

**News and Political Information Consumption in Brazil:  
Mapping the First Round of the 2018 Brazilian Presidential Election on Twitter**

COMPROP DATA MEMO 2018.4 – ONLINE SUPPLEMENT / OCTOBER 5, 2018  
*comprop@oii.ox.ac.uk*

Caio Machado  
Oxford University  
[@caiocvm](#)

Beatriz Kira  
Oxford University  
[@beakira](#)

Gustavo Hirsch  
Oxford University  
[@gustavoh1985](#)

Nahema Marchal  
Oxford University  
[@nahema\\_marchal](#)

Bence Kollanyi  
Oxford University  
[@bencekollanyi](#)

Philip N. Howard  
Oxford University  
[@pnhoward](#)

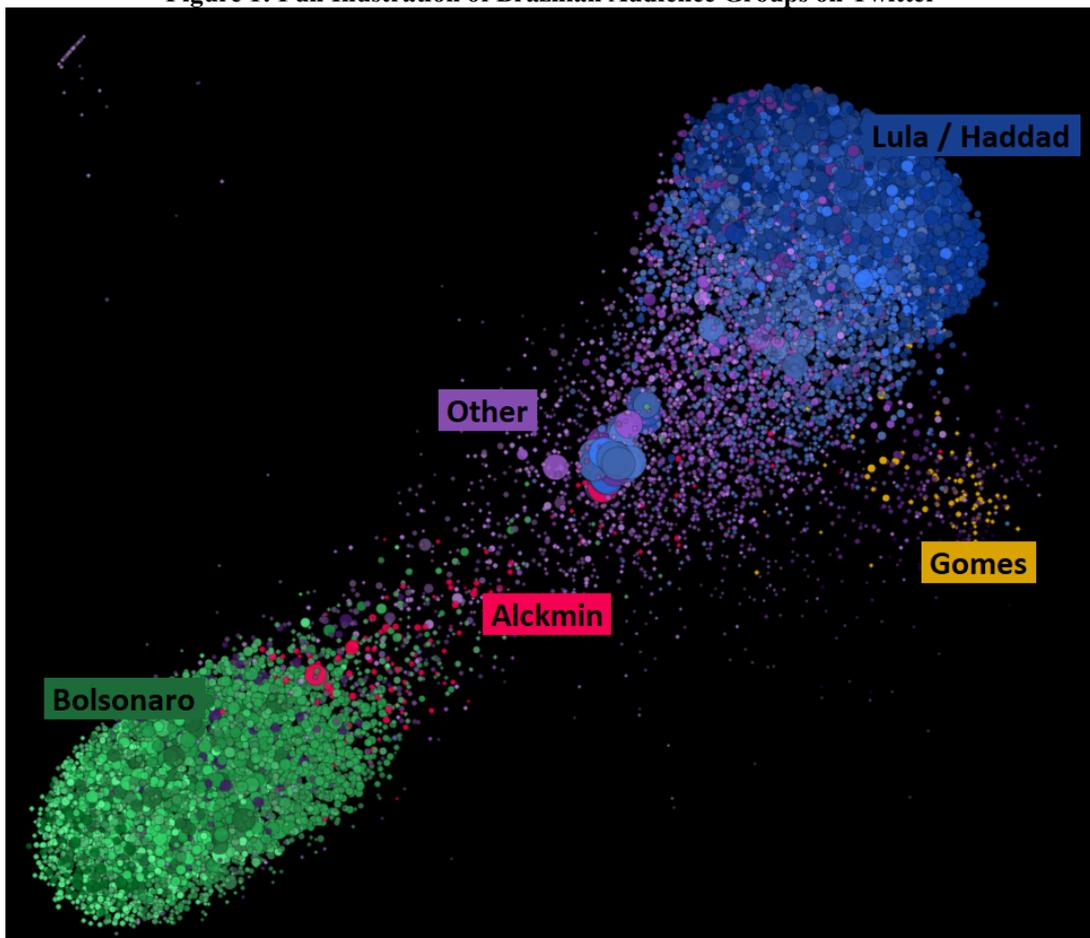
Thomas Lederer  
Graphika

Vlad Barash  
Graphika  
[@vlad43210](#)

**1. Brazilian Audience Groups on Twitter**

Each node in this network (*Figure 1*) represents an account on Twitter. Each node belongs to both a broad group and a smaller cluster within that group. A cluster is a collection of nodes with a shared pattern of interest while a group is a collection of clusters that are politically, culturally, or socially similar. The size of each node is proportional to the number of other nodes that follow it on Twitter. The color of each node is based on its parent cluster. *Figure 1* is the full color visualization of the Brazilian audience groups we found on Twitter.

**Figure 1: Full Illustration of Brazilian Audience Groups on Twitter**



*Source: Authors' calculations from data sampled 19/08/18-28/08/18. Note: Groups are determined through network association and our interpretation of the kinds of content these users distribute.*

The nodes are placed within the map using a Fruchterman-Reingold visualization algorithm. This works to place nodes on the map according to two principles: first, a “centrifugal force” acts upon each node to push it to the edge of the canvas; second, a “cohesive force” acts upon every connected pair of nodes to pull them closer together. Table 1 gives a full list of Twitter groups and associated clusters considered in this study.

<b>Table 1: Groups and Clusters for the Twitter Visualization</b>	
<i>Groups</i>	<i>Clusters</i>
Geraldo Alckmin	Journalists and Market Specialists
Ciro Gomes	Political Support Accounts   Bot-Like
Lula / Haddad	Lula and Dilma Supporters   Bot-Like PT Supporters and Activists Left Activists   Bot-Like Lula Supporters   Oppose Dilma’s Impeachment Ordinary People   Lula Livre Supporters PT Institutional Accounts PT Institutional and Individual Supporters PT Official Accounts, Politicians and Youth Wing Ordinary People   PT Supporters TV Stars Left-wing Supporters   Lula Livre Supporters News Media and Political Blogs Ordinary People and Protester   Left-wing PT and Left-Wing Politicians Lula Supporters   Bot-Like Left Influencers and Bloggers Left-wing Cultural Influencers and Scholars Anti-Temer   Follow Back Leftists   PT Supporters
Other	International   Pop Music Stars and Pop Culture Sport Journalists International Pop Music Stars   Pop Culture Cultural Left Turkish Social Media Influencers Love, Beauty and Travelling   Social Media Rio Institutions Influencers and Celebrities Latin American   Left-wing journalists Feminist Activists Brazilian TV Celebrities Quotes, Memes, Pets and Travel   Social Media Internet Influencers   Memes, Quotes and Videos Venezuela   Pro-Chavez / Maduro Football Fans
Jair Bolsonaro	Anti-PIT and Anti-Lula   Nationalists Bolsonaro Supporters   Women Anti-Left   Conservatives Suspended   Hate Speech Mainstream Conservative News Media PSL Politicians and Supporting Celebrities Women With Bolsonaro Anti-PT Nationalists   Bot-Like Anti-Socialism   Anti- PT Cultural Right and Right-wing Writes Suspended   Far Right   Junk News Ordinary People   Anti-PT Bolsonaro Supporters   Right Follows Right Conservative Politicians and Right-Wing Journalists Right-wing Liberals   Anti-PT

## 2. Heterophily Index

For every pairing of groups within a network map, a value of heterophily can be calculated. This is a measure of the level of connection between the groups. In order to determine this a ratio is calculated between the actual ties between two groups, compared to the expected ties between the groups if all the accounts in the map were evenly distributed. The natural log of these ratios is then taken, along with a zero correction to create a balanced index and to ensure that all values are displayed in a positive form.

$$Ratio\ of\ Ratios_T = \frac{\frac{Connections_{pairing}}{\sum_{all\ pairings} Connections}}{Connections_{pairing}}}{\frac{Connections_{pairing}}{\sum_{all\ pairings} Connections}}$$

Expression A: Ratio of Two Ratios

This heterophily index is therefore created through a ratio of two ratios. This ratio reveals whether two nodes have about the proportion of links they should have given its size. This is displayed in Expression A, where a pairing of groups is calculated as having a measure of connections in balance with its share of all the connections.

