

FUCAPE FUNDAÇÃO DE PESQUISA E ENSINO

ZUIL ANTONIO PIROLA FILHO

@JAIRBOLSONARO: A ARMA DE ATAQUE DO PRESIDENTE

**VITÓRIA
2019**

ZUIL ANTONIO PIROLA FILHO

@JAIRBOLSONARO: A ARMA DE ATAQUE DO PRESIDENTE

Dissertação apresentada ao Programa de Pós-Graduação em Administração, Fucape Fundação de Pesquisa e Ensino, como requisito parcial para obtenção do título de Mestre em Administração.

Orientador: Prof. Dr. Daniel Modenesi de Andrade.

**VITÓRIA
2019**

ZUIL ANTONIO PIROLA FILHO

@JAIRBOLSONARO: A ARMA DE ATAQUE DO PRESIDENTE

Dissertação apresentada ao Programa de Mestrado em Administração da Fucape Fundação de Pesquisa e Ensino, como requisito parcial para obtenção do título de Mestre em Administração.

Aprovada em 12 de dezembro de 2019.

COMISSÃO EXAMINADORA

Prof. Dr. DANIEL MODENESI DE ANDRADE
Fucape Fundação de Pesquisa e Ensino

Prof. Dr. FABIO YOSHIO SUGURI MOTOKI
Fucape Fundação de Pesquisa e Ensino

Prof. Dr. PATRICK MARQUES CIARELLI
UFES – Universidade Federal do Espírito Santo

AGRADECIMENTOS

Aos meus pais Eli e Zuil, pelo apoio em todos os âmbitos. Por diversas vezes me vi desistindo deste trabalho e nunca seguiria em frente sem estas pessoas motivadoras e parceiras.

A minha esposa Amanda, todo carinho, conforto e abrigo nos momentos de fadiga. Obrigado por acreditar no meu potencial.

Em especial ao meu orientador Daniel Modenesi por aceitar orientar um aluno quase desistente, por acreditar no meu trabalho e por me incentivar a buscar o conhecimento nos lugares corretos.

.

You shall know a word by the company it keeps.
(J. R. Firth 1957:11)

RESUMO

O objetivo do presente trabalho é aplicar uma rotina em *latent Dirichlet allocation* (LDA), para modelar tópicos e associá-los a repercussão (engajamento) das postagens no Twitter do atual presidente do Brasil, Jair Bolsonaro (JB), compreendendo o período de 2014 a 2018. Inicialmente foi utilizado LDA para a modelagem dos tópicos e então regressões para tentar explicar o número de curtidas a partir dos tópicos inferidos: apoio, ataque e campanha. O estudo indica uma relação positiva entre mensagens de ataque e o número de curtidas. O estudo também indica que o processamento de Linguagem Natural (NLP) e *machine learning* podem contribuir para análise de dados de redes sociais em processos eleitorais.

Palavras-chave: Processamento de Linguagem natural; *Machine learning*; LDA; Twitter, Bolsonaro, eleições.

ABSTRACT

The aim of this paper is to apply a Dirichlet Latent Allocation (LDA) routine, to model topics and associate the repercussion (engagement) of the Twitter posts of the current president of Brazil, Jair Bolsonaro (JB), covering the period from 2014 to 2018. Initially, LDA was first used for topic modeling and then regressions to try to explain the number of likes from inferred topics: support messages, attack messages, and campaign messages. The study indicates a positive relationship between attack messages and number of likes. The study also indicates that Natural Language Processing (NLP) and machine learning can contribute to the analysis of social network data in electoral processes.

Keywords: NLP, *Machine learning*, LDA, Twitter, Bolsonaro, election.

LISTA DE TABELAS

Tabela 1 - Gastos e tempo de campanha	18
Tabela 2 - Eventos de campanha.....	18
Tabela 3 - Tópicos lda.....	25
Tabela 4 - Estatística descritiva.....	29
Tabela 5 - Modelo empírico - Likes x Topic_0 x Topic _1	30
Tabela 6 - Matriz de correlação palavras Topic_1.....	34
Tabela 7 - Total de ocorrências Tópico 1	35
Tabela 8 - Modelo empírico - RTS x Topic_1	35

LISTA DE FIGURAS

Figura 1: Evolução do número de seguidores @jairbolsonaro	19
Figura 2 - Publicações de tweets JB e FH	19
Figura 3 - Exemplo de artigo com a demonstração dos tópicos.....	22
Figura 4: Tweet mais curtido tópico 0.....	26
Figura 5: Tweet mais curtido tópico 1.....	26
Figura 6: Tweet mais curtido tópico 2.....	27
Figura 7: Análise horizontal - número de tweets por tópico.....	32
Figura 8 - Representação gráfica Word2Vec	38

LISTA DE SIGLAS

LDA – Latent Dirichlet Allocation

JB – Jair Bolsonaro

FH – Fernando Haddad

ICT – Information and communication technology

RTS – Retweets

API – Application Programming Interface

MPDFT – Ministério Público do Distrito Federal e Territórios

PT – Partido dos trabalhadores

CBOW – Continuous Bag-of-Words

SUMÁRIO

Capítulo 1	11
1.INTRODUÇÃO	11
Capítulo 2	14
2.REFERENCIAL TEÓRICO	14
2.1.E-CAMPAIGNING	14
2.2.CONTEXTO DA CAMPANHA PRESIDENCIAL 2018	17
2.3.FREE MEDIA COVERAGE	20
Capítulo 3	22
3.METODOLOGIA	22
Capítulo 4	28
4.RESULTADOS	28
4.1.TÓPICOS LDA	28
4.2.REGRESSÕES	29
4.3.ABRANGÊNCIA DO TÓPICO 1	33
4.4.WORD2VEC	36
Capítulo 5	39
5.CONCLUSÃO	39
REFERÊNCIAS	42

Capítulo 1

1. INTRODUÇÃO

O objetivo do presente trabalho é aplicar uma rotina em *latent Dirichlet allocation* (LDA), para modelar tópicos e associá-los a repercussão (engajamento) das postagens no Twitter do atual presidente do Brasil, Jair Bolsonaro (JB), compreendendo o período de 2014 a 2018. Adicionalmente, foi aplicado o método word2vec para identificar as palavras mais utilizadas e a relações entre seus vetores.

A literatura sugere que a política no Twitter não é um substituto para campanhas tradicionais fortes (Stateman, 2014). Porém, diversos atores têm observado na internet um método de comunicação eficaz durante o processo eleitoral (Perestre-Mendizábal, 2018). O *e-campaigning* e a utilização de tecnologia de comunicação em informação (*Information and communication technology* - ICT) podem mobilizar e influenciar indivíduos (Lilleker & Vedel, 2013). Uma palavra individual ou uma *hash-tag* podem ter mais peso retórico que uma palavra em um discurso ou em um artigo convencional (Basilakis, Diermeier, & Goecke, 2018).

Para Paulo (2011), as tecnologias digitais de comunicação no uso político têm a capacidade variável de contribuir para resolver problemas de comunicação, isso significa que estas contribuições não são uma mera peculiaridade. Sendo assim, resta reconhecer que as dimensões do jogo democrático são refratárias a modificações bruscas de tecnologia, sendo sujeitas muito mais a uma cultura política do que um simples avanço tecnológico ou uma ferramenta.

Se compararmos o ciberespaço com as mídias convencionais, podemos observar que além da velocidade de circulação de dados nos meios digitais, este meio

permite que o usuário navegue e recupere dados uma vez postados. Isto faz que a informação perca sua característica efêmera muito observada no meio televisivo, por exemplo (Valente & Silva, 2010).

Este estudo é motivado pela noção de que não se pode desprezar os conteúdos digitais em um processo eleitoral. Assim, ao tratar da classificação de conteúdo de redes sociais de forma efetiva, Wang (2016) aplicou *Latent Dirichlet Allocation* (LDA) para extrair tópicos das postagens do Twitter, do então presidente americano Donald Trump. Os resultados indicam que, dentre os tópicos encontrados, os que envolvem ataque aos democratas foram os mais curtidos pelos seguidores.

A modelagem por tópicos não abrange somente temas envolvendo política, há uma variedade de estudos envolvendo a área. No Brasil, o LDA também tem sido utilizado para modelar tópicos em corpus jurídicos e textos como o tema agropecuária (Basilio & Pereira, 2018; Drury, Fernandes, & Lopes 2017). Seu propósito é habilitar um modo eficiente de processamento de uma grande quantidade de dados preservando essencialmente a relação estatística que é utilizada em classificações (Blei, Ng, & Edu, 2003).

O segundo método utilizado neste trabalho tem como finalidade avaliar a correlação entre as palavras a partir da diferença entre os valores de seus vetores. A partir da utilização do modelo Word2vec, proposto por Mikolov (2013), uma representação vetorial pode capturar tanto relações sintáticas quanto semânticas das palavras. No Brasil, Lopes (2015) aplicou o word2vec para a língua portuguesa e o método se tornou eficaz também para identificar sintática e semanticamente palavras do português brasileiro.

Com a representação vetorial é possível plotar as palavras em um gráfico e mensurar a distância destas palavras, sendo palavras mais próximas mais ligadas

dentro do contexto da amostra e palavras mais distantes mais longe dentro do contexto da amostra.

Para atender aos objetivos desse estudo, foram utilizados dados extraídos via API do Twitter do perfil @jairbolsonaro. Na extração foram coletados 4001 tweets, seus links, o número de curtidas (likes) e retweets (RTS) de cada um deles.

O presente trabalho contribui com a literatura de modelagem de tópicos sendo utilizada como ferramenta efetiva de análise de dados em redes sociais. A ferramenta pode ser utilizada tanto em um processo eleitoral para monitorar um candidato, como para avaliar a atuação de uma empresa na rede medindo a efetividade de seu desempenho. A representação gráfica do word2vec gera associações que representam o comportamento do candidato durante o processo eleitoral e expressam visualmente o comportamento na emissão das mensagens.

Os resultados encontrados apontam para o LDA como método efetivo para descoberta de tópicos. Tem-se observado que o tópico de ataque também foi o mais estatisticamente relacionado com as curtidas, mesmo com as dificuldades da implantação do LDA para o idioma em português. Para a base de dados, a utilização de likes para medir preferência de tópicos foi mais efetiva que o RTS.

Capítulo 2

2. REFERENCIAL TEÓRICO

2.1. *E-CAMPAIGNING*

A ideia de utilizar internet para impulsionar campanha não é nova. Desde 2010, na campanha do ex-presidente americano Barack Obama, a importância das redes sociais no âmbito eleitoral tem crescido sistematicamente. Políticos têm aproveitado os benefícios da internet para divulgação de material de campanha e comunicação direta com os eleitores (Azmi, Sylvia, & Mardhiah, 2018).

No Brasil, as campanhas eleitorais também vem se caracterizando pelo uso massivo da internet e das mídias sociais como ferramenta de propaganda para os candidatos, sendo observada a consolidação de tal tendência nas Eleições 2014 e 2016 (Braga & Carlomagno, 2018). Em 2012, Cremonese estudou a utilização do Twitter pelos candidatos que concorriam as eleições presidências do Brasil de 2010. O autor descreveu o Twitter como ferramenta favorável a produção de capital social e civismo e a ferramenta foi empregada principalmente para a divulgação de agendas, tentando aproximar os eleitores de seus candidatos. O Twitter pode ser utilizado por políticos para informar sobre suas atividades pessoais ou político-partidárias, anúncio de políticas públicas, expressar opinião sobre questões atuais, agradecimentos e felicitações (feriados nacionais, por exemplo) e encorajar engajamento civil (Azmi, Sylvia, & Mardhiah, 2018)

O Twitter, foco deste estudo (também considerado um microblogging), é uma rede que permite que usuários transmitam seus status ou atualizações (tweets) através de 280 caracteres. A limitação de caracteres impede uma comunicação

sofisticada ou completa (Ott, 2017). Por outro lado, palavras individuais ou hashtags têm mais peso retórico que um discurso convencional (Bassilakis, Diermeier, & Goecke, 2018).

Os tweets são exibidos no perfil do usuário emissor e são também enviados para todos os usuários que o seguem (*followers*), sendo este o maior potencial da rede em ambiente político. O candidato é seguido por eleitores (ou potenciais eleitores), assim, quando uma mensagem é elaborada e enviada pelo candidato, todos seus seguidores recebem. Seus seguidores podem inclusive encaminhar (*Retweet*) as mensagens para seus próprios seguidores, tornando assim uma poderosa rede de disseminação, podendo funcionar como um *Bypass* aos meios de comunicação convencional (Bassilakis, Diermeier, & Goecke, 2018; Schoeder, 2018).

Sobre este potencial, Paulo e Almeida (2011) alertam para a utilização das redes sociais no sentido de aderir a listas e grupos de discussão específicas onde sempre seus próprios colegas comentam e divulgam, como o Twitter por exemplo. Para o autor, isto pode reforçar e reproduzir comportamentos típicos. Estas “bolhas” tendem a isolar atores dentro de grupos gerando uma falsa impressão de que o que todos falam e o que circula ali é opinião pública, silenciando o contraditório (Recuero, Zago, & Soares, 2017). Como contraponto, para Tufekci (2015) o Twitter pode ser uma boa escolha de estudo, pois tem menos algoritmos de filtragem do que o Facebook, por exemplo.

Ott, em 2017, propõe que o Twitter pode ser definido em 3 características básicas: simplicidade, impulsividade e incivilidade. Em seu trabalho, o autor critica o Twitter a um câncer social. Ele afirma que a rede “destrói o diálogo e a deliberação, promove a farsa e o fanatismo e contribui para a insensibilidade e o desprezo (p.60).”

Outra perspectiva seria pensar no Twitter capaz de produzir conteúdo informativo, opinativo, testemunhais e convocatórios (Malini & Antoun, 2013).

Hosch-Dayican (2014) constataram que pelo menos metade das mensagens postadas de cunho eleitoral no Twitter trabalham ou para persuadir seguidores a votar em um partido/candidato específico ou trabalham com campanha negativa.

As semelhanças entre as campanhas do atual presidente americano Donald Trump (@realDonaldTrump) e Bolsonaro são diversas. Depois de avaliar 2500 tweets do presidente americano, Ott (2017) descreve características observadas: Tweets com escrita simples e palavras monossilábicas, uma esmagadora quantidade de tweets com conotação negativa e com insultos e o uso frequente de exclamações e textos em caixa alta. Dentre as características descritas por Ott, inquestionavelmente, o aspecto ofensivo é semelhança entre os dois candidatos.

Para Cioccarri e Persischetti (2018), o então deputado federal Bolsonaro é um dos parlamentares mais atuantes nas redes, e utiliza-se desta para se aproximar dos eleitores alternando entre promessas, compromissos, leviandades e acusações.

O mesmo acontece com Trump, que emerge como uma rede de campanha negativa, como um vitríolo da mídia, incentivando inclusive o *New York Times* a publicar uma lista completa de pessoas insultadas pelo candidato no Twitter¹ (Gross & Johnson, 2016).

A ofensividade é inerente ao processo eleitoral, isto é, tanto pode ser considerada campanha negativa (Hosch-dayican et al., 2014), quanto um

¹ <https://www.nytimes.com/interactive/2016/01/28/upshot/donald-trump-twitter-insults.html>

comportamento de ataque (Wang et al., 2016). A incivilidade pode ser resumida como discurso indelicado, insultuoso ou em alguns casos ofensivo (Ott, 2017).

As curtidas (likes) podem ser interpretadas como atratividade dos seguidores e se algumas mensagens recebem mais likes que outras isto pode ser inferido como um fator de preferência (Wang et al., 2016). Smith (2015) discorda que likes possam ser uma boa métrica para engajamento. Seu principal argumento é que para que haja engajamento é necessário interatividade para gerar consumo de informação, senso de presença, imersão e conexão social.

A relevância do Twitter como uma plataforma de aproximação com o eleitorado - para o então eleito presidente da república, Jair Bolsonaro – é reforçada quando na fase de transição, Bolsonaro divulgou nomes de diversos ministros e presidentes de estatais (Lopes, 2018). Apesar da força da plataforma, Braga e Carlomagno (2018) apontam para um decréscimo do uso do Twitter e websites, e um crescimento do Facebook. Ainda que tenham observado um aumento médio de interações entre mídias sociais em diversas regiões do Brasil e uma associação positiva entre a presença online dos candidatos e o percentual de votos.

2.2. CONTEXTO DA CAMPANHA PRESIDENCIAL 2018

Volpatti e Lima (2019) defendem que Jair Bolsonaro (JB) foi eleito graças ao uso eficiente das redes sociais, propondo uma relação direta com seus eleitores por meio de mensagens do próprio candidato e *lives* (transmissões ao vivo frequentes). Isto porque diferentemente de outras eleições brasileiras, onde o tempo de TV e o total de verba de campanha eram determinantes para eleger presidentes, em 2018 diversos fatores no processo desbalancearam esta relação.

TABELA 1 - GASTOS E TEMPO DE CAMPANHA

CANDIDATO	VOTOS (1º Turno)	% VOTOS	TEMPO DE TV (1º Turno)	TOTAL DE DESPESAS CAMPANHA
JAIR BOLSONARO	49.277.010	46.03	8 s	BRL 2,456,215.03
FERNANDO HADDAD	31.342.051	29.28	2 min 23 s	BRL 37,503,104.50
CIRO GOMES	13.344.371	12.47	38 s	BRL 24,359,713.60
GERALDO ALCKMIN	5.096.350	4.76	5 min 32 s	BRL 53,350,139.97
JOAO AMOÊDO	2.679.745	2.5	5 s	BRL 2,884,581.03

Fonte: TSE (2019)

Dentre os 5 candidatos mais bem colocados, Jair Bolsonaro (JB) teve menos despesas de campanha e o segundo menor tempo de TV (ver Tabela 1).

A relação de dinheiro e tempo de TV foi subvertida por fatores muito peculiares de 2018 (Tabela 2): um dos candidatos impedido de concorrer pela Lei da Ficha Limpa, outro candidato sofrendo um atentado à vida (facada), possível impulsionamento ilegal de mensagens via Whatsapp, o fator Internet e suas múltiplas facetas como #elenão #elesim, dentre outros. Falando de internet, e mais especificamente do candidato vencedor, Jair Bolsonaro (JB), houve um uso efetivo e sistêmico dos meios e das redes para promoção².

TABELA 2 - EVENTOS DE CAMPANHA

DATA	EVENTO
22/07/2018	Oficialização
05/08/2018	Mourão
09/08/2018	Rede Bandeirantes
17/08/2018	Rede TV
28/08/2018	Jornal nacional
06/09/2018	Facada
29/09/2018	#Elenao
30/09/2018	#Elesim
07/10/2018	Primeiro turno
09/10/2018	Roger Water
10/10/2018	Whatsapp
28/10/2018	Segundo turno

Fonte: Elaborado pelo autor (2019)

² <http://familiabolsonaro.blogspot.com.br>

O crescimento do número de seguidores no Twitter do candidato (Figura 1) foi significativo no período que antecedeu o pleito, mais especificamente, no mês de outubro de 2018, mês no qual foram realizados o primeiro e o segundo turnos.

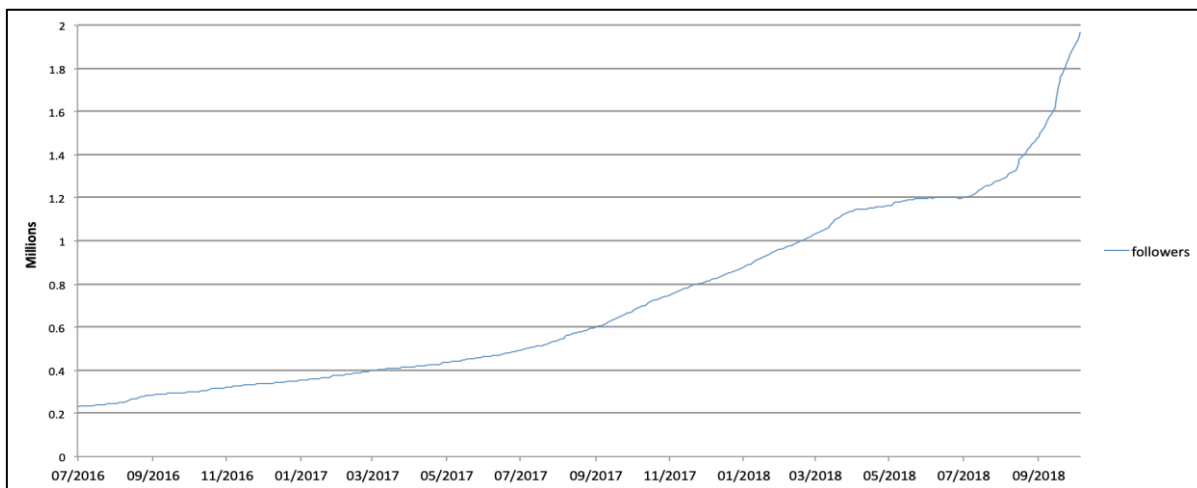


Figura 1: Evolução do número de seguidores @jairbolsonaro

Fonte: socialblade.com

Diante da ausência de tempo de TV, JB utilizou as redes como mecanismo de resposta. Este contorno e a característica da rede deram corpo à candidatura do candidato. Abaixo, na Figura 2, segue o fluxo de postagens dos dois candidatos Jair Bolsonaro (JB) e Fernando Haddad (FH) durante o processo eleitoral.

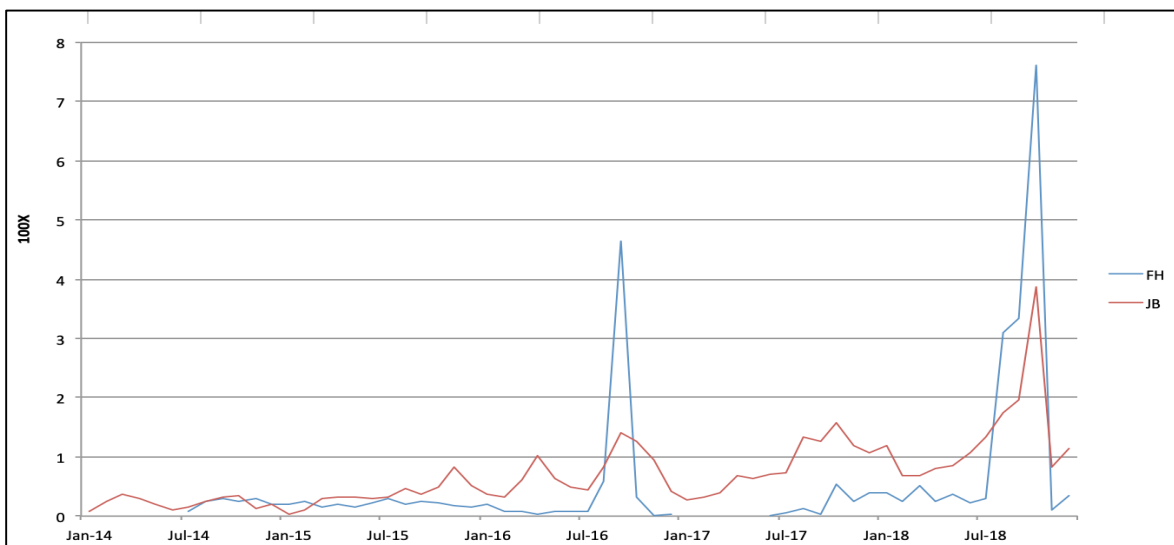


Figura 2 - Publicações de tweets JB e FH

Fonte: Elaborado pelo autor

Na Figura 2, acima, pode-se observar uma utilização constante da ferramenta, sendo esta intensificada nos meses que antecederam o pleito. Não há como relacionar tweets elaborados por FH antes de janeiro de 2017 ao processo eleitoral de 2018. Isto devido ao candidato em questão estar concorrendo a prefeitura de São Paulo no período. Somente após o hiato de janeiro de 2017 a julho de 2017 que podemos afirmar que a campanha do candidato na rede iniciou.

2.3. *FREE MEDIA COVERAGE*

No Brasil o horário político obrigatório é regulamentado pela lei n 9.504/97 com alterações baseadas na lei n 13.165/2015, que calcula o tempo de TV proporcionalmente a partir do tamanho da bancada do partido do candidato no congresso, por este critério a campanha televisiva do candidato Bolsonaro ficou limitada a 8 segundos.

É comum que as postagens do Twitter acabem extravasando os meios digitais, uma vez que os meios de comunicação constantemente trataram o *feed* do Twitter de candidatos como notícia (Ott, 2017). O Twitter permite que o candidato contorne as barreiras dos meios de comunicação convencionais, fazendo que estes meios ajudem a disseminar cada *tweet*, como observado por Gross e Johnson (2016) na eleição do então candidato Donald Trump.

Schoeder (2018), descreve que, no caso da eleição de Trump, foi essencial para a candidatura, a capacidade do candidato de continuar postando mensagens mesmo sendo coberto negativamente pela grande mídia, o que reforça as alegações do candidato que a mídia era tendenciosa contra ele.

No Brasil, a estratégia de Bolsonaro quase que seguiu a mesma cartilha: grande volume de postagens, mídias noticiando os *tweets* e as mensagens polêmicas e a característica de ser uma eleição declaradamente barata em relação aos demais candidatos. Para a eleição americana o “*free extensive media coverage*” também permitiu que o candidato gastasse menos com propaganda eleitoral. (Schoeder, 2018)

Involuntariamente Bolsonaro se manteve na grande mídia a partir do atentado contra sua vida, ocorrido em 6 de setembro de 2018, fato que acabou lhe favorecendo.

Capítulo 3

3. METODOLOGIA

Latent Dirichlet Allocation (LDA) é um modelo de probabilidade generativa (gerador de dados). Nele, cada documento representa uma mistura de tópicos e cada tópico uma distribuição de palavras (Blei, Ng & Edu, 2003), conforme demonstrado na Figura 3, abaixo:

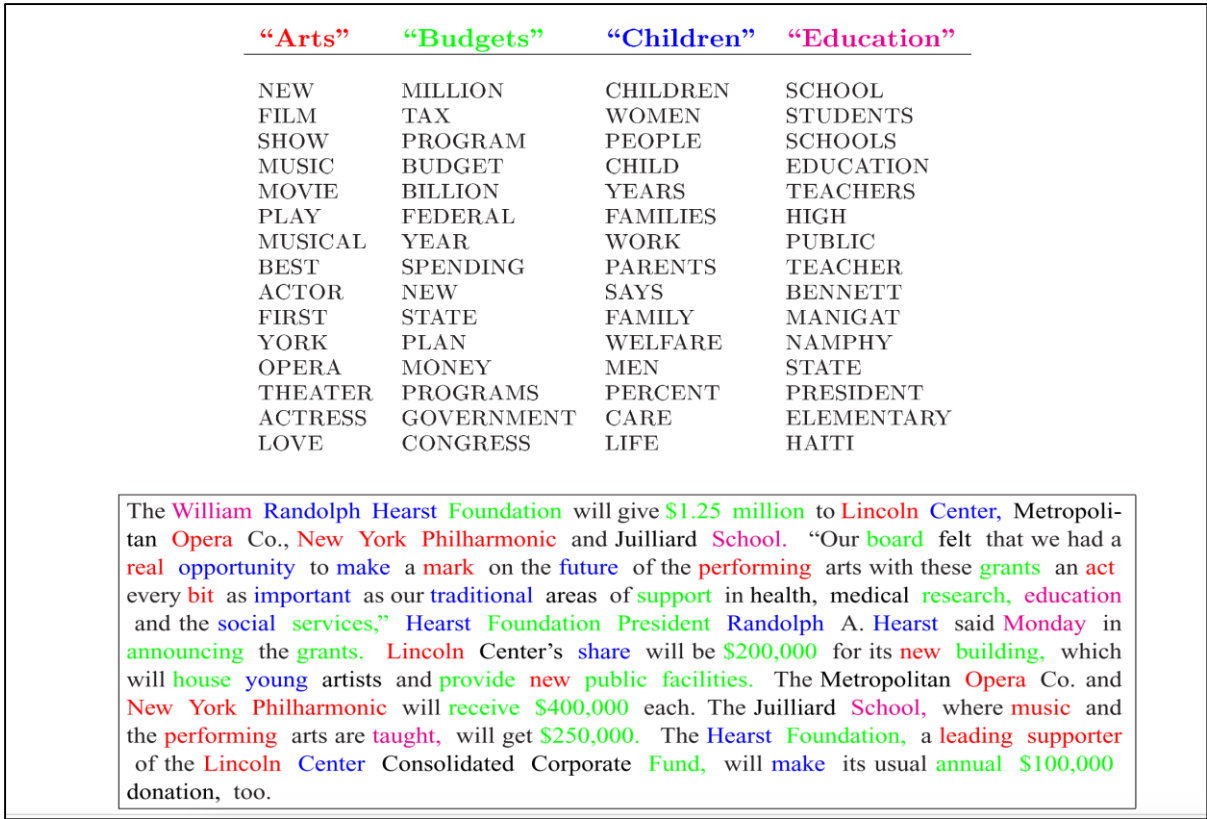


Figura 3 - Exemplo de artigo com a demonstração dos tópicos
Fonte: Blei, Ng & Edu, 2003

A técnica tem sido aplicada de forma sistêmica no meio acadêmico. Basilio (2018) utilizou LDA e técnicas de mineração de texto para obter tópicos sobre criminalidade em um banco de dados do serviço de atendimento do 190. Drury (2017) aplicou o LDA em um corpus em português-brasileiro da área de agricultura. Pedroso

(2018) aplicou LDA com o objetivo de identificar casos semelhantes em um corpus do Ministério Público do Distrito Federal e Territórios (MPDFT).

Métodos de *machine learning* em análise de texto no Twitter também têm sido utilizados. Bassilakis (2018) avaliou o conteúdo do Twitter em relação ao debate do *Brexit*, comparando mensagens dos atores (agentes da mídia, parlamentares e etc) das campanhas de *Remain* (contra o *Brexit*) e *Leave* (a favor do *Brexit*) a partir de um pré treinamento utilizando as campanhas oficiais '*Leave.eu* and '*Open Britain*. Os pesquisadores identificaram uma taxa de aprovação crescente para o lado *Remain*; e uma divergência adicional no tópico *Brexit* entre diferentes políticos no período estudado.

Este trabalho, busca-se compreender os motivadores das curtidas, por meio da equação adaptada de Wang (2016) abaixo:

$$\text{Ln}(\text{likes}) = \beta_0 + \beta_1 \text{ataque} + \beta_2 \text{campanha} + \beta_3 \text{weekend} + \text{Ln}(\text{followers}) + \mu$$

likes = número inteiro, quantidade de curtidas do tweet.

ataque = variável binária proveniente da classificação do texto via LDA para tweets do tipo ataque.

campanha = variável binária proveniente da classificação do texto via LDA para tweets do tipo campanha.

weekend = variável binária, classificação do tweet se foi enviado ao final de semana ou não.

followers = número inteiro, número de seguidores na data no qual o tweet foi postado.

Para definir os tópicos do LDA, foram utilizadas as bibliotecas gensim³ e nltk⁴.

Inicialmente, foi necessário montar o dicionário de palavras. Nesta fase utilizou-se de todos os tweets captados, sem a limitação dos *followers* (4001 tweets). Os

³ <https://pypi.org/project/gensim/>

⁴ <https://www.nltk.org>.

tweets captados foram separados e tokenizados utilizando a biblioteca nltk. O mecanismo de tokenizar tem como objetivo decompor os tweets na dimensão do objeto de estudo.

É importante salientar que o Twitter tem uma limitação de 280 caracteres, o que força o usuário a escolher quais termos utilizar de forma a expressar o sentimento da mensagem com poucas palavras. Quando tokenizado, se faz necessário excluir as *stopwords*.

Stopwords são palavras com muita ocorrência em textos e com pouco significado para o contexto estudado, como “o”, “em”, “Assim” e etc. A remoção destas palavras ocorre para que só permaneçam palavras funcionais e para melhorar o processamento. Para Aguiar (2016) *stopword* são palavras de pouco ou nenhum valor semântico, e sua retirada é uma atividade comum para o processamento de textos.

Para a linguagem em português, os dicionários de *stopwords* são limitados, assim foi utilizado o dicionário nativo da biblioteca nltk complementado com o dicionário colaborativo descrito no github⁵.

Uma vez os dados limpos, foi criado um dicionário e introduzido este no modelo de LDA proposto pela biblioteca gensim (python). A rotina lerá todas as palavras tokenizadas e relevantes dos tweets e fará um agrupamento em tópicos.

A definição do número total de temas do LDA foi baseada no valor de *perplexity* do modelo e na recorrência de palavras à medida que o número de tópicos aumentava. Valores baixos de *perplexity* indicam uma melhor performance (Blei, Ng, & Edu, 2003).

⁵ <https://gist.github.com/alopes/5358189>

O quantitativo de tópicos é sugerido pelo programador, ou seja, se faz necessário realizar simulações calculando tanto os tópicos, quanto o valor de *perplexity*. As simulações seguem até que se obtenha tópicos com palavras coerentes à base analisada. Para este trabalho, foram realizadas simulações progressivas iniciando em 2 tópicos e seguindo até 7. Foi observado pouca modificação no valor de *perplexity* entre 3 e 4 tópicos, indicando que o valor de 3 tópicos já seria suficiente.

A decisão, portanto, foi de assumir 3 tópicos. Na Tabela 3, são apresentados grupos de palavras que se enquadram em cada um deles: 0= apoio; 1=ataque; 2= campanha. Tal caracterização foi elaborada a partir das leituras das palavras recorrentes, considerando, especialmente, as 10 top palavras recorrentes de cada tópico. A palavra “Jair Bolsonaro” foi o único bigrama do trabalho, pois pós processamento e tokenização houve ocorrência significativa das palavras em separado. Foi modificado o pré processamento para que as duas palavras fossem considerada juntas quando sequenciais.

TABELA 3 - TÓPICOS LDA

ID	Tópico	Palavras
0	apoio	abraço obrigado forte parabéns liberdade dia considerações chegada esquerda confiança
1	ataque	mídia pt contra esquerda poder governo corrupção noite outros qualquer
2	campanha	“jair Bolsonaro”* pt verdade obrigado assista entrevista rio acima deus fala

Fonte: Elaborado pelo autor (2019)

Abaixo seguem, como exemplos, os tweets mais curtidos de cada um dos tópicos:

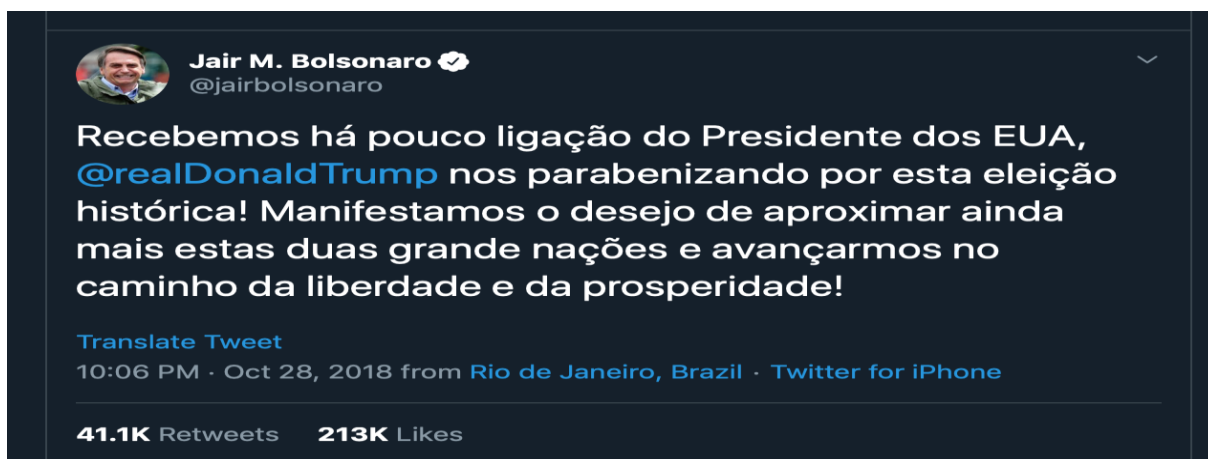


Figura 4: Tweet mais curtido tópico 0

Fonte: Twitter @jairbolsonaro

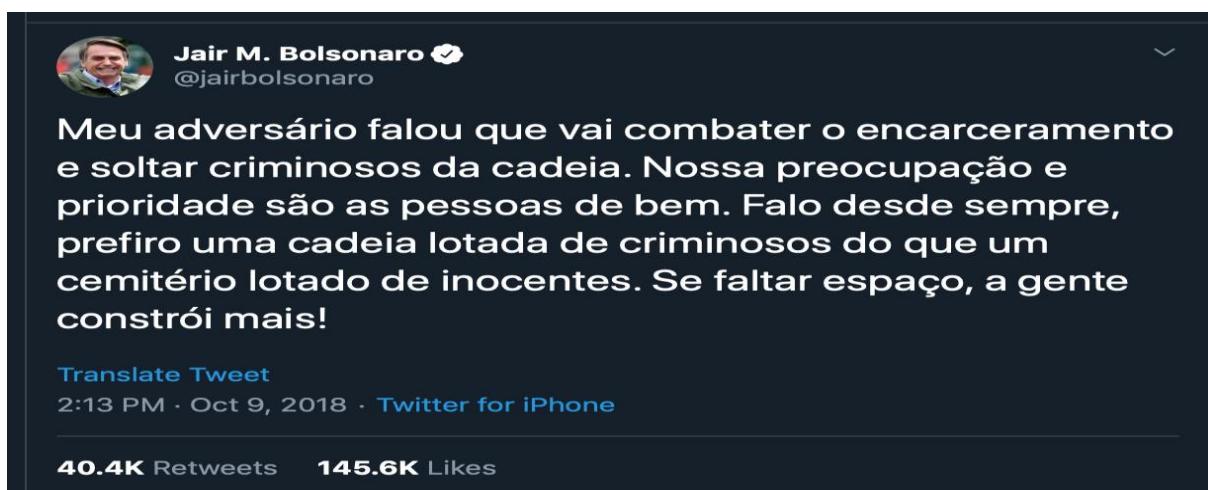


Figura 5: Tweet mais curtido tópico 1

Fonte: Twitter @jairbolsonaro



Figura 6: Tweet mais curtido tópico 2

Fonte: Twitter @jairbolsonaro

Antes da utilização do LDA para modelagem de tópicos, foi feita a classificação manual tweet a tweet. Apesar de esta modalidade implicar em maior probabilidade de acerto, o tempo para sua execução é muito longo (3 meses). Além do mais, a utilização de tal classificação acarretaria em forte viés, uma vez que representaria um olhar singular.

Capítulo 4

4. RESULTADOS

4.1. TÓPICOS LDA

Conforme demonstrado na Tabela 3, foram obtidos 3 tópicos através da rotina de LDA. Para cada uma das rotinas é apresentado também 10 palavras que caracterizam os tópicos.

A classificação dos tweets por tópicos se deu utilizando a biblioteca gensim. Em uma das saídas da biblioteca, pode-se obter a probabilidade de um tweet pertencer ou não a um tópico. Esta probabilidade é calculada dentre outras coisas pelo número da ocorrência de palavras conforme demonstrado na Figura 3. Os Tweets foram classificados utilizando o tópico com maior probabilidade. Os tópicos foram classificados em:

- Apoio: muito comum durante o processo eleitoral, através de mensagens apresentadas pelo candidato de demonstrações de apoio vindas do eleitorado e em diversas destas mensagens o candidato aparece agradecendo. A palavra chegada também caracteriza as viagens do candidato quando ele era recebido no aeroporto por diversos eleitores.
- Ataque: caracteriza as mensagens de ataque direto a blocos ou grupos opostos ao do candidato. Em muito destes ataques o candidato direcionava-se principalmente a partidos de esquerda e ao combate a corrupção. Outro alvo comum do candidato durante todo o processo eleitoral foi a mídia.

- Campanha: as palavras deste tópico são agrupadas principalmente no intuito de divulgar material, por exemplo, “entrevista” e “assista”. Parte do slogan do candidato também é contido no tópico com a presença das palavras “Deus” e “acima”.

4.2. REGRESSÕES

Para adquirir os valores necessários para a equação proposta no item 3, foi adotado a rotina em Python GetOld⁶, rotina esta que extrai, utilizando a API (*Application Programming Interface*) do Twitter, os tweets do então candidato, nos 5 anos que antecederam o pleito. Os dados extraídos compreendem o período entre 01 de Janeiro de 2014 e 28 de outubro de 2018 (segundo turno). Foram um total de 4001 tweets.

Contando com o conteúdo dos tweets, datas de postagem e número de curtidas já extraídos, foi criada a variável *dummy* “weekend”, que classifica se a postagem foi feita em um final de semana ou não, a partir da data da postagem.

Para obter o número de seguidores, no dia da postagem foi utilizado o acesso ao site socialblade.com. Os valores contidos no site estão entre 24/07/2016 a 08/09/2019 (dia da extração), com isso os dados de 01/01/2014 a 23/07/2016 foram desprezados por não conterem a informação *followers*. O período excluído não compreende nenhuma data importante para o processo eleitoral e ainda com a exclusão permaneceram 2925 *tweets* para serem avaliados.

Segue abaixo, na Tabela 4, a estatística descritiva da amostra final:

TABELA 4 - ESTATÍSTICA DESCRITIVA	
Painel A: Estatística Descritiva das Variáveis	

⁶ <https://pypi.org/project/GetOldTweets3>

Variáveis ^a	Obs.	Média	S.D.	Mínimo	Máximo
likes	2925	9370.83	16233.84	10	223098
weekend	2925	0.2680342	0.4430112	0	1
followers	2925	890055.5	500515.1	231396	1967115

Fonte: Elaborado pelo autor (2019)

Em um primeiro momento, optou-se por inserir o número de curtidas e seguidores com o valor absoluto, porém é observado que estes valores representam coeficientes da casa de milhão e dispersão de valores. Assim foi utilizado o LogN para que fosse feito uma análise de elasticidade. Com a mudança, os valores de R^2 aumentaram significativamente indicando que a curva de *likes* pode atender a uma curva do tipo exponencial.

Foram feitas 3 regressões individuais, uma para cada tópico, visando encontrar qual tinha a melhor relevância para o método conforme fórmula proposta no item 3. Uma vez identificando que somente o tópico de ataque foi estatisticamente significativo e diretamente correlacionado com as curtidas, será demonstrado na Tabela 5 abaixo o tópico 0 e o tópico 1 na mesma regressão para avaliarr a efetividade da variável *topic_1*.

Nem todos os tópicos foram inseridos nesta regressão, uma vez que foram construídas variáveis *dummy*, e o método eliminaria uma das variáveis de qualquer maneira.

TABELA 5 - MODELO EMPÍRICO - LIKES X TOPIC_0 X TOPIC_1

Variável Dependente ^a	log_likes	
Variável Independente ^b e Controles ^c	Coef.	t-stat ^d
Constante	-6.97636	-41.2800***
topic_0	-0.00252	-0.1300*
topic_1	0.12676	6.8600***
weekend	0.01184	0.6700*
log_followers	1.78102	61.8500***
Nº de Obs.		2925.0000
R ²		0.5669

Esta tabela contém os coeficientes estimados para o método dos mínimos quadrados, a variável *topic_0* é a variável independente que sinaliza ter sido ou não classificada pelo método computacional como do *topico_0*, sendo 1 fazendo parte do tópico e 0 não fazendo parte do tópico. A variável *topic_1* é a variável independente que sinaliza ter sido ou não classificada pelo método computacional como do *topico_0*, sendo 1 fazendo parte do tópico e 0 não fazendo parte do tópico. A Variável *weekend* é uma variável independente, onde 1 o tweet foi postado em um final de semana, e 0 durante a semana (não foram considerado feriados). A variável *followers* é uma independente de valor inteiro que representa o número de seguidores que o emissor possuía no dia da postagem da mensagem. *t-stat^d* representa a (estatística T) se a variável, seja ela a independente ou as de controle, são significativas, sendo que *, ** e *** indica estatisticamente significativo com 10%, 5% e 1%, respectivamente.

Fonte: Elaborado pela autor (2019)

Conforme observado na Tabela 5, observa-se um R^2 de 0,5669. O tópico 0(apoio), não foi significativo e apresentou uma relação negativa em relação ao número de curtidas. O tópico 1 (ataque) foi significativo e apresentou relação positiva com o Log de curtidas. Weekend apresentou relação positiva porém foi pouco significante. Já *log_followers* apresentou relação positiva e significância como já se esperava, quanto mais seguidores mais curtidas.

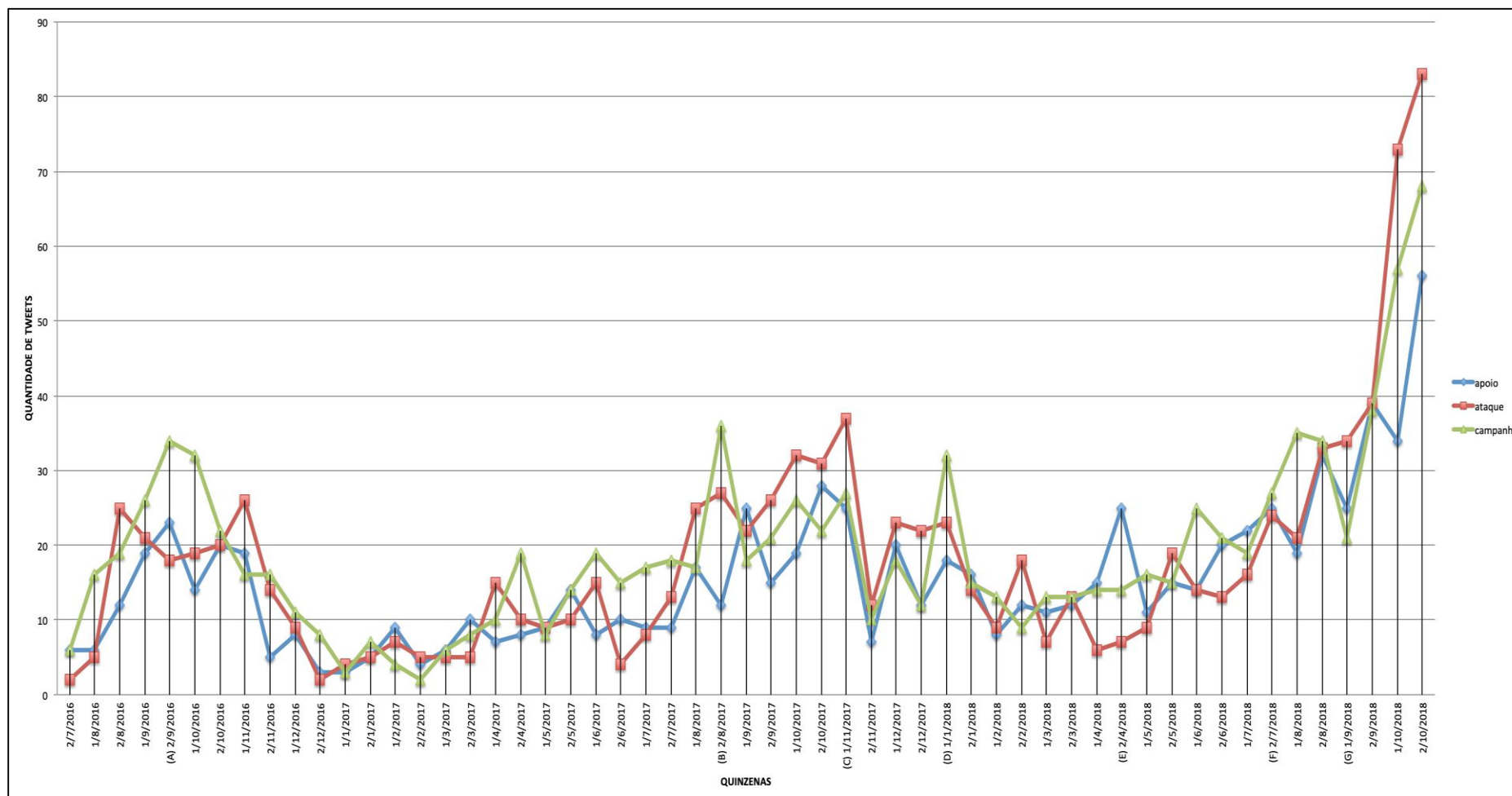


Figura 7: Análise horizontal - número de tweets por tópico

Fonte: elaborado pelo autor (2019)

Pode-se observar, na Figura 7, que o ápice de mensagens classificadas como campanha concentra-se na segunda quinzena de setembro/2016 e primeira quinzena de outubro/2016 (caracterizado pela letra A no gráfico). Este volume de postagens é decorrente das eleições municipais daquele ano, onde o filho do presidente concorreu ao cargo de vereador da cidade do Rio de Janeiro.

Para os campos identificados com as letras B e D, no gráfico, não foi possível identificar um motivador do aumento de mensagens classificadas como campanha. O aumento de mensagens do tipo ataque (caracterizado no gráfico com C), foi motivado por publicações em revistas no período criticando o candidato. Nesse período, foram observados diversos ataques à mídia em geral. Em “E”, o candidato intensificou postagem de mensagens de apoio devido a viagens feitas para diversos locais do Brasil.

A oficialização da candidatura do candidato foi feita no dia 22/07/2018 (F na Figura 7), observa-se que a partir desta data os tweets se intensificam até o desfecho da eleição. Por fim, não há como caracterizar uma perda de conteúdo significativo durante o período que o candidato esteve se recuperando do atentado que sofreu no dia 06/09/2018 (G na Figura 7), houve uma queda do material de campanha e apoio retornando imediatamente na quinzena seguinte.

4.3. ABRANGÊNCIA DO TÓPICO 1

Diante do observado na regressão (Tabela 5), o tópico 1 foi fortemente associado aos números de curtidas. Ataques feitos pelo candidato obtiveram reações positivas de seu eleitorado mais do que os tópicos de apoio e de campanha.

O tópico 1 (ataque) permaneceu com relação diretamente proporcional ao número de curtidas.

TABELA 6 - MATRIZ DE CORRELAÇÃO PALAVRAS TOPIC_1

	mídia	pt	contra	esquerda	poder	governo	corrupção	Noite	outros	qualquer
mídia	1.000 0									
pt	0.017 6	1.000 0								
contra	0.089 9*	0.055 8*	1.000 0							
esquerda	0.106 7*	- 0.002 7	0.033 7*	1.0000						
poder	0.001 0	0.032 9	0.011 4	0.0606 *	1.000 0					
governo	0.038 8*	0.051 4*	0.027 4	- 0.0205	0.084 0*	1.000 0				
Corrupção	0.036 9*	0.078 1*	0.059 3*	- 0.0153	0.087 3*	0.064 8*	1.0000			
noite	0.002 7	- 0.027 3	- 0.021 0	- 0.0078	0.023 7	- 0.004 5	-0.0156	1.000 0		
outros	0.021 7	0.012 5	0.029 0	0.0064	0.025 7	0.010 9	0.0175	0.008 3	1.000 0	
qualquer	0.071 9*	0.038 7*	0.098 6*	0.0299	0.061 4*	0.037 0*	0.0319*	0.007 3	0.049 5*	1.0000

Ocorrência das Palavras nos tweets: mídia, pt, contra, esquerda, poder, governo, corrupção, noite, outros, qualquer. * significância a 95%

Fonte: Elaborado pela autor (2019)

Na Tabela 6, observa-se a relação entre as top 10 palavras do tópico 1. A palavra “esquerda” tem relação significativa com a palavra “mídia”, tal como a palavra “contra” e a palavra “mídia”. Estes dados demonstram parte da estratégia do candidato que através das redes sociais e até mesmo após o pleito segue rotinas de ataque aos meios convencionais de comunicação.

As palavras “contra” e “pt” são outras marcas da campanha ofensiva do candidato. O Partido dos trabalhadores (PT) ficou no governo por 16 anos antes do pleito e por fim foi associado a casos de corrupção que inclusive desencadearam o impeachment da presidente Dilma Rousseff. Isso corrobora outro grupo de associação de palavras: “corrupção” e “pt”.

A palavra “qualquer” pode ser considerada sem significado para o contexto do estudo, como ela não estava em nenhuma das listas de palavras disponíveis (*stopwords*), acabou sendo incluída no método. Contudo, não pode ser recomendada sua manutenção.

Outro questionamento sobre o método é qual o fator de abrangência, isto é, do quantitativo de tweets qual foi o total de tweets influenciados pelo tópico 1, ou em quantos tweets as palavras aparecem. Se faz necessário avaliar, dentre o total de mensagens postadas pelo candidato, em quais delas aparecem as 10 palavras referidas da modelagem de tópico de ataque.

