia Brasileira). As will be argued in this article, the use of digital technologies was particularly relevant in this context.

To better understand the role played by Google as information provider during the electoral process, we developed an innovative data gathering method. Our first goal is to present this method, its results and limitations. We also present an analysis of the data collected considering the information sources presented by Google as well as the contents of the texts within the pages of the search results. For the content analysis, we applied the

---

<sup>2</sup> The first round of the Brazilian Presidential election was held on October 7, and the second round on October 28, 2018.

<sup>3</sup> List available in <https://www.alexa.com/topsites/countries/BR>, last accessed February 1, 2020.

Word2Vec algorithm, a Natural Language Processing technique that allowed us to identify patterns of association between key topics.

Our research design was organized in three consecutive steps. We first created fifteen Google accounts, which emulated different user profiles. Six of these accounts represented left and right-wing users; other six were male and female, and the last three were accounts without a search history. We then collected the results of the searches for keywords related to the electoral process made through each logged account, between August 17 and October 28, 2018. This resulted in a database of over 235 thousand URL links. Finally, we accessed the contents of these links, which resulted in a database of approximately 2 million words.

The analysis shows that the results displayed by Google for our logged accounts were highly concentrated on a small number of online mainstream media sources. At the same time, social media links and alternative media sources were, by comparison, much less visible. The partial exception is the case of Wikipedia. The content analysis of the texts within the URLs underscores the fact that violent rhetoric dominated the electoral debates. The results of searches associated with the names of the two candidates that clashed in the second round of the Presidential election showed that words such as "fear" and "aggression" were closely associated with Fernando Haddad, the candidate of the Workers' Party, while "enemy" and "nazi" appeared close to Jair Bolsonaro, the candidate of the Social Liberal Party.

This article is organized in four sections. The first offers an introduction to the discussion around search engines and their political impacts. In the second section, we discuss the specific context in which our research took place. We also describe the attempts by electoral institutions to regulate mechanisms such as "boosting content" in digital platforms during the electoral process. The third section comprises the results and analysis from our database of search results from keywords consulted on Google. We analyze the type of media and contents delivered by Google. We conclude by describing some of the key challenges we faced during our research and identifying future possibilities related to this research agenda.

## 1. Search Engines and Politics

According to the Regional Center for Studies for the Development of the Information Society (Cetic.br) survey, in 2008 approximately 28% of Brazilian households in rural and urban areas had some type of Internet connection. By 2017, the percentage of Brazilian households with some type of connection reached the milestone of 61% of households (CETIC, 2017: 114). This number, however, may underestimate Internet use and its impacts in the country. Increasingly, people access the Internet through smartphones, without making a clear distinction whether they are on or offline.

This is especially true if we consider the use of search engine services. As several authors have argued, doing online searches has become a naturalized action in daily interactions with electronic devices and thus, contradictorily, ends up being underestimated in its potential to influence access to political information (see, for example, Pan et al., 2007; Epstein and Robertson 2015). As Hillis et al have put it: “To search has become so natural and obvious a condition of using the Web, and the Web such a natural and obvious feature of the internet, that the specific contingency of these everyday practices has become obscured” (2013: 2).

The logic by which search engines order and present results involves a complicated set of processes. As Google itself explains,<sup>4</sup> there are five aspects that are taken into account when displaying a result. First, the algorithms analyze the meaning of the query using linguistic models “to understand the type of query you’ve entered by applying some of the latest research on natural language understanding”<sup>5</sup>. Then, algorithms analyze the correspondence between keywords and websites, checking the relevance of the latter for the research. According to Google, subjective concepts are not considered in the operation:

It’s important to note that, while our systems do look for these kinds of quantifiable signals to assess relevance, they are not designed to analyze subjective concepts such as the viewpoint or political leanings of a page’s content.<sup>6</sup>

---

<sup>4</sup> See the documentation available in: <https://www.google.com/search/howsearchworks/algorithms/> , last accessed February 14, 2020. The next two paragraphs rely on this source.

<sup>5</sup> *Idem*.

<sup>6</sup> *Ibidem*.

The next steps are "Ranking useful pages" - the evaluation of a website on a given subject against other sites that provide similar content - and "Usability of the webpages", that is, analyzing whether the settings allow an adequate experience: "whether the site appears correctly in different browsers; whether it is designed for all device types and sizes, including desktops, tablets and smartphones; and whether the page loading times work well for users with slow Internet connections". Last but not least, algorithms use users' information like country, location and recent search activity that, supposedly, can be controlled and restricted by each user in privacy settings.

Also, part of the search results is guided by the logic of selling ad spaces by the company. This advertising service, according to Google, intends to improve the company's search service by coordinating - from the "sale" of keywords used in the search - two interests: the user's, who seek information, and the advertiser's, who have information of interest to the user (Vogel and McCaffrey, 2000). The coordination of these interests depends, in part, on the processing of information by Google and its view about the interests of the user. Zuboff (2019: 134-138) argues that the ad-selling mechanism may also be based on the collection of information from other Google services and products, such as the Android operating system for mobile phones. The relationship between the operating system and the search service is not well-known. In Brazil, more than 110 million people have access to the Internet through mobile devices, such as mobile phones (Belizario, 2018). Within this universe, according to a survey by data analysis company Kantar (2019), about 95.4% carry the Android operating system.

In recent years, a few scholars have strived to better understand how search engines work and their impacts in political life in general, and elections in particular. In 2015, Epstein and Robertson coined the term "search engine manipulation effect" (SEME), to call attention to the possible impacts of search rankings on elections (2015). Based on the results of five experiments held in the United States and India, the authors argue that biased search rankings can shift the preferences of undecided voters. Although this research did not use actual data from search results during elections (but relied on the manipulation of data in experiments), their findings called the attention to the dangerous potential impacts of search engine manipulation. This manipulation can be a subtle one, deriving, for instance, from a change in the order in which results appear. Because the majority of users tend to click only

on the first three URLs on the first page of the search results, such a change could have a manipulation impact that users would not be aware of (Idem; see also Granka et al. 2004).

According to Pan et. al (2007), the preference for the first-ranked results is explained by the trust that users place on the criteria used by the company to provide the best information in a quick and efficient manner. The behavior of users, who do not usually question the results obtained, and the concentration of traffic in one or two search engines, have led scholars to question the lack of transparency of companies and point to their potential misuses. These claims have provoked public reactions from representatives of Google on a few occasions. For instance, in 2015 Amitabh Kumar "Amit" Singhal, then vice president and head of the Google research group,<sup>7</sup> published an op-ed in which he strongly denied that Google would manipulate search results, calling Epstein's arguments a "flawed elections conspiracy theory" (2015).

Other studies have pointed to further types of effects of search engines during elections. Also based on data from experiments, Hindman (2009) and Halavais (2009) have expressed concerns at Google's tendency to favor dominant information sources, operating a model in which they "both contribute to the selection of more prominent sites, and in turn are more influenced by them" (Halavais 2009: 59)<sup>8</sup>. This article contributes to analyzing the potential impacts of Google's search results during electoral processes, by proposing a methodology that complements studies based on experiments.

## **2. The Brazilian 2018 Presidential Election**

In late 2018, Brazilian voters chose a new President, Vice-President, representatives for the House and two-thirds of the Senate, as well as new State Governors and State legislators. The results of this particular election have been characterized by political scientists as "disruptive" (e.g. Abranches, 2018: 7; Moura and Corbenllini, 2019: 30), because they represented a breach with the organization of political forces and the party system of the last two and a half decades.

---

<sup>7</sup> Singhal worked at Google until February 26, 2016.

<sup>8</sup> Yet other studies have focused not on what Google offers, but on what users demand. Based on data from Google Trends (a tool that provides archived search records that show fluctuations in the popularity of keywords), Trevisan et al. (2018) show that, both in the United States and in the United Kingdom, high-profile mediated events such as TV debates between Presidential candidates constitute key drivers of Internet searches in elections.

Nothing represents this rupture better than the election of the new President, Jair Bolsonaro, candidate of the until then insignificant Partido Social Liberal (Social Liberal Party – PSL). Bolsonaro, formerly a Congressman elected from the state of Rio de Janeiro, won a highly contentious second round dispute against Fernando Haddad, candidate of the coalition led by the Partido dos Trabalhadores (Workers’ Party - PT). Results also showed high turnover rates in the composition of the National Congress, and an increase of the already highly fragmented political party system. We do not aim to analyze the complex and multiple causes that led to these results. It is important to point out, however, that we cannot understand the disruptive impacts of the 2018 elections without considering the role of digital technologies.

In the 2018 elections there were a new set of regulations in place, which, together with the increased access to the Internet, helped make digital arenas more relevant than in previous years. For the first time, candidates could lawfully pay for online electoral advertising. The Electoral Courts typified this practice of advertising through the services of social media platforms as "boosting content" (TSE, 2017: 1), which consists, under the terms of Resolution 23.551 of the Superior Electoral Court (TSE), of "the mechanism or service that, by contracting with Internet application providers, leverage the reach and dissemination of information to reach users who would not normally have access to its content"<sup>9</sup> (Idem). In addition to social media platforms, the Electoral Justice included within the definition of payment of "boosting content" (Ibidem) the "content resulting from Internet search application [e.x. Google]" (Ibidem).

In this context, Facebook and Google officials signed a memorandum of understanding with the Brazilian Electoral Court, in which they agreed to "fight against the misinformation generated by third parties" (TSE, 2018: 1). In addition, these platforms had to meet the requirements of the Brazilian authorities regarding the constitution of publicity mechanisms for the contracts of their services during the electoral process. While Facebook, Instagram and Google agreed to meet these demands, Twitter decided not to provide advertising services for campaigns, because it could not offer the "appropriate tools to facilitate this transparency" (TWITTER, 2018, apud Wakka, 2018:1) required by Brazilian law.

---

<sup>9</sup> What the Resolution 23.551 of the TSE defines as "internet application providers", we refer to in this article as "social media platform".

For candidates' electoral strategies, the "boosting content" in Google's search engine was important in order to publicize the campaign's official sites and social media pages, and also to present counter-information about a certain news or attack that the candidate was facing.<sup>10</sup> According to official data on campaign expenditures (presented by the candidacies), 571 candidates paid Google the total amount of R\$ 6,910,013.86 - approximately US\$ 1.7 million - to "boost content" during the election period (TSE, 2019). They also attempted to influence ranking results, and strived to change the contents by editing websites such as Wikipedia entries.<sup>11</sup>

These efforts are justified by Google's importance as Brazil's most popular search engine. Its relevance can also be assessed in terms of searches related to the 2018 elections. According to data from Google Trends, the names of three Presidential candidates were among the most searched terms that year: "Lula" (former President Luis Ignacio Lula da Silva, who was barred from running by electoral officials), "Fernando Haddad", and "Jair Bolsonaro". In addition to those, the term "Elections 2018" was one of the most searched in the year, along with questions related to the electoral process: "Why vote for Bolsonaro?", "Why not vote for Bolsonaro?", "Lula's Prison" and "Lula's Trial"<sup>12</sup>.

### **3. Interacting with Google**

In this section we present our research design and results.<sup>13</sup> First, we specify the process of creation of Google accounts. Next, we show the types of information sources offered as results by Google for the search of terms related to the 2018 elections. We then explain why and how Natural Language Analysis was used to analyze the proximity of words to the search terms. Finally, we compare results of searches using the names of the two second-round Presidential candidates, "Bolsonaro" and "Haddad".

#### **3.1 Data Gathering Method**

---

<sup>10</sup> Information obtained through semi-structured interviews with actors involved with local and national campaigns, undertaken in Brasilia, Rio de Janeiro and Sao Paulo between October of 2018 and January of 2019.

<sup>11</sup> Idem.

<sup>12</sup> See more in Brazil Google Trends, "Year Research of 2018", in "Search" (Busca), and "News" (Acontecimentos). Available at: <<https://trends.google.com.br/trends/yis/2018/BR/>>

<sup>13</sup> For a more detailed presentation of our research design, see Stabile et al 2019.



The methodology developed to analyze Google's role as information provider during the elections consisted of systematically collecting the search results displayed for given keywords to a set of logged accounts during the official electoral period. This task was accomplished in two sequential steps. The first consisted of creating fifteen different accounts that simulated different types of Google users, and the second was to collect data on searches done through these accounts.

The fifteen Google accounts created for this research varied in terms of ideology and gender: "left", "right", "woman", "man" and "neutral". For each of the four first types, we generated a list of websites and specific search terms, which were used to record a browsing "history". This "training" process occurred from June 18, 2018, until the end of the Brazilian elections, on October 28 of the same year. The algorithm script for capturing data from searches performed by the created accounts was developed through scraping data from the Google results page. The script was executed with Selenium, a program that automates web browser navigation and is used to emulate users' interaction with Google's search service. The logical procedure that the algorithm performed is shown in Figure 1.

**Figure 1 - Google Data-Capturing Procedure**

1. **For each user account:**
  - 1.1. Access <https://accounts.google.com>
  - 1.2. Authenticate with login and password
  - 1.3. **For each keyword:**
    - 1.3.1. Type keyword in search box
    - 1.3.2. Wait for result, take a screenshot and save the page's HTML
  - 1.4. **Logout**

Source: authors' own elaboration.

**Figure 2 - Textual-Content Capturing Procedure**

1. **For each URL stored:**
  - 1.1. Retrieve the HTML content of the page
  - 1.3. **For each paragraph tag in the HTML page:**
    - 1.3.1. Extract textual content from <p> tag.
    - 1.3.2. Store data in comma-separated file

**Source:** authors' own elaboration.

We ran the procedure in Figure 1 between August 17 and October 28, 2018. This resulted in 10,396 searches and 235,570 URLs. Each of these URLs provides a link to a specific page, for a specific search query, and for each of the 15 accounts created. From this body of URLs, 8,883 were classified as unique. However, a relevant portion of them were offline or displayed an error message when we tried to access the address. After cleaning the dataset, we ended up with 5,199 unique URLs. The text collected from these pages (procedure in Figure 2) resulted in 2,734,999 words, of which 160,756 were unique.

**Table 1** - Number of Searches According to Keywords, August 17-October 28, 2018

<b>Keywords</b>	<b>First Round</b>	<b>Second Round</b>	<b>Total</b>
<b>Presidential candidates keywords</b>			
<b>Haddad</b>	<b>844</b>	<b>75</b>	<b>919</b>
<b>Bolsonaro</b>	832	78	910
<b>Lula</b>	865	-	865
<b>Alckmin</b>	852	-	852
<b>Marina Silva</b>	829	-	829
<b>Cabo Daciolo</b>	805	-	805
<b>Ciro Gomes</b>	798	-	798
<b>João Amoedo</b>	783	-	783
<b>Álvaro Dias</b>	726	-	726
<b>Boulos</b>	634	-	634
<b>Other keywords</b>			
<b>Candidatos à Presidência (presidential candidates)</b>	<b>785</b>	<b>62</b>	<b>847</b>
<b>Lugar de votação (Voting place)</b>	662	36	698
<b>Eleições 2018 (Elections 2018)</b>	613	57	670
<b>fraude nas urnas (fraud in voting machines)</b>	-	60	60

Source: authors' own elaboration.

### 3.2 Where does Google take searchers?

The analysis of the URLs from the search database shows the sources of information potential voters have access to through the search engine, when typing specific election-related keywords. Tables 2 and 3 present search results according to a typology of media outlets: 1. Traditional or mainstream media - includes established vehicles with large audiences and recognition, and that are present in more than one type of platform (i.e. television channel and website); 2. Blog or journalistic page - comprehends pages or blogs by journalists, which may or may not be part of a mainstream vehicle, but do not necessarily share its communication strategy or ideological profile; 3. Social media page - this category includes links to social media platforms, such as a Facebook page or Twitter account; 4. Page or blog associated with political parties - channels linked to official political party media and their supporters; 5. Wikipedia - includes all the URLs from the collaborative research platform; and 6. Public portals - includes links to public administration websites, such as the House of Representatives, the Senate, electoral officials, and others (Valente and Franco, 2019).

As shown in Table 2, the distribution of results privileged mainstream online outlets. In total, a little over half of the links (51.4%) lead to traditional, established national media outlets - G1 (13.8%), Veja (8.1%), Estadão (5.7%), Folha de São Paulo (5.4%), UOL (5.1%), El País (2.8%), Exame (2.7%), BBC (2.5%), Gazeta do Povo (1.7%), O Globo (1%), Terra (1%), Carta Capital (0.8%) and IstoÉ (0.8%). Wikipedia, a collaborative research platform, was the third most common URL shown. Social media platforms - Facebook (4%), Twitter (2.6%) and YouTube (1.6%) - concentrated 8.2% of the results, and the website from the Superior Electoral Court (TSE) concentrated 1.3% of the total.

These results show that, at least in the case of the accounts we created in the context of the Brazilian 2018 elections, Google tended to direct users to established, mainstream media outlets for more information on election-related searches. This is perhaps not a surprising result, given the fact that these outlets are owned by corporations that have the resources that allow them to be better positioned in the ranking of search engines. Perhaps more striking is the fact that, as Table 2 shows, distribution is highly concentrated in only a

few of these outlets. The domains associated with two media vehicles, G1 and Veja, were the most common ones<sup>14</sup>. Together, they concentrated almost 22% of all URLs displayed in our database. These findings confirm experiment-data on Google results (Hindman, 2009; Halavais, 2009). On the one hand, this may be seen as a positive result, because of the absence of the plethora of alternative media outlets that were created during the election, many of which channeled misinformation campaigns (Hindman, 2009; Halavais, 2009). On the other hand, this result echoes recent researchers' concerns that search engines favor dominant information sources and do not present a plurality of opinions and perspectives. As said by Hindman, *some sites consistently rise to the top of Yahoo!'s and Google's search results; some sites never get indexed by search engines at all* (Halavais, 2009, p. 15).

**Table 2** – Most Frequent Domains in the Search Results Database, by Type of Media

<b>Domain</b>	<b>Type of Domain</b>	<b>N</b>	<b>%</b>	<b>% acum.</b>
<b>g1</b>	Mainstream Media	30,107	13.8	13.8
<b>veja</b>	Mainstream Media	17,787	8.1	21.9
<b>wikipedia</b>	Wikipedia	16,193	7.4	29.3
<b>estadao</b>	Mainstream Media	12,490	5.7	35.0
<b>folha</b>	Mainstream Media	11,704	5.4	40.4
<b>uol</b>	Mainstream Media	11,213	5.1	45.5
<b>facebook</b>	Social Media Page	8,726	4.0	49.5
<b>elpais</b>	Mainstream Media	6,159	2.8	52.3
<b>exame</b>	Mainstream Media	5,995	2.7	55.0
<b>infomoney</b>	Mainstream Media	5,568	2.5	57.6
<b>bbc</b>	Mainstream Media	5,399	2.5	60.1
<b>brasil247</b>	Blog or journalistic page	4,970	2.3	62.3
<b>twitter</b>	Social Media Page	4,474	2.0	64.4

<sup>14</sup> G1 is a daily online news portal owned by the largest national media corporation, Rede Globo (g1.globo.com), and Veja is a weekly printed and online magazine owned by the communication group Abril (veja.abril.com.br).

<b>gazetadopovo</b>	Mainstream Media	3,812	1.7	66.1
<b>em</b>	Mainstream Media	3,573	1.6	67.8
<b>youtube</b>	Social Media Page	3,448	1.6	69.3
<b>esmaelmoraes</b>	Blog or journalistic page	3,038	1.4	70.7
<b>tse</b>	Public Portal	2,741	1.3	72.0
<b>poder360</b>	Mainstream Media	2,431	1.1	73.1
<b>oglobo</b>	Mainstream Media	2,233	1.0	74.1
<b>terra</b>	Mainstream Media	2,219	1.0	75.1
<b>revistaforum</b>	Blog or journalistic page	2,168	1.0	76.1
<b>catracalivre</b>	Blog or journalistic page	2,155	1.0	77.1
<b>istoe</b>	Mainstream Media	1,753	0.8	77.9
<b>cartacapital</b>	Mainstream Media	1,653	0.8	78.7
<b>Other</b>		46,674	21	100.0
<b>Total</b>		218,683	100	-

**Source:** authors' own elaboration.

Mainstream media outlets are the most frequently URLs for all logged accounts, not varying significantly either according to ideology or gender (see Table 3).<sup>15</sup> However, the extent to which this type of media appears for each search term varied. The ratio of mainstream media and other media types reached 86% for searches of the term "Presidential candidates". In comparison, when searches included the names of the candidates, regardless of their ideological position the results generated a greater plurality of sources. The search for the names of the two candidates who reached the second round of the election, Jair Bolsonaro and Fernando Haddad, resulted in mainstream media outlet sources in approximately 62.7% and 60.6% of the cases, respectively.

The second most frequent type of media varies. For seven of the ten potential candidates in the first round of the electoral process, blogs and journalistic pages appear among the resulting sources of information. Social media pages, such as Facebook or Twitter, form the second most frequent type for two of these candidates. In the case of Jair Bolsonaro, social media pages showed a tie with journalistic pages as the second most

<sup>15</sup> For a more detailed analysis of this data, see Valente and Franco, 2019.

frequent type of media. In one case, that of Marina Silva (candidate of the Green Party), the second most frequent type of media was the collaborative encyclopedia site, Wikipedia.

As shown in Table 3, pages or blogs associated with political parties and public portals were less frequent among the search results on the platform. Consistently, public portals were relevant for searches with the term "voting place" (74.3%), and were the second most frequent type when searching for "Elections 2018" and "Presidential candidates" (10.9 and 7%, respectively).

**Table 3 - Average Presence of Types of Media Domains, by Search Keyword (%)**

<b>Search Keywords (names of candidates and four key topics related to the election)</b>	<b>Mainstream Media</b>	<b>Journalistic Blogs and Pages</b>	<b>Social Media Pages</b>	<b>Wikipedia</b>	<b>Pages or Blogs linked to Political Parties</b>	<b>Public domains</b>	<b>Others</b>
Alckmin	67.0	10.0	5.0	9.0	4.0	--	5.0
Álvaro Dias	57.0	16.0	10.0	8.0	--	7.0	2.0
Boulos	58.4	20.8	7.9	9.9	1.0	--	2.0
Bolsonaro	62.7	8.8	8.8	7.8	5.9	--	5.9
Cabo Daciolo	56.0	13.0	9.0	10.0	--	7.0	5.0
Ciro Gomes	69.0	8.0	11.0	8.0	2.0	--	2.0
Haddad	60.6	19.2	7.1	7.1	--	--	6.1
João Amoedo	54.0	8.0	14.0	8.0	7.0	1.0	8.0
Lula	56.6	16.2	9.1	10.1	6.1	--	2.0
Marina Silva	68.0	7.0	8.0	9.0	6.0	--	2.0
Candidatos a Presidência (Presidential candidates)	86.0	5.0	--	--	--	7.0	2.0
Eleições 2018 (Elections 2018)	75.2	8.9	3.0	--	--	10.9	2.0
Fraude nas urnas (fraud in voting machines)	76.8	13.1	8.1	--	--	--	2.0
Lugar de votação (voting place)	23.8	1.0	--	--	--	74.3	1.0

**Source: authors' own elaboration.**

### **3.3 The Proximity of Terms in Google Searches**

To better understand the contents of the information Google displayed for its users during the election period, it is important to go beyond an analysis of the distribution of links per type of media outlets, to analyze the texts. In order to do so, we generated a second dataset: we visited each unique URL in the dataset discussed in previous sections and saved their textual content to a key-value database. This step generated a text body of approximately 2 million words. We used the Word2Vec algorithm to understand which terms came closest to the key players in the second round of the Presidential election: Jair Bolsonaro and Fernando Haddad.

### **Introduction to Natural Language Processing and Word2Vec**

One of the most fundamental characteristics of any human society is the language its members use to communicate with one another as it has the potential to show the inner societal structure of a community. Analyzing language, nonetheless, is a very challenging task: aspects such as grammar, semantic structures and subjective aspects such as irony represent key sources of difficulty. However, recent developments in cloud computing and big data led to the development of Natural Language Processing (NLP) tools and techniques which allow us to work with huge sets of data. Uses of NLP techniques are manifold: from sentiment analysis to automated translation between languages, NLP allows us to transform audio into text and vice-versa, parts-of-speech tagging, among others.

To analyze the texts in our content database, we used a specific NLP algorithm, Word2Vec, which allows us to build proximity relationships among words, considering the context. The concept of “n-grams” enables us to go beyond simply counting words and is one of the fundamental ideas behind Word2Vec. An n-gram is a set of words which are likely to occur together in a body of text, where n is the number of words. For instance: “like a baby” is a 3-gram. Furthermore, the Word2Vec algorithm considers the context in which a word or sentence occurs in a body of text, by determining the distance between these words or n-grams in relation to the rest of the text. This, in turn, enables us to understand the



semantics of a language better than we would be able to do by relying only on frequency analysis.

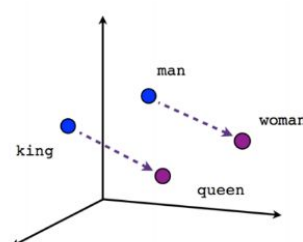
Word2Vec is described in its original presentation to the scientific community as an algorithm which "presents two new model architectures for learning distributed representations of words that try to minimize computational complexity" (Mikolov et al 2013: 4): The *Continuous Bag-of-Words Model* and the *Continuous Skip-gram Model*. Both of them minimize computational complexity by improving upon the shared point of failure found in other proposals: the non-linear hidden layer. A neural network is able to learn both linear and nonlinear functions such that its learning is based on the amount of training it receives and the neural network itself. The non-linear layer can be considered as an element which allows a neural network to learn non-linear functions. They are very powerful but also very consuming in terms of computing power, such that we might be able to perform with higher efficiency by using simpler models.

This is where both architectural models gain the spotlight: the Continuous Bag-of-Words Model allows for a projection layer which is shared for all words instead of a hidden non-linear layer, which means that all words will be projected to a single position while having their weights averaged, so that the order of appearance of a word in a text does not influence its impact when evaluating other words, only the context in which it appears. The Continuous Skip-gram Model goes somehow in the opposite direction: it works in a more focused manner, in which it tries to predict the context based on a target word instead of trying to predict the word based on the context. The result of running these models against a body of text generate word embeddings, which are vector representations of the relationship between words in a body of text in a matricial space. Thus, if a word tends to appear in a certain context that is similar to another word appearing in the same context, they will have proximate word embeddings – meaning that they are closer to each other in an N-dimensional space. The following example illustrates this: if we consider the phrase "It is a beautiful day in Chicago" and the sentence 'It is a beautiful day in New York', the Word2Vec algorithm will understand Chicago and New York as being closely related due to the fact that they appear in very proximate contexts. Therefore, the word embeddings for both will be similar and the distance between the vectors that represent both words will be very short.

This is where both architectural models gain the spotlight: *Continuous Bag-of-Words Model* allows for a projection layer which is shared for all words instead of a hidden non-linear layer, which means that all words will be projected to a single position while having their weights averaged so that the order of appearance of a word in a text does not influence its impact when evaluating other words, only the context in which it appears. The *Continuous Skip-gram Model* goes somehow in the opposite direction: it works in a more focused manner, in which it tries to predict the context based on a target word instead of trying to predict the word based on the context. The result of running these models against a body of text generate what we call *word embeddings*, which are vector representations of the relationship between words in a body of text in a matricial space, such that if a word tends to appear in a certain context that is similar to another word appearing in the same context, they'll have proximate word embeddings - meaning they're closer to each other in an N-dimensional space. The following example illustrates this: if we consider the phrase "*It is a beautiful day in Chicago*" and the sentence '*It is a beautiful day in New York*', the Word2Vec algorithm will understand Chicago and New York as being closely related due to the fact that they appear in very proximate contexts. Therefore, the word embeddings for both will be similar and the distance between the vectors that represent both words will be very short.

This point leads us to one of the most impressive features in Word2Vec: it allows us to perform mathematical operations such as addition, subtraction, sine and cosine between words as if they were numbers. This feature led to the already classic Word2Vec example for the English language:

$$\text{King} - \text{Man} + \text{Woman} = \text{Queen}$$



Male-Female

16

---

<sup>16</sup> Image taken from <https://medium.com/scaleabout/a-gentle-introduction-to-doc2vec-db3e8c0cce5e> which exemplifies an example we address later on the text.

The ability to operate upon words as if they were vectors in a N-dimensional space allows us to understand subjective aspects of a given language (assuming we dispose of a large enough dataset) which were computationally imperceptible prior to Word2Vec - the sheer analysis of n-grams will give hints in the direction of how a certain language works, but Word2Vec is powerful enough to let us understand nuances that reflect how a given society is organized in a much more subjective manner.

### 3.3.2 Applying Word2Vec to the Content Dataset

Before we delve into the Word2Vec analysis of our dataset, we will outline the procedure used in order to acquire the data for our study<sup>17</sup>. First, we built a scraper for the pages listed, which visited and searched for paragraph elements on each page. When it found one, we downloaded its content to a new dataset containing the URL visited and the textual content. While we considered this heuristic to be quite fragile - page owners could store text in different HTML tags, or change the structure of the page while we were scraping - we manually visited the 15 most frequent unique domain names in Google's results and made sure they were storing their content in a way that we could access and download. Since they represented approximately 65% of all results (as shown in the previous section), we believe we were able to build a representative database of textual content, consisting of approximately 2 million words.

**Figure 3** - Sample of Data Collected, by Type of Result, URL Address and Result Title

result_type	result_positi	result_url	result_title
noticia	1	<a href="https://g1.globo.com/politica/eleicoes/2018/noticia">https://g1.globo.com/politica/eleicoes/2018/noticia</a>	Bolsonaro diz que vai tirar Brasil da ONU se for
noticia	2	<a href="https://www1.folha.uol.com.br/colunas/celso-rocha">https://www1.folha.uol.com.br/colunas/celso-rocha</a>	Marina contra Bolsonaro - 20/08/2018 - Celso R
noticia	3	<a href="https://brasil.elpais.com/brasil/2018/08/18/politica">https://brasil.elpais.com/brasil/2018/08/18/politica</a>	Vitória de Bolsonaro pode tirar investimentos
noticia	4	<a href="https://www.poder360.com.br/eleicoes/alckmin-ter">https://www.poder360.com.br/eleicoes/alckmin-ter</a>	Alckmin terá 11 minutos diários na TV e Bolso
noticia	5	<a href="https://veja.abril.com.br/blog/noblat/tem-cheiro-ou">https://veja.abril.com.br/blog/noblat/tem-cheiro-ou</a>	Tem cheiro (ou mau cheiro) de Bolsonaro no ar
noticia	6	<a href="http://www3.redeTV.uol.com.br/blog/reinaldo/bolsc">http://www3.redeTV.uol.com.br/blog/reinaldo/bolsc</a>	Bolsonaro prova: Trump não é de nada; mo

Source: author's own elaboration.

Then we proceeded to run the Word2Vec algorithm on our new database. We used the following algorithm to do so:

<sup>17</sup> For a more detailed explanation, see Moura 2019.

**Figure 4 – Word2Vec Algorithm Applied to the Database**

Algorithm	Applies Word2Vec to a corpus of text
1:	<b>procedure</b> WORD2VEC
2:	<i>bodyOfText</i> $\leftarrow$ <i>loadBodyText().lower().removeAccents().applyIramuteqFilter()</i>
3:	<i>wordTokens</i> $\leftarrow$ <i>nlTK.word_tokenize(bodyOfText, language = "portuguese")</i>
4:	<i>word2vec_holder</i> $\leftarrow$ <i>Word2Vec(wordTokens, size = 10, min_count = 2)</i>
5:	<i>word2vec_holder.train(wordTokens, total_examples = len(wordTokens), epochs = epochs)</i>
6:	<i>print(word2vec_holder.wv.most_similar('some_query_term', topn = 50))</i>

Source: Moura, 2019.

## **Nazi, incompetent, authoritarian, rapist: the language of a disruptive election**

For this article, we analyzed the proximity of words in the texts of the results of searches for the names of the two most relevant Presidential candidates: Bolsonaro (Jair Bolsonaro) and Haddad (Fernando Haddad). The results, presented in Figure 4, offer a general birds-eye view of the way in which the election was debated. We see both negative and positive words associated with the two candidates, such as "nazista" (nazi) and "autoritarismo" (authoritarianism), closely associated with Bolsonaro, and "Incompetente" (incompetent) and "corruptor" (which comes from the word "corrupt"), associated to Haddad. At the same time, Bolsonaro appears related to "pragmatismo" (to be pragmatic, a trait which the candidate supposedly possessed, according to his supporters), while Haddad appears close to "escolarizar" (word related to schools and education - which might be explained due to the fact that he has been an academic for 35 years and a former Ministry of Education).

Other proximities among words provide insights into the election: "louvor" (to pray) is associated to Bolsonaro, revealing the relevance of the candidate's alliance with the Evangelical Churches, whose members are expected to compose the majority of the Brazilian population by 2032<sup>18</sup>. In order to explore a bit further the religious influence in the election, we also plotted which word would be the counterpart to Bolsonaro's "louvor". Our finding is that Bolsonaro is to "louvor" as Haddad is to "aula" (class), which fits with the reality of the electoral process: while Bolsonaro received major support from

<sup>18</sup> For further discussion see

<http://www.ihu.unisinos.br/78-noticias/585245-transicao-religiosa-catolicos-abaixo-de-50-ate-2022-e-abaixo-do-percentual-de-evangelicos-ate-2032>, access on October 13th, 2019

aforementioned evangelical churches, Haddad was popular among university students. This relationship can also be expressed in this equation:

$$\text{Haddad} + \text{Bolsonaro} - \text{louvor} = \text{aula}$$

**Table 4 Words Most Proximate to “Bolsonaro” and “Haddad”**

Words proximate to "Bolsonaro"			Words proximate to "Haddad"		
Word		Coefficient	Word		Coefficient
descumprir	disobey	0.9965	incompetente	incompetent	0.9829
capaz	capable	0.9935	doleiro	money changer	0.9813
solidarizar	sympathize	0.9902	ideologia	ideology	0.9773
inimigo	enemy	0.9768	Correios	Correios	0.9769
Dirceu	Dirceu	0.9737	hebraico	hebrew	0.9758
Kim	Kim	0.9618	alencar	alencar	0.9449
mente	lies	0.9257	maluf	maluf	0.9310
pragmatismo	pragmatism	0.9217	escolarizar	school	0.9265
indenizar	indemnify	0.9182	Doria	Doria	0.9257
nazista	Nazi	0.9169	corrupto	corrupt	0.9225
rato	rat	0.9145	enriquecer	enrich	0.9220
museu	museum	0.9032	Buarque	Buarque	0.9192
gadelha	gadelha	0.9019	Estuprador	Rapist	0.9115
rouco	hoarse	0.9007	Diesel	Diesel	0.9087
autoritarismo	authoritarianism	0.8991	prepotente	boastful	0.9086
desequilibrado	unbalanced	0.8966	disperso	scattered	0.9074

**Source:** Moura 2019.

The words "gay" and "ideologia" (ideology), which appear proximate to “Haddad”, refer to another polarizing issue of the election. While the first one displays the proximity of the LGBT community with the leftist candidate, the second might be related to an expression made famous by people who oppose LGBT rights: "ideologia de gênero" (gender ideology). While further research is needed in order to verify this claim, Haddad openly declared his support for minorities and as such was criticized by his opponents.

Other aspect of the election is also clearly stated by the words most related to both candidates: The debate around corruption. Haddad's Party (Partido dos Trabalhadores - Workers' Party) has been a target of investigation for connection with major fraudulent schemes for years, which recently resulted in the arrest of former President and PT leader Luiz Inacio Lula da Silva<sup>19</sup>. These incidents affected Haddads' campaign, as we can see in the words "doleiro" (a person who exchanges foreign currency, often illegally), "corruptor", "enriquecer" (to get rich, which is fair to assume has a negative meaning), "preso" (to be locked up) and "Maluf" (a Brazilian politician famous for being on the interpol list due to his corruption crimes). Although the term doesn't appear on our list, it is important to remember that operation Car Wash - one of the major anti-corruption investigations ever carried by a Brazilian authority - was strongly active during the campaign, and might have an influence on the words related to corruption we see on our list.

Further analysis was done to understand whether the results varied according to the ideological profile of the Google accounts created for this research: left-wing and right-wing users. The following tables show the terms most associated with "Haddad" and "Bolsonaro" for both of them:

**Figure 5 - Proximity of words to "Bolsonaro" and "Haddad"**

---

<sup>19</sup> See <https://www.terra.com.br/noticias/brasil/politica/liberdade-prisao-domiciliar-ou-continuar-na-cadeia-os-possiveis-desfechos-do-novo-julgamento-de-lula.3d704f91980c2c339023ed17b6d7effadc4kn98b.html>, access on October 13th, 2019.

Words proximate to "Bolsonaro" <i>left accounts</i>			Words proximate to "Haddad" <i>left accounts</i>		
Word		Coefficient	Word		Coefficient
projeto	project	0.9054	levantar	rise	0.9681
jobim	jobim	0.8635	covarde	coward	0.9680
ironizar	irony	0.8536	receio	ideology	0.9613
manipular	manipulate	0.8408	disseminar	Correios	0.9549
teatro	theater	0.8380	boneco	hebrew	0.9529
telejornal	newscast	0.8316	propriedade	alencar	0.9372
homenagear	honor	0.8174	aliar	maluf	0.9197
progresso	progress	0.8155	despudor	school	0.9185
mole	soft	0.7827	obra	Doria	0.9178
correio	correio	0.7787	protesto	corrupt	0.9039
indiciar	indict	0.7760	ministerial	enrich	0.8894

Source: Database of the Research Google Results: Brazilian General Election 2018, Resocie, 2018.

**Figura 6 - Proximity of words in right-wing accounts to "Bolsonaro" and "Haddad"**

Words proximate to "Bolsonaro" <i>right accounts</i>			Words proximate to "Haddad" <i>right accounts</i>		
Word		Coefficient	Word		Coefficient
presidenciável	presidential	0.9997	tranquilao	too calm	0.9869
odebrecht	odebrecht	0.9989	ironia	irony	0.9606
processo	trial	0.9988	Hitler	Hitler	0.9568
unir	unite	0.9984	hipocrisia	hipocrisy	0.9329
privatizar	privatize	0.8663	mamar	mamar*	0.8936
caos	chaos	0.8695	ensino	teaching	0.8795
mito	mito	0.8652	saneamento	sanitation	0.8717
desenvolvimentista	developmentalist	0.8203	renovar	renew	0.8519
mito	mito	0.8173	assertivo	assertive	0.8424

Source: Database of the Research Google Results: Brazilian General Election 2018, Resocie, 2018.

As we can see from Figures 5 and 6, the polarization occurred on both sides of the discussion, with recurrent violent attacks between them along with several other interesting aspects. For left-wing voters, terms associated with Bolsonaro suggest in general a discussion around the weaker characteristics of the opposing candidate: words such as "Ironizar" (to ironize), "manipular" (to manipulate), "telejornal" (to appear in a news programme tv show) - very probably in reference to his participation in the most watched news programme in the country, *Jornal Nacional*, which generated strong opinions in the audience specially for his mentioning of the dictatorship period the country went through from 1964 to 1985<sup>20</sup> - and "homenagear" (to honor), most likely in reference to his mentioning of admitted torturer Carlos Brilhante Ustra. Surprisingly, however, the terms associated with "Haddad" for left-wing accounts are not necessarily supportive of the candidate. "Covarde" (coward) and "despudor" (shameless), for example, open the discussion for multiple possibilities as to why such words appear for the candidate even among his supporters.

The analysis for right-wing accounts shows that words related to Bolsonaro are more supportive of the candidate. Words such as "presenciável" (presidential), "mito" (legend, an expression used by Bolsonaro's supporters in order to praise their candidate) and "desenvolvimentista" (developmentalist) all suggest a favourable attitude towards the candidate. There are also words which raise questioning, such as "caos" (chaos) and "petismo" (a word related to the Workers' Party). Previous research suggests<sup>21</sup> that one of the main topics of discussion for Bolsonaro's supporters was the opposing party. Another word that catches the eye is "algoritmo" (algorithm). Bolsonaro is accused of having used automated social media posting in favor of his campaign<sup>22</sup> and of using Whatsapp bots in order to support his presidential campaign<sup>23</sup>, all themes strongly associated to discussions world-wide around misinformation and its effects on democracies.

---

<sup>20</sup> See

<https://g1.globo.com/jornal-nacional/noticia/2018/08/28/jair-bolsonaro-psl-e-entrevistado-no-jornal-nacional.g.html>, access on October 13th, 2019.

<sup>21</sup> See

<https://institutodademocra.wixsite.com/meusite/feed-de-posts/o-debate-sobre-a-seguran%C3%A7a-da-urna-eletronica-brasileira-no-youtube>, access on October 17th, 2019.

<sup>22</sup> See

<https://www1.folha.uol.com.br/poder/2018/12/apos-eleicao-perfis-falsos-e-robos-pro-bolsonaro-continuam-ativos-aponta-estudo.shtml>, access on October 13th, 2019.

<sup>23</sup> See

<https://www1.folha.uol.com.br/poder/2018/10/estudo-aponta-para-automacao-no-envio-de-mensagens-e-orquestracao-entre-grupos-de-whatsapp-pro-bolsonaro.shtml>, access on October 13th 2019.



In parallel, for right-wing accounts words associated with "Haddad" put the candidate under unfavourable light: "Ironia" (Irony), "Hitler", "Hipocrisia" (hypocrisy), "mamar" (a pejorative term for the relation between the Workers Party and the government) all show the violent speech found in the electoral period. Terms that could be positive and appear in this list such as "Ensino" (to teach), "Saneamento" (sanitary), "renovar" (to renovate) and "assertivo" (to be assertive) need further research in order for us to determine if they are associated with "Haddad" in a positive or negative manner.

It is also important to keep in mind the limitations of the analysis presented. First, Word2Vec algorithm possesses an inherent degree of uncertainty, which means that two different people running the same algorithm<sup>24</sup> on the same dataset might see two different results. This is expected and does not necessarily represent a programming error or a human mistake, but the intrinsic nature of the algorithm itself. As we mentioned earlier, the fact both models used for projecting words before training them in an N-gram Neural Network Language Model (N-gram NNLM)<sup>25</sup> do not try to represent the original data with perfect accuracy produces varying results which are not wrong - only different from each other. In addition, the algorithm expects to work with enormous amounts of data. Although we have collected a significant dataset of about 2 million words, the smallest dataset used as an example in the original paper which describes Word2Vec (Mikolov, Tomas, Chen, Corrado, Dean, 2013: 4) contains approximately 24 million records, a significant difference which certainly affects the effectiveness of our analysis process.

## Conclusion

Google acts as a bridge between internet users and information sources, a kind of *gatekeeper among gatekeepers*. In spite of its relevance, its role remains understudied, in a field that has tended to privilege the analysis of social media platforms such as Facebook, Twitter or YouTube. This article presents a first effort in developing an innovative method to analyze internet search engines and its effects on the provision of information during political events.

---

<sup>24</sup> Which is open-source and available online at <https://github.com/teogenesmoura/pesquisaGoogle>

<sup>25</sup> A good discussion around N-gram NNLMs can be found here: <https://arxiv.org/abs/1708.07252>

The analysis of search results pages shows that Google overwhelmingly tended to send our logged accounts to mainstream media outlets during the months leading to and during the Brazilian 2018 elections. More interestingly, results were highly concentrated in only a handful of these outlets. On the one hand, this could be considered a positive result, because of the absence of the plethora of new alternative media outlets, many of which were created during the election to channel misinformation campaigns. On the other hand, this result echoes recent researchers' concerns that search engines favor dominant information sources and do not present a plurality of opinions.

In order to better understand what kind of information Google users access, we also did a content analysis of the texts of the links that appeared in the search results for our logged accounts. The analysis of approximately 2 million words portrays an electoral process delineated by a narrative of violent polarization, where there is no clear space for building dialogue among the various parties involved in the election.

The content analysis also shows that the search engine presents different results for users coming from varying political leanings. This finding is relevant, considering that other studies have not found significant differences in Google search results among users with various access habits (*apud* Hannak *et al*, 2013). Future studies should take a closer look into the potential effects of restricted information access according to ideology, based not only on the type of media that appears on the search results, but also on the contents of the texts within links.

We faced several obstacles to systematically gather search results pages over a period of several weeks. Specifically, in future works we will invest on robust computational infrastructure in order to gather data and run the analysis. We also need to do further research in order to better understand the effects that our training of ideological personas had on the results provided by Google. Finally, we intend to apply computational techniques with higher efficiency in smaller datasets such as ours.

In spite of these limitations, this work represents an important achievement. We indicated that it is possible as well as desirable to develop techniques to analyze the potential effects of search engines in democratic elections. We understand that building knowledge in this research field could boost interest and encourage new studies. It is important to note that our analysis was situated in Brazil, during the 2018 electoral period. As (*apud* Hannak *et al*, 2013) rightly pointed out in the discussion about their findings, the research results are

representative of this period, but they may not be in the future, because of the intrinsically dynamic characteristics of the search algorithms. Future studies will have to consider the difficulties in analyzing the structural changes in Google's search engine. This lack of transparency represents a constant challenge in understanding how their service operates.

## References

Abranches, Sérgio (2018). “Polarização radicalizada e ruptura eleitoral”. In: *Democracia em risco? 22 ensaios sobre o Brasil hoje*. Editora Companhia das Letras. São Paulo. p.07-26

Alexa (2019). “Top Sites Brazil”. Last access in 11.27.2019. Available at:<  
<https://www.alexa.com/topsites/countries/BR>>

BAND TV (2018). “Debate na Band: reveja na íntegra o 1º confronto entre os presidenciais”. 09/08/2018. *YouTube*. Last access in 08.19.2019, Available at:  
<[https://youtu.be/9EnJeUKwX\\_c?t=18](https://youtu.be/9EnJeUKwX_c?t=18)>

Belizário, Jefferson (2018). “Pesquisa do IBGE revela que 138 milhões de brasileiros possuem um smartphone”. *TudoCelular*. February 22nd, 2018. Last access in 08.29.2019, Available at  
<<https://www.tudocelular.com/android/noticias/n120658/Pesquisa-revela-indice-uso-smartphones-brasil.html>>

CETIC (2017). *TIC Domicílios: Pesquisa Sobre o Uso das Tecnologias da Informação e Comunicação nos Domicílios Brasileiros*. CGI.br. São Paulo, available at  
<[https://www.cetic.br/media/docs/publicacoes/2/tic\\_dom\\_2017\\_livro\\_eletronico.pdf](https://www.cetic.br/media/docs/publicacoes/2/tic_dom_2017_livro_eletronico.pdf)>