Half the distribution of possible values from this ratio of ratios ranges from 0 to 1 (corresponding to disproportionately small share of connections in a group given its size) and the other half ranges from 1 to +infinity (a disproportionately large share of connections in a group given its size). However, by taking the natural log of the ratio of ratios the index will become more balanced: from -infinity to 0 becomes less than proportionate share, and from 0 to +infinity becomes more than proportionate share. For example, take a three-group network (A, B and C). If nodes in group A have a total of ten connections, and there are ten nodes in each group, then the expected connections between A and B will be 3.33. If, in reality, the nodes in group A actually have all ten connections to nodes in group B then this connection is stronger than expected. The heterophily score for groups A and B =  $10/3.33 = 3.0$ . The natural log of this is then taken along with a zero correction across the range of heterophily values.

A greater heterophily index indicates a denser pattern of connections between the two groups. It is important to note however that these scores indicate only first order connections, not second or third order connections. The pairwise Heterophily scores for the Twitter groups are given in Table 2.

**Table 2: Heterophily Indexes for Junk News Audiences Groups on Twitter**

Groups	Ciro Gomes	Geraldo Alckmin	Jair Bolsonaro	Lula & Haddad	Other
Ciro Gomes	54.3	0.7	0.0	0.5	1.0
Geraldo Alckmin	0.0	4.0	1.2	0.4	0.6
Jair Bolsonaro	0.0	0.0	2.7	0.1	0.4
Lula & Haddad	0.0	0.0	0.0	2.3	0.9
Other	0.0	0.0	0.0	0.0	0.8

Source: Authors' calculations from data sampled 22/08/2018 - 21/09/2018

### 3. List of Hashtags and Twitter Handles Tracked

**Table 3: Hashtags Tracked for Twitter Data Collection on Brazilian Election**

Candidate	Candidate Hashtags	Candidate Handle	Party	Party Handle
Alvaro Dias	#alvarodias19, #alvarodias, #adcomunicacao, #alvarodiaspresidente, #abreoolhobrasil	@alvarodias_	PODE	@podemos_19
Cabo Daciolo	#cabodaciolo, #votedaciolo, #daciolopresidente2018, #daciolopresidente, #daciolo	@CaboDaciolo	PATRI	@partidoPEN51
Ciro Gomes	#ciropresidente, #ciro12, #ciropresidente12, #cirogomes, #spciro, #livrecomciro	@cirogomes	PDT	@PDT_Nacional
Geraldo Alckmin	#alckmin, #equipeGA, #equipePSDB, #psdb, #alckminpresidente, #geraldoalckmin	@geraldoalckmin	PSDB	@Rede45
Guilherme Boulos	#psol, #BoulosESonia50, #vamossemmedo, #psol2018, #equipeBoulos, #boulosesonia, #boulos, #guilhermeboulos	@GuilhermeBoulos	PSOL	@psol50

Henrique Meirelles	#chamaomeirelles, #chamaofichalimpa, #fichalimpa, #meirelles15, #meirellesgeek	@meirelles	MDB	@MDB_Nacional
Jair Bolsonaro	#bolsonaro2018, #estoucombolsonaro, #bolsonaropresidente, #euvotobolsonaro17, #bolsonaro17, #somostodosbolsonaro, #mulherescombolsonaro, #bolsonaropresidente17, #brasilcombolsonaro, #mudabrasilverdade	@jairbolsonaro	PSL	@PSL_Nacional
Lula da Silva	#lulalivre, #lulaécandidato, #obrasilfelizdenovo, #lulapresidente, #marchalulalivre, #lulalivrejá, #lulalivre, #freelula	@LulaOficial	PT	@ptbrasil
Marina Silva	#somosmarina, #marina2018, #marinapresidente, #marina18, #marinasilva, #marinasilva18, #votemarina18	@MarinaSilva	REDE	@REDE_18
Fernando Haddad	#lulamanuhaddad, #haddad, #lulahaddadmanu, #fernandohaddad, #lulamanuelahaddad, #lulahaddadmanuela, #fernandolulahaddad, #lulahaddadmanu13pt, #haddad13, #lulahaddad	@Haddad_Fernando	PT	@ptbrasil
General	#eleições2018, #eleições, #eleicoes, #eleicoes2018, #eleicoesbrasil			

#### 4. Clustering Algorithm for Determining Groups and Clusters

In order to generate clusters and groups for each map it is necessary to employ a clustering algorithm. This involves first building a bipartite graph between nodes in the map and the rest of the social medium in question. This bipartite graph provides a structural similarity metric between nodes in the map. This was then used in combination with a hierarchical agglomerative clustering algorithm in order to cluster a map into distinct communities. This is a ‘bottom up’ approach whereby each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. Twitter maps are clustered based on follower-relationships, since mention-relationships have been shown to overemphasize the news cycle and salient external events. Facebook networks are clustered based on page likes.

#### 5. K-core reduction

To identify and map the ‘discussion core’ of the most active, connected, and influential users, we performed a k-core reduction to reduce the total collected set of Twitter users from the initial data collection into a set of well-connected accounts. This produces a maximally connected subgraph of active nodes with degree of connection at least ‘k’. This degree of connection, k, can be thought of as the number of links between each node in the graph. For example, selecting a k value of 0 for the reduction does not remove any nodes from the graph, since each node must have 0 connections or greater. Selecting a k value of 1 would remove all of the nodes that have no connections to other nodes in the graph. Selecting a k value of 2 would remove all nodes with fewer than 2 connections, and so on.

#### 6. Junk News Classification

These sources deliberately publish misleading, deceptive or incorrect information purporting to be real news about politics, economics or culture. This content includes various forms of propaganda and ideologically extreme, hyper-partisan, or conspiratorial news and information. For a source to be labelled as junk news at least three of the following five characteristics must apply:

- Professionalism: These outlets do not employ the standards and best practices of professional journalism. They refrain from providing clear information about real authors, editors, publishers and owners. They lack transparency, accountability, and do not publish corrections on debunked information.
- Style: These outlets use emotionally driven language with emotive expressions, hyperbole, ad hominem attacks, misleading headlines, excessive capitalization, unsafe generalizations and fallacies, moving images, graphic pictures and mobilizing memes.
- Credibility: These outlets rely on false information and conspiracy theories, which they often employ strategically. They report without consulting multiple sources and do not employ fact-checking methods. Their sources are often untrustworthy and their standards of news production lack credibility.
- Bias: Reporting in these outlets is highly biased and ideologically skewed, which is otherwise described as hyper-partisan reporting. These outlets frequently present opinion and commentary essays as news.
- Counterfeit: These outlets mimic professional news media. They counterfeit fonts, branding and stylistic content strategies. Commentary and junk content is stylistically disguised as news, with references to

news agencies, and credible sources, and headlines written in a news tone, with bylines, date, time and location stamps.

Table 4 gives a list of the top 10 junk news sources used for this analysis.

---

---

**Table 4: List of Top 10 Junk News Sources**

---

<i>Domain name</i>	<i>Frequency</i>
diariodocentrodomundo.com.br	148
reaconaria.org	95
falandoverdades.com.br	47
blogdacidadania.com.br	46
tercalivre.com.br	43
folhacentrosul.com.br	36
clickpolitica.com.br	31
contextolivre.com.br	28
republicadecuritiba.net	25
apostagem.com.br	22

---

---

*Source: Authors' calculations from data sampled 19/08/18 — 28/08/18*

---

---