

Uma avaliação do potencial de uso dos dados do Twitter para a predição do resultado de eleições: O caso das eleições presidenciais brasileiras de 2018

Mariana Lana Sales¹, Marcelo Werneck Barbosa²

Resumo

As eleições presidenciais brasileiras de 2018 mostraram forte presença dos candidatos em redes sociais com a intenção de abranger a maior parcela possível do eleitorado. Os 5 candidatos que tinham mais intenção de voto em pesquisas eleitorais usaram assiduamente o Twitter como ferramenta para alcançar visibilidade das opiniões e projetos que propuseram. Por meio de um estudo exploratório, esse trabalho busca investigar o potencial do uso de dados do Twitter como ferramenta de predição dos resultados da eleição. Para isso, foram coletados dos perfis dos candidatos alguns dados relacionados a tweets e seguidores, e em seguida, estes dados foram utilizados para construir previsões que foram comparadas com a principal pesquisa de intenção de voto e com o resultado eleitoral. Os resultados mostram que o uso destes dados do Twitter tem um bom potencial de predição.

Palavras-chave: Análise preditiva. Redes sociais. Eleições presidenciais. Análise com Twitter. Coleta de dados.

Abstract

The Brazilian presidential elections of 2018 showed a strong presence of candidates in social networks with the intention of reaching the largest possible portion of their electorate. The top 5 candidates, according to polls, have frequently used Twitter as a tool to give visibility to their opinions and projects. Through an exploratory study, this work investigates the potential of using Twitter data as a tool to predict the results of the election. In order to do this, some data related to tweets and followers were collected from the candidate profiles, and then these data were used to construct forecasts that were compared with polls and with the election result itself. The results show that using this Twitter data has a good predictive potential.

Keywords: Predictive analysis. Social media. Presidential elections. Twitter analysis. Data collection.

¹ Pontifícia Universidade Católica de Minas, Minas Gerais, Brasil. E-mail: marianasalan@outlook.com. Orcid: <https://orcid.org/0000-0002-9438-9963>

² Pontifícia Universidade Católica de Minas, Minas Gerais, Brasil. E-mail: mwerneck@pucminas.br. Orcid: <https://orcid.org/0000-0003-3993-5375>

1 Introdução

As redes sociais são plataformas de interação entre pessoas nas quais elas criam, compartilham ou trocam informações e ideias em comunidades virtuais. O uso de redes sociais aumentou demasiadamente nos últimos anos, como afirma o instituto de pesquisa Statista (2016) e como nós mesmos pudemos atestar ao longo dos últimos anos observando os ciclos sociais em que convivemos.

Uma das mais conhecidas redes sociais é o Twitter, que surgiu em março de 2006. Seu funcionamento e estrutura serão cobertos na próxima seção. Apesar de ter perdido o crescimento expressivo de usuários que tinha em seus anos de maior popularidade segundo o instituto Statista (2019), o Twitter ainda é uma rede social de enorme importