

Uma análise da rede dos bots no Twitter nas eleições presidenciais brasileiras de 2018

Gustavo Germano
Departamento de Ciência da
Computação
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil
gustavog@dcc.ufmg.br

Fabricio Murai
Departamento de Ciência da
Computação
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil
murai@dcc.ufmg.br

Marcus Abilio Pereira
Departamento de Ciência Política
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil
marcusabilio@fafich.ufmg.br

ABSTRACT

A utilização das redes sociais como ferramenta de propagação de ideias políticas tem sido um dos fatores mais marcantes nas últimas eleições pelo mundo. E, cada vez mais, a utilização de soluções automatizadas, como bots, está presente. Motivado pelas eleições presidenciais no Brasil de 2018, esse trabalho tenta entender melhor o *modus operandi* dos bots no Twitter. nesse contexto. Além dos bots propriamente ditos, analisamos toda a sua rede de interação/influência. Assim, esperamos compreender que tópicos são discutidos pelos bots e de que forma eles estão sendo propagados.

1 INTRODUÇÃO

Nos últimos anos, a utilização da internet alterou drasticamente as campanhas eleitorais, e as redes sociais vem sendo usadas extensivamente para promoção e divulgação de candidatos a eleições em todo mundo.

Desde as primeiras utilizações da internet como plataforma eleitoral, que remontam a 1992, passando pelo caso emblemático da eleição de Barack Obama nos Estados Unidos em 2008, e culminando nas eleições presidenciais estadunidenses de 2016, marcada pela ampla utilização desse ferramental pelo candidato vencedor, Donald Trump, o processo de campanha online vem se modificando a modo de se aproveitar das possibilidades e potencialidades específicas da internet [1]. E, com a evolução da própria internet, essas estratégias vem se alterando.

Nos anos 2004, quando Al Gore e George Bush Jr. disputaram a presidência dos Estados Unidos, a utilização e-mails para estabelecer contato com os eleitores e tentar criar engajamento para a campanha, com aumento do fluxo de mensagens enviadas em períodos importantes das eleições, foi então uma das principais estratégias utilizadas [19].

O panorama quando Barack Obama se elegeu pela primeira vez, em 2008, já havia se alterado. A campanha do então candidato se baseou principalmente na ampla presença nas redes sociais, que estavam em ampla ascensão nos Estados Unidos, além de um site com bastante conteúdo relacionado ao candidato e à campanha. A atuação em cada rede social era personalizada, de forma a se aproveitar das especificidades de cada uma delas, como *MySpace*, *Flickr* e *Youtube*. Além disso, estimam-se gastos mais de quatro vezes maiores em publicidade on-line em comparação ao concorrente republicano [8].

Além de ajudar a promover os candidatos, as redes sociais começaram a ser utilizadas como vetor de propagação de rumores e notícias falsas já nas eleições presidenciais estadunidenses de 2012 [17].

Paralelamente, com o surgimento de várias ferramentas que facilitaram extração de grande quantidade de dados, que passaram a possibilitar abordagens automatizadas com o uso desses dados [18].

Mais recentemente, a utilização de bots para automatizar esse processo de promoção/divulgação, para gerar popularidade e engajamento para os candidatos ou para promover suas ideias se tornou mais frequente. Em 2016, na eleição presidencial estadunidense, bots no Twitter, presumidamente de origem russa, propagaram massivamente mensagens de apoio ao agora presidente Donald Trump e podem ter influenciado diretamente o resultado da eleição [3].

Esse cenário não está restrito aos Estados Unidos. Um estudo de 2018 [9] mostra que bots no Twitter tiveram uma influência considerável no referendo do Brexit, promovendo o voto “sim”. Um cenário similar se repetiu nas eleições presidenciais francesas de 2017, com um estudo sugerindo a existência de um mercado negro de bots para gerar desinformação [6]. Já no Brasil, um estudo realizado pela Universidade de Oxford mostra que desde 2014 os bots exercem um papel importante na política nacional, com alguns partidos gastando dezenas de milhões de reais em sua utilização [2].

Motivados por todo esse movimento, enxergamos a necessidade de analisar a utilização e o *modus operandi* dos bots no twitter, no contexto das eleições presidenciais brasileiras de 2018.

O Twitter, uma das mais famosas e populares redes sociais, tem sido central nas campanhas políticas por todo mundo, e sua estrutura permite a extração de dados de forma razoavelmente simples. Por isso, o Twitter foi o foco desse trabalho.

Monitoramos, durante o período eleitoral (7 de julho a 28 de outubro), as hashtags mais populares que eram relacionadas diretamente com as eleições presidenciais brasileiras. Os principais temas das hashtags coletadas foram apoio ou rejeição explícita a um candidato e debates realizados na televisão. Uma vez escolhidas as hashtags, coletamos usuários que as utilizavam e os classificamos entre bots ou não-bots, e dessa forma obtemos uma lista de usuários presumidamente bots que tweetavam sobre política.

Utilizando esses usuários presumidamente bots que tweetavam com hashtags relacionadas à política, construímos uma rede, onde eles eram os nós iniciais, e utilizamos a técnica Frontier Sampling [15] para realizar um processo de amostragem na rede. Dos usuários obtidos pelo processo de amostragem, coletamos o texto e a data de criação de seus tweets.

Utilizaremos na análise apenas aqueles tweets que foram realizados durante o período eleitoral. Esses foram agrupados em grupos, conforme os tópicos presentes em seus tweets. Através disso, realizamos uma análise utilizando ferramentas de análise de discurso. Através disso, esperamos:

- Entender melhor a composição da rede dos bots que tweetam sobre política nacional no Twitter.
- Descobrir quais tópicos são predominantemente discutidos nessa rede.
- De que forma esses tópicos são abordados.

Esse trabalho foi estruturado da seguinte forma: A Seção 3 descreve o ferramental utilizado para a coleta e análise de dados. A Seção 4 detalha o processo de obtenção dos dados, bem como as técnicas utilizadas. Nas Seções 5 e 6, discutimos os resultados.

2 TRABALHOS RELACIONADOS

Existem vários estudos com foco no impacto e influência dos bots no Twitter no contexto estadunidense. Citando alguns, temos uma análise da atividade dos bots no primeiro debate presidencial nas eleições de 2016 [12] e no período eleitoral como um todo [11]. Uma análise do tipo de conteúdo compartilhado via Twitter no estado de Michigan [10], também no contexto eleitoral de 2016, também aborda a influência dos bots e fake news nas eleições estadunidenses.

No contexto brasileiro, temos discussões sobre as interferências ilegítimas e o os riscos da utilização dos bots nas eleições de 2018 [16].

Fora do contexto eleitoral, existe um estudo que analisa estratégias para tornar bots relevantes e influentes no Twitter [13].

3 REFERENCIAL TEÓRICO

O Twitter, como uma rede social, propicia facilmente uma modelagem em forma de rede/gráfo, onde os vértices representam usuários e uma aresta de um usuário A para um usuário B indica que A “segue” B. Através desta abstração, utilizaremos de técnicas de Redes Complexas e amostragem para modelar e obter estimar estatísticas globais da rede de bots do Twitter.

Para a obtenção dos dados, utilizamos a API do Twitter, que é disponibilizada pelo mesmo. A API do Twitter, que agora faz parte da Twitter Developer Platform, fornece todo um ferramental para obtenção de dados da rede social. Através da sua utilização, podemos obter as mais diversas informações sobre os usuários e seus tweets.

Para a amostragem da rede, utilizamos o algoritmo Frontier Sampling, cujo pseudocódigo está reproduzido no Algoritmo 1, que consiste em realizar múltiplos caminhamentos aleatórios em um grafo.

Algorithm 1 Frontier Sampling

```

1:  $n \leftarrow 0$  { $n$  é o número de passos}
2: Inicializar  $L = (v_1, \dots, v_m)$  com  $m$  vértices
3: do
4:   Selecione  $u \in L$ , com probabilidade  $\deg(u)/\sum_{v \in L} \deg(v)$ 
5:   Selecione uma aresta de saída de  $u$ ,  $(u, v)$ , uniformemente
6:   Substitua  $u$  por  $v$  em  $L$ 
7:   Adicione  $(u, v)$  a sequência de arestas amostradas
8:    $n \leftarrow n + 1$ 
9: while  $n \leq B - mc$ 

```

A única mudança em relação ao algoritmo original foi utilizar um conjunto escolhido manualmente como o conjunto inicial, ao invés

de escolher os membros desse conjunto aleatoriamente de forma uniforme, como é proposto no artigo original. Murai et al. [14] mostra que esta modificação não impacta significativamente os resultados quando o número de passos realizados por cada passeio aleatório é grande.

Para análise dos tópicos discutidos, utilizamos a técnica de Latent Dirichlet allocation (LDA) [4], um modelo estatístico que permite a extração dos tópicos presentes num texto. Para analisar o teor do discurso, utilizamos também a API HateSonar [5], que permite identificar a presença de discurso de ódio ou agressivo em um texto, e a biblioteca Vader, uma ferramenta para análise de sentimentos [7].

4 METODOLOGIA

Nesta seção descrevemos como os dados utilizados foram coletados (Seção 4.1), como serão analisados os tópicos (Seção 4.2) e o teor do discurso (Seção 4.3).

4.1 Coleta de dados

A coleta de dados se iniciou com a coleta de hashtags populares que estavam sendo utilizadas no período eleitoral e que eram relacionadas as eleições presidenciais. Foram escolhidas 16 hashtags, listadas abaixo:

Table 1: Hashtags

Hashtags
#Bolsonaro17, #bolsonaroCagao, #BolsonaroNao
#Eleições2018, #BolsonaroSim, #EleNao
#EleSim, #ViraViraCIR0, #DebateNaBand
#DebateSBT #ForaPT, #DebateNaGlobo
#HaddadÉLula, #HaddadSim, #DebateNaRecord
#VemProDebate.

O próximo passo foi identificar bots que estavam tweetando utilizando essas hashtags. Para isso, utilizamos a ferramenta *Botometer*, disponibilizada gratuitamente e desenvolvida na Universidade de Indiana [5]. Uma vez classificados, foram escolhidos 100 bots que serviram como nós iniciais para a rede. Nessa rede, os vértices representam os usuários (sendo o usuário inicial/central na construção da rede um dos bots), e as arestas representam as relações de seguidores/seguindo.

Com os bots definindo os nós iniciais da rede, utilizamos o Frontier Sampling, com o intuito de realizar uma amostragem, já que percorrer cada um dos usuários seria computacionalmente inviável.

Através desse, obtemos cerca de 2000 usuários, dentre bots e não-bots, que estão relacionados com o grupo de bots iniciais, obtidos via hashtags. Desses usuários coletamos dados relacionados ao perfil, como número de seguidores, e os Tweets. Com esses dados, inicia-se a próxima fase.

4.2 Análise de tópicos

Com os tweets de todos os usuários obtidos através do processo de amostragem coletado, utilizamos o LDA para extrair os tópicos de toda a massa coletada. Extraímos também, da mesma forma, os tópicos de cada usuário individualmente. Lembrando que apenas os

tweets realizados durante o período eleitoral foram considerados nessa análise.

Dessa forma, cada usuário pode ser representado através das palavras que descrevem os tópicos que ele discute. A configuração utilizada no LDA foi 10 tópicos com 15 top-words. Essa representação foi utilizada para como entrada para o algoritmo de clusterização. A técnica utilizada foi o k-means com PCA para redução de dimensionalidade. Assim, obtivemos clusters e os tópicos que eles discutem.

4.3 Análise de discurso

Utilizamos a API HateSonar e a biblioteca Vader para realizar a análise dos tweets de cada usuário. As duas ferramentas dão uma nota para o texto em três categorias. O HateSonar possui as seguintes categorias: discurso de ódio, linguagem ofensiva e neutro. Já o Vader classifica o texto entre positivo, negativo e neutro. Dessa forma, cada usuário foi descrito por três valores por cada uma dessas ferramentas.

5 RESULTADOS

Nesta seção apresentamos resultados em relação à caracterização dos usuários, dos tópicos e o teor de discurso presente nos Tweets.

5.1 Dados gerais dos usuários

Primeiramente, analisamos os dados relacionados aos usuários e a seus tweets. Sobre o usuário, foram considerados: número de seguidores, número de amigos (usuários que o seguem), número de tweets e número de retweets. Sobre os tweets, consideramos: número de likes recebidos, número de retweets recebidos e número de urls tweetadas. Comparando essas métricas entre os usuários bots e não-bots, obtivemos os resultados mostrados nas Figuras 1 a 4.

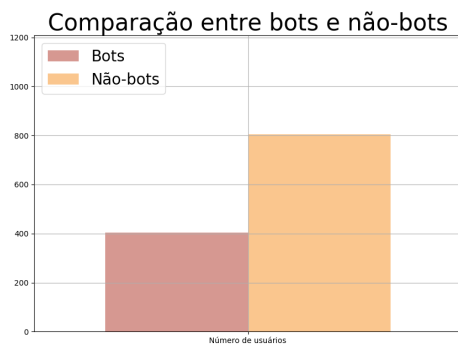


Figure 1: Comparação entre o número de bots e não-bots

Comparação entre bots e não-bots (em média)

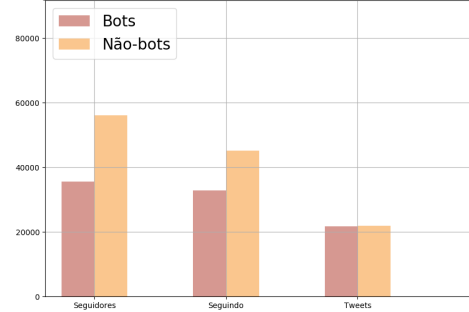


Figure 2: Comparação do número de seguidores, seguindo e tweets entre bots e não-bots

Comparação entre bots e não-bots (em média)

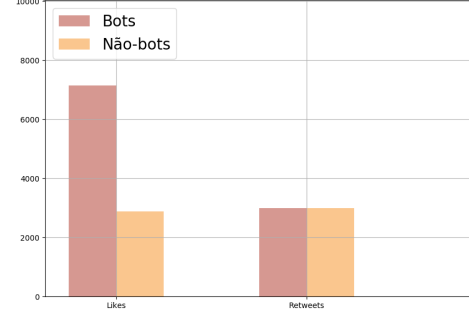


Figure 3: Comparação do número de likes e retweets entre bots e não-bots

Comparação entre bots e não-bots (em média)

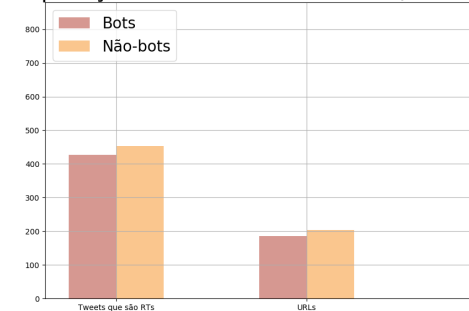


Figure 4: Comparação do número de tweets que são retweets e urls utilizadas entre bots e não-bots

5.2 Tópicos e clusterização

Table 2: Tópicos em todo texto

Tópico	Top words de cada tópico
Topic 0	pop bola tambem vai nao segundo historia agora bolsonaro brasil pra video voce sempre sobre
Topic 1	nao brasil bolsonaro todos ser ano povo presidente ha pt sao ja hoje voto bem
Topic 2	video vida brasil dia todos aqui amigo voce na toda grande nao amor agora no
Topic 3	pra nao ta vc voce dia vou vai pessoa ser ja pessoas gente pq vida
Topic 4	rt bolsonaro nao vai pt brasil ser sobre todos lula presidente contra ja agora pra
Topic 5	nao bolsonaro pra pt vai brasil haddad ja vamos ser dia voce todos agora pais
Topic 6	bola pop segundo tambem vai historia video agora voce sao sempre la estao dia ai
Topic 7	nao pra voce vai ser ta ja dia ano tudo ter gente ate sao melhor
Topic 8	bolsonaro nao brasil contra vai lula sobre diz haddad sao pt paulo ser presidente governo
Topic 9	eh nao pra pq ta ser vai sobre ne ja ano vcs sao ter gente

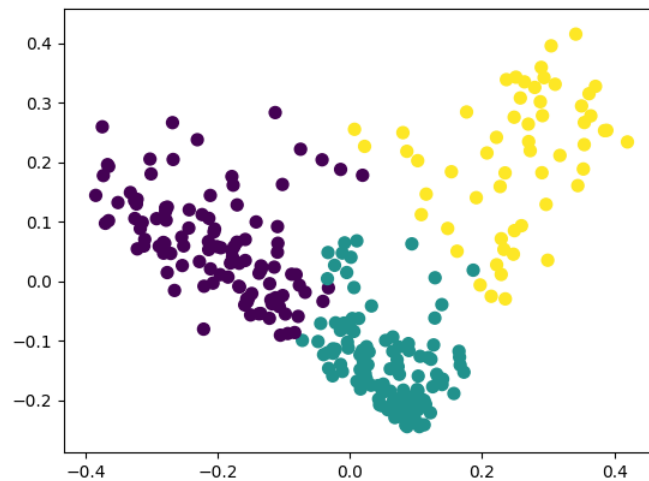


Figure 5: Grupos obtidos com a clusterização

Table 3: Tópicos por cluster

Cluster	Tópicos
Cluster roxo	eu nao vida pra ficar alguem ja pessoa ter
Cluster amarelo	link nao pra voce vai ser dia brasil ta
Cluster ciano	bolsonaro pt presidente candidato contra pais haddad campanha povo

5.3 Teor do discurso

As Figuras 6 e 7 mostram os resultados obtidos pelo Hatesonar e Vader, respectivamente.

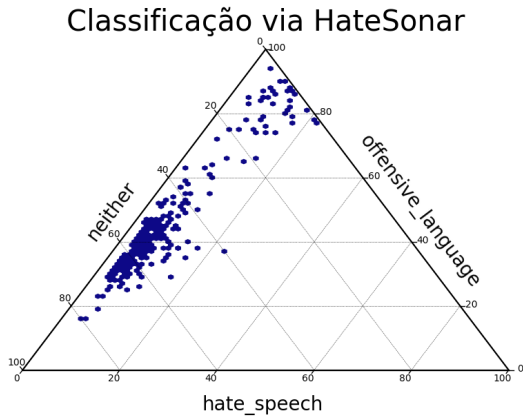


Figure 6: Resultados obtidos com o HateSonar

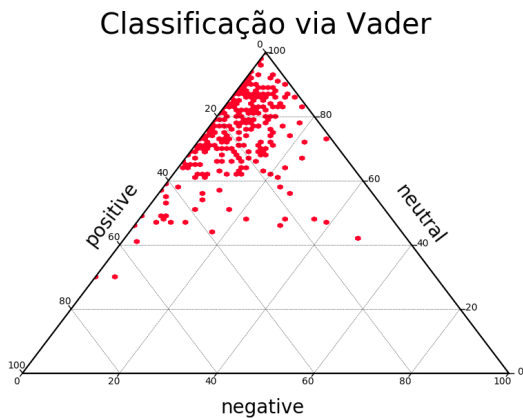


Figure 7: Resultados obtidos com o Vader

6 DISCUSSÃO DOS RESULTADOS

Analisando os resultados obtidos, vemos que uma parcela considerável dos usuários amostrados da rede dos bots também é bot (cerca de 33%). Vemos também que, embora bots e não-bots tenham número de tweets similares, em média, os usuários não-bots possuem mais seguidores e pessoas os seguindo. Vemos também que os bots, em média, receberam um número consideravelmente maior de likes em seus tweets, e um número equivalente de retweets em comparação aos usuários não-bots.

Nas demais métricas utilizadas, baseadas em metadados sobre os usuários e sobre os tweets, os usuários bots e não-bots obtiveram resultados iguais ou bastante similares.

Através da análise dos tópicos, vemos que, na análise geral, 4 dos 10 tópicos obtidos são exclusivamente sobre as eleições presidenciais de 2018. Palavras relacionadas a temática também apareceram em outros tópicos. Dentro da temática das eleições, não é possível separar totalmente os tópicos discutidos, com palavras como "bolsonaro", "brasil", "lula" e "haddad" aparecendo com frequência bastante similar.

Analisando por clusters, obtivemos um cluster (o cluster ciano) que aborda exclusivamente as eleições presidenciais. Novamente, palavras que se referem aos lados antagônicos no segundo turno das eleições aparecem lado a lado, não sendo possível assim verificar se o cluster trata de uma temática específica dentro do contexto geral das eleições. A predominância da temática das eleições foi similar entre a abordagem de análise de tópicos em todo os tweets e por clusters (40% na primeira, 33% na segunda).

Através da análise de discurso, vemos que o teor do discurso propagado na rede é predominantemente neutro. O HateSonar, porém, detectou uma parcela de usuários com um score alto de linguagem ofensiva. Analisando os dados obtidos com o Vader, também vemos alguns usuários com discurso positivo. Mesmo com esses grupos, a classe predominante do discurso, em ambos os algoritmos, foi uma classe de neutralidade.

Assim, vemos que, em sua maioria, a rede dos bots que discutiram as eleições presidenciais de 2018 é composta, em sua maioria, por usuários não-bots, com discurso neutro e que não abordam, majoritariamente, temáticas relacionadas a política em seus perfis.

7 TRABALHOS FUTUROS

Alguns pontos interessantes foram levantados nesse trabalho e podem ser desenvolvidos futuramente. Analisando o número de bots na rede, vemos que o valor é bastante similar com a parcela da rede que discute política. Assim, seria interessante analisar se a temática política é majoritariamente discutida pelos bots ou pelos não-bots.

Com uma análise num contexto político geral, não apenas no contexto das eleições, talvez seja possível identificar as sub-temáticas dentro da temática política que são predominantes na rede dos bots.

REFERENCES

- [1] Camilo Aggio. 2010. Campanhas Online: O percurso de formação das questões, problemas e configurações a partir da literatura produzida entre 1992 e 2009. *Opinião Pública* 16, 2 (2010), 426–445.
- [2] Dan Arnaudo. 2017. Computational propaganda in Brazil: Social bots during elections. *Computational Propaganda Research Project, Working Paper* 2017.8 (2017).
- [3] Alessandro Bessi and Emilio Ferrara. 2016. Social bots distort the 2016 US Presidential election online discussion. (2016).
- [4] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.
- [5] Thomas Davidson, Dana Warmley, Michael Macy, and Ingmar Weber. 2017. Automated hate speech detection and the problem of offensive language. *arXiv preprint arXiv:1703.04009* (2017).
- [6] Emilio Ferrara. 2017. Disinformation and social bot operations in the run up to the 2017 French presidential election. (2017).
- [7] CJ Hutto Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Available at (20/04/16) <http://comp.social.gatech.edu/papers/icwsml4.vader.hutto.pdf>.
- [8] Wilson Gomes, Breno Fernandes, Lucas Reis, and Tarcizio Silva. 2009. "Politics 2.0": Barack Obama's on-line 2008 campaign. *Revista de Sociologia e Política* 17, 34 (2009), 29–43.
- [9] Yuriy Gorodnichenko, Tho Pham, and Oleksandr Talavera. 2018. *Social media, sentiment and public opinions: Evidence from# Brexit and# USElection*. Technical

- Report. National Bureau of Economic Research.
- [10] Philip N Howard, Gillian Bolsover, Bence Kollanyi, Samantha Bradshaw, and Lisa-Maria Neudert. 2017. Junk news and bots during the US election: What were Michigan voters sharing over Twitter. *Computational Propaganda Research Project, Oxford Internet Institute, Data Memo 2017.1* (2017).
 - [11] Philip N Howard, Bence Kollanyi, and Samuel Woolley. 2016. Bots and Automation over Twitter during the US Election. *Computational Propaganda Project: Working Paper Series* (2016).
 - [12] Bence Kollanyi, Philip N Howard, and Samuel C Woolley. 2016. Bots and automation over Twitter during the first US presidential debate. *Comprop data memo 1* (2016), 1–4.
 - [13] Johnnatan Messias, Lucas Schmidt, Ricardo Augusto Rabelo de Oliveira, and Fabricio Rodrigues Benevenuto. 2013. You followed my bot! Transforming robots into influential users in Twitter. (2013).
 - [14] Fabricio Murai, Bruno F. Ribeiro, Don Towsley, and Pinghui Wang. 2018. Characterizing Directed and Undirected Networks via Multidimensional Walks with Jumps. *TKDD (to appear)* (2018).
 - [15] Bruno Ribeiro and Don Towsley. 2010. Estimating and sampling graphs with multidimensional random walks. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. ACM, 390–403.
 - [16] Marco Aurélio Ruediger, Amaro Grassi, Ana Freitas, Andressa Contarato, Carolina Taboada, Danilo Carvalho, Humberto Ferreira, Lucas Roberto da Silva, Pedro Lenhard, Rachel Bastos, et al. 2017. Robôs, redes sociais e política no Brasil: estudo sobre interferências ilegítimas no debate público na web, riscos à democracia e processo eleitoral de 2018. (2017).
 - [17] Jieun Shin, Lian Jian, Kevin Driscoll, and François Bar. 2017. Political rumoring on Twitter during the 2012 US presidential election: Rumor diffusion and correction. *new media & society* 19, 8 (2017), 1214–1235.
 - [18] Tarcizio Silva and Max Stabile. 2016. Monitoramento e Pesquisa em Mídias Sociais–Metodologias, aplicações e inovações. *Análise de redes em Mídias Sociais. São Paulo: Uva Limão* (2016), 237–260.
 - [19] Andrew Paul Williams and Kaye D Trammell. 2005. Candidate campaign e-mail messages in the presidential election 2004. *American Behavioral Scientist* 49, 4 (2005), 560–574.