TABELA 7 - TOTAL DE OCORRÊNCIAS TÓPICO 1

total_ocorrência	total_tweets	%
0	3244	81.08%
1	612	15.30%
2	117	2.92%
3	26	0.65%
4	2	0.05%
	4001	100.00%

Fonte: Elaborado pelo autor (2019)

As palavras que compõem o tópico 1 aparecem em quase 18,92% da amostra total conforme demonstrado na Tabela 7. Mesmo com este fator a relação entre o tópico e as curtidas foi fortemente explicada. Adicionalmente, o tópico 1 também apresentou relação positiva com retweets, conforme Tabela 8, abaixo.

TABELA 8 - MODELO EMPÍRICO - RTS X TOPIC_1

Variável Dependente ^a	Retweets	
Variável Independente ^b e Controles ^c	Coef.	t-stat ^d
Constante	-2280.24200	-17.8500*
topic_1	1137.43000	9.3700*
Weekend	148.31150	1,14***
Followers	0.00453	39.3400*
Nº de Obs.		2925.0000
R ²		0.3629

Esta tabela contém os coeficientes estimados para o método dos mínimos quadrados, a variável *topic_0* é a variável independente que sinaliza ter sido ou não classificada pelo método computacional como do *topico_0*, sendo 1 fazendo parte do tópico e 0 não fazendo parte do método. A Variável *weekend* é uma variável independente onde 1 o tweet foi postado em um final de semana, e 0 durante a semana (não foi considerado feriados). A variável *followers* é uma independente de valor inteiro que representa o número de seguidores que o emissor possuía no dia da postagem da mensagem. ^d t-stat representa se a variável, seja ela a independente ou as de controle, são significativas, sendo que *, ** e *** indica estatisticamente significativo com 1%, 5% e 10%, respectivamente.

Fonte: Elaborado pela autor (2019)

4.4. WORD2VEC

O word2vec funciona tomando como input um corpus de palavras e retornando vetores de palavras (Aguiar, 2016). Assim, ao transformar as palavras em vetores, é como se ocorressem operações aritméticas para calcular as distâncias ou simplesmente a visualização gráfica.

De posse do conjunto de palavras separada e tokenizada (*corpus*) bastava aplicar o word2vec na mesma amostra para obter o valor dos vetores. Visando melhorar a visualização no gráfico foi aplicado também o TSNE⁷, outro algoritmo de *machine learning* que visa reduzir os dados multidimensionais e o plota em 2 duas ou mais dimensões mantendo a relação entre as palavras da amostra. O TSNE cria uma nova representação mantendo as relações vetoriais entre as palavras.

Assim a partir da representação demonstrada na Figura 8, pode ser vista a proximidade das palavras “flaviobolsonaro” e “carlosbolsonaro”. O nome do terceiro filho de Bolsonaro apresenta-se mais distante dos outros dois, porém, mais relacionado com o nome do candidato, “Jair Bolsonaro”. Há também uma distância

⁷ <https://scikit-learn.org/stable/modules/manifold.html>

pequena entre as palavras, “pt”, comparado à “segurança”, “país”, “liberdade” e “jair Bolsonaro”.

“Poder” e “corrupção” aparecem bem como “população”, “esquerda” e “diretos” retornaram com relativa proximidade. Isto pode se dar pelos ataques diretos, associando os governos de esquerda aos casos de corrupção e esquerda e diretos reforçam o ataque massivo em campanha aos diretos humanos, bandeira que Bolsonaro atribui principalmente aos partidos de esquerda.

Outra proximidade interessante é “Israel”, “respeito”, “confiança” e “amigos”. É declarado o posicionamento do candidato quanto à Israel tendo inclusive visitado o país em março de 2019. É interessante notar, também, que as palavras “TV”, “jornal”, “programa entrevistas” ficaram próximas e distantes das ferramentas online utilizadas pelo candidato, “youtube” e “facebook”.

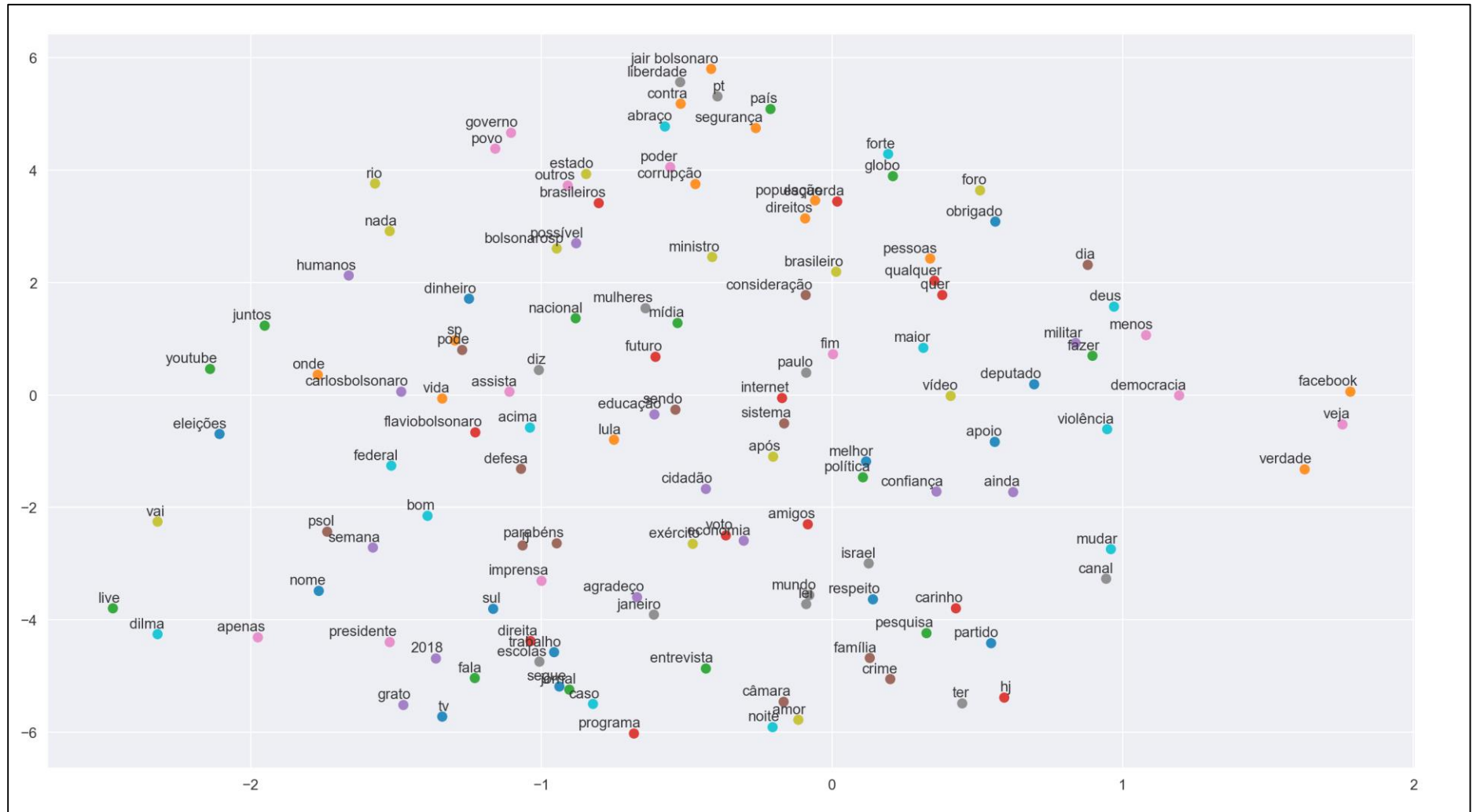


Figura 8 - Representação gráfica Word2Vec

Capítulo 5

5. CONCLUSÃO

JB fez uso de conteúdos com características de simplicidade, impulsividade e em alguns casos, incivilidade. Opinou e convocou, mantendo um volume significativo de postagens, quase sempre controversas aos meios de comunicação convencionais, que acabavam noticiando estas mensagens, na maioria das vezes, criticando o candidato. Assim, mantendo um volume de postagens maior que os adversários, ele acabou dominando, também, a mídia convencional, fenômeno conhecido como *free media coverage* (FMC). Uma estratégia semelhante a campanha do presidente americano Donald Trump, que também se apoiou na propagação e divulgação do conteúdo gerado originalmente no Twitter.

A partir dos resultados pode-se inferir que o, agora, presidente Jair Bolsonaro também se apoiou na FMC. Assim, o presente trabalho tem sua conclusão alinhada com a pesquisa de Hrenechen e Gadini (2018): o Twitter é uma ferramenta onde o conteúdo gerado torna-se “fonte” para a produção de notícias por veículos tradicionais.

Os tópicos que mais motivaram curtidas no perfil do candidato foram os envolvendo ataques diretos à oposição ou mídia. Se definirmos os 3 tópicos propostos como 3 estratégias de campanha, a divulgação de material de campanha e mensagens demonstrando apoio motivam menos o eleitorado que mensagens de confronto.

Analisando a conta de Jair Bolsonaro no Twitter, há uma impressão de que ainda que a sua audiência na plataforma um pouco mais de 10% do quantitativo de

votos obtidos no primeiro turno (5,5M de seguidores na plataforma em 2019 para 49M de votos no primeiro turno), este não foi o canal que atingiu grande parte dos seu eleitorado de forma direta. De outra maneira, aparentemente o conteúdo ali gerado foi levado à população por diferentes veículos de comunicação.