Epstein, Robert; Robertson, Ronald E (2015). “The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections”. *PNAS*. August 4th, 2015. Last accessed 12.12.2019, available at  
<https://www.pnas.org/content/early/2015/08/03/1419828112?sid=2096de02-a382-4aeb-89e0-df794a4ca03a>

Epstein, Robert (2015). “How Google Could Rig the 2016 Election”. *Politico Magazine*, August 19th, 2015. Last accessed in 03.13.2019, available at  
<https://www.politico.com/magazine/story/2015/08/how-google-could-rig-the-2016-election-121548>

Epstein, Robert (2016). “The new mind control”. *Aeon*. Essay, February 18th, 2016. Accessed in 03.13.2019, available at  
<https://aeon.co/essays/how-the-internet-flips-elections-and-alters-our-thoughts>

Granka, L., Joachims, T., & Gay, G. (2004). “Eye-tracking analysis of user behavior in WWW search”. *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (p. 478–479). New York: ACM Press.

Halavais, A (2009). *Search engine society*. Cambridge: Polity Press.

Hannak, A. et al (2013). "Measuring personalization of Web search", proceedings of the 22nd International Conference on the World Wide Web, 527-537.

Hillis, Ken, Michael Petit and Kylie Jarrett (2013). *Google and the Culture of Search*. New York: Routledge.

Hindman, M (2009). *The myth of digital democracy*. Princeton, NJ: Princeton University Press.

Kantar (2019). "Android vs. iOS". *Smartphone OS Market Share*. Accessed in 29.08.2019, available at <<https://www.kantarworldpanel.com/global/smartphone-os-market-share/>>

Moura, Maurício; CORBELLINI, Juliano (2019). *A Eleição Disruptiva: por que Bolsonaro venceu*. Editora Record. Rio de Janeiro.

Moura, Teógenes.(2019). *Monitorando e entendendo a eleição brasileira por meio de Processamento de Linguagem Natural*. Unpublished work, part of the requirements to finish the Bachelor's Degree in Computer Engineering, University of Brasília, Brasília.

Pan, Bing; Hembrooke, Helene; Joachims, Thorsten; Lorigo, Lori; Gay, Geri; Granka, Laura (2007). "In Google We Trust: Users' Decisions on Rank, Position, and Relevance". *Journal of Computer-Mediated Communication*. Vol. 12, Nº 3, de 2007. pp.801-823. Versão eletrônica consultada 13.03.2019, em <https://academic.oup.com/jcmc/article/12/3/801/4582975>

Pessoa, Gabriel Sá (2018). "Lula lidera intenções de voto, seguido por Bolsonaro, aponta pesquisa CNT". *Folha de São Paulo*, August 20th, 2018. Last accessed in 19.08.2019, available at <https://www1.folha.uol.com.br/poder/2018/08/lula-lidera-intencoes-de-voto-seguido-por-bolsonaro-aponta-pesquisa-cnt.shtml>

Pinto, Celi (2017). "A Trajetória Discursiva das Manifestações de Rua no Brasil (2013-2015)". *Lua Nova*. (100). São Paulo. pp.119-153. Last accessed in 19.08.2019, available at <http://www.scielo.br/pdf/ln/n100/1807-0175-ln-100-00119.pdf>

REDETV! (2019). "Debate presidencial na RedeTV!", aired in 08.17.2019. *YouTube*. Last accessed in 08.19.2019, available at <https://youtu.be/99SmMo1XqzQ?t=375>

Robertson, Ronald E. (2018). "When will Google Defend Democracy?". *The Conversation*, June 11th, 2018. Last accessed in 03.13.2019, available at <https://theconversation.com/when-will-google-defend-democracy-96838>

Singhal, Amit.(2015) "A Flawed Elections Conspiracy Theory". *Politico Magazine*, August 26th, 2015. Last accessed in 10.13.2019, available at <https://www.politico.com/magazine/story/2015/08/google-2016-election-121766>

Stabile, Max, Alexandre Arns, Alexandre Gomes & Teógenes Moura (2019). "E o Google? Técnicas de coleta dos resultados de pesquisa do Google", paper presented at the *VIII Compolítica 2019*, Brasília, 15-17 May, 2019.

Trevisan, Fillippo, Andrew Hoskins, Sarah Oates & Dounia Mahlouly.(2018). “The Google voter: search engines and elections in the new media ecology”, *Information, Communication & Society*, 21:1, 111-128. DOI: 10.1080/1369118X.2016.1261171

TSE.(2018). *Memorando de Entendimento*, June 28th, 2018. Tribunal Superior Eleitoral. Brasília. Last accessed 08.19.2019, last accessed <http://www.justicaeleitoral.jus.br/arquivos/memorando-de-atendimento>

Resolução nº 23.551, December 18th,2017. Coordenadoria de Jurisprudência. Tribunal Superior Eleitoral. Brasília. Last accessed in 08.19.2019, available at <http://www.tse.jus.br/legislacao-tse/res/2017/RES235512017.html>

TSE (2018). *TSE indefere pedido de registro de candidatura de Lula à Presidência da República*, notícia de 1 de setembro. Brasília. Last accessed in 08.19.2019, available at <http://www.tse.jus.br/imprensa/noticias-tse/2018/Setembro/tse-indefere-pedido-de-registro-d-e-candidatura-de-lula-a-presidencia-da-republica>

TSE (2019). *Consulta de Doadores e Fornecedores das Eleições de 2018: Google Brasil Internet Ltda*, from October 13, 2019. Available at: <<http://www.tse.jus.br/eleicoes/eleicoes-2018/divulgacandcontas#/consulta/doadores-fornecedores/2022802018>>

UOL.(2018). “Debate dos Candidatos à Presidência da República - 26/09”, from September 26. *YouTube*. Accessed 19.08.2019, available at: <<http://www.youtube.com/watch?v=MSgq-tZiAkU>>

Valente, Alana and Beatriz Franco. (2019). “‘Dá um Google aí’: buscando a relevância do Google nas eleições brasileiras de 2018”, paper presented at *VIII Compolítica 2019*, Brasilia, 15-17 May.

Vogel, Kimberly; MCCAFFREY, Cindy. (2000). “Google Launches Self-Service Advertising Program”, from 23 October. *GooglePress*, available at: <https://googlepress.blogspot.com/2000/10/google-launches-self-service.html>

Wakka, Wagner.(2018). “Twitter não vai permitir post pagos de campanhas eleitorais”. *Canaltech*, from 28 May. Available in: <https://canaltech.com.br/internet/twitter-nao-vai-permitir-post-pagos-de-campanhas-eleitorais-114708/>

Zuboff. Shoshana.(2019). *The Age of Surveillance Capitalism*. New York. Ed. Public Affairs.