Utilizar um sistema de *machine learning* para leitura de diversos tweets pode ser efetivo para um candidato ou para uma empresa medir qual é o engajamento de seus potenciais eleitores ou compradores sobre um produto ou utilizar este mecanismo para monitorar mensagens da concorrência.

É essencial entender no processo eleitoral de 2018 o impacto de outros fatores, tais como o Whatsapp e melhor entender principalmente o *free media coverage*, pois após o atentado ao candidato, ele praticamente dominou os noticiários no período.

Este estudo possui algumas limitações. Não foi possível implantar duplas e triplas neste trabalho. A implantação ou a conjugação de tokens transmitem melhor a ideia descrita no texto. Um exemplo seria o token “direitos” e a dupla “direitos humanos”. Os dois casos descritos podem representar ideias totalmente divergentes. O uso de duplas e triplas serve como uma recomendação para estudos futuros, pois pode melhorar a margem de acerto do método ou utilizar em conjunto com o LDA outros métodos de *machine learning*, tais como o word2vec.

Avaliar dados emitidos no passado também pode ter suas limitações. Perfis podem ser deletados e mensagens e a dinâmica política do momento não é captada. Um candidato no auge de sua popularidade no período de campanha pode reverter em uma rejeição nos primeiros meses de mandato. Por isso, se faz necessário que estudos futuros tenham um monitoramento mais próximo do dado coletado.

O número de curtidas pode ser um fator de engajamento como foi descrito acima, mas se faz necessário medir o teor das mensagens respostas ao tweets. Um processo de análise de sentimento em paralelo com o LDA poderia melhor descrever qual o sentimento dos usuários da rede para com as mensagens de um determinado tópico postado. Outra possibilidade seria avaliar a quantidade de palavras por publicação.

REFERÊNCIAS

- Aguiar, E. M. de. (2016). *Aplicação do Word2vec e do Gradiente descendente dstocástico em tradução automática*, Doctoral dissertation, Fundação Getúlio Vargas. Rio de Janeiro-RJ.
- Azmi, A., Sylvia, I., & Mardhiah, D. (2018). Indonesian Politicians' Social Media Role in Encouraging Civic Engagement. *Advances in Social Science, Education and Humanities Research*, 251, 305–310. Recuperado em 02 junho, 2019, <https://doi.org/10.2991/acec-18.2018.70>.
- Basilio, m., & Pereira, V. (2018). *A Escolha da Estratégia de Policiamento em Função da Demanda Criminal: um Modelo Probabilístico de Tópicos*. ENEGEP 2018 - Encontro Nacional de Engenharia de Produção. Recuperado em 02 de junho, 2019, https://doi.org/10.14488/enegep2018_tn_stp_263_511_35839.
- Bassilakis, A., Diermeier, M., & Goecke, H. (2018). Converging Media versus Diverging Politics - the Brexit Twitter on Debate. *CESifo Forum*, 19(4), 29–35.
- Blei, D. M., Ng, A. Y., & Edu, J. B. (2003). Latent Dirichlet Allocation Michael I. Jordan. *Journal of Machine Learning Research*, 3, 993-1022.
- Braga, S., & Carlomagno, M. (2018). Eleições como de costume? Uma análise longitudinal das mudanças provocadas nas campanhas eleitorais brasileiras pelas tecnologias digitais (1998-2016). *Revista Brasileira de Ciência Política*, 26, 7–62. Recuperado em 05 de junho, 2019, <https://doi.org/10.1590/0103-335220182601>.
- Cioccarri, D., & Persischetti, S. (2018). A política e o espetáculo em Jair Bolsonaro, João Doria e Nelson Marchezan. *Revista de La Asociación Argentina de Estudios de Cine y Audiovisual*, 18(2), 54–84.
- Cremonese, D. (2012). On-line politics: The use of Twitter as a tool of social capital in the 2010 presidential elections. *Sociedade e Cultura*, 15(1), 135–149. Recuperado em 05 de junho, 2019, <https://doi.org/10.5216/sec.v15i1.20679>.
- Drury, B., Fernandes, R., & De Andrade Lopes, A. (2017). *BrAgriNews: Um Corpus Temporal-Causal (Portugues-Brasileiro) para a Agricultura*. *Linguamatica*, 9(1), Recuperado em 06 de julho, 2019, 41–54. <https://doi.org/10.21814/lm.9.1.245>.
- Gross, J. H., & Johnson, K. T. (2016). Twitter Taunts and Tirades: Negative Campaigning in the Age of Trump. *PS - Political Science and Politics*, 49(4), 748–754. Recuperado em 06 de julho, 2019, <https://doi.org/10.1017/S1049096516001700>.
- Hosch-Dayican, B., Amrit, C., Aarts, K., & Dassen, A. (2014). How Do Online Citizens Persuade Fellow Voters? Using Twitter During the 2012 Dutch Parliamentary Election Campaign. *Social Science Computer Review*, 34(2), 135–152.

- Hrenechen, V. C. de A. T., & Gadini, S. L. (2018). O papel do Twitter no agendamento jornalístico em mídia regional no Paraná. *Novos Olhares*, 7(2), 64–77.
- Lilleker, D. G., & Vedel, T. (2013). The Internet in Campaigns and Elections. In *The Oxford Handbook of Internet Studies* (pp. 1–38). Recuperado em 06 de julho, 2019, <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199589074.001.0001/oxfordhb-9780199589074-e-19>.
- Lopes, E. D. (2015). Utilização do modelo skip-gram para representação distribuída de palavras no projeto Media Cloud Brasil. Doctoral dissertation. Fundação Getúlio Vargas. Rio de Janeiro-RJ.
- Lopes, M. F. (2018). Bolsonaro pode bloquear pessoas que o criticam no Twitter? *El Pais - Opinião*, 218.
- Malini, F., & Antoun, H. (2013). *@internet e #rua*.
- Mikolov, T., Yih, W., & Zweig, G. (2013). Linguistic Regularities in Continuous Space Word Representations Tomas. *Association for Computational Linguistics*, 746–751. Recuperado em 05 de junho, 2019, <https://doi.org/10.3109/10826089109058901>.
- Ott, B. L. (2017). The age of Twitter: Donald J. Trump and the politics of debasement. *Critical Studies in Media Communication*. Routledge. Recuperado em 1 de janeiro, 2019, <https://doi.org/10.1080/15295036.2016.1266686>.
- Paulo, F., & Almeida, J. (2011). Internet e eleições 2010 no Brasil: rupturas e continuidades nos padrões mediáticos das campanhas políticas online. *Galáxia*, 22(1), 208–221.
- Pedroso, D. de S. C. (2018). *Identificação automática de casos repetitivos no MPDFT*. Universidade de Brasília. Brasília-DF.
- Percastre-Mendizábal, S. J. (2018). Twitter en campaña. Actores mediáticos en Twitter durante la campaña electoral intermedia en México en el año 2015. *Hipertext.Net: Revista Académica Sobre Documentación Digital y Comunicación Interactiva*, 17. Recuperado e, 5 de junho, 2018, <https://doi.org/10.31009/hipertext.net.2018.i17.09>.
- Recuero, R., Zago, G., & Soares, F. B. (2017). *Conversations on Twitter, 1–27. Mídia social e filtros-bolha nas conversações políticas no Twitter*. Associação Nacional de Programas de Pós-Graduação em Comunicação. Encontro Anual (COMPÓS). São Paulo, SP.[Anais]. São Paulo: Faculdade Cásper Líbero, 2017.
- Schroeder, R. (2018). Digital media and the rise of right- wing populism Studies. In *Social Theory after the Internet*. UCL Press. *Theory and Society*, 46(5), 357-385.
- Smith, B. G., & Gallicano, T. D. (2015). Terms of engagement: Analyzing public engagement with organizations through social media. *Computers in Human Behavior*, 53, 82–90. Recuperado em 08 de junho, 2018,

<https://doi.org/10.1016/j.chb.2015.05.060>.

- Statesman, N. E. W. (2014). Twitter politics is no substitute for ideas and strong campaigns. Recuperado em 08 de julho, 2018, <https://www.newstatesman.com/politics/2014/04/leader-twitter-politics-no-substitute-ideas-and-strong-campaigns>.
- Tufekci, Z. (2015). Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency. *Journal on Telecommunications & High Tech Law*, 13(23), 203–216.
- Valente, M. R. M., & silva, M. L. H. da. (2010). *A utilização do Twitter na campanha política e sua aplicação no Tocantins: estudo de caso do perfil do candidato a Governador eleito Siqueira Campos*. Congresso Panamericano de Comunicação. Brasília-DF.
- Wang, Y., Luo, J., Niemi, R., Li, Y., & Hu, T. (2016). *Catching Fire via “Likes”*: Inferring Topic Preferences of Trump Followers on Twitter. Recuperado em 09 de junho, 2018, <http://arxiv.org/abs/1603.03099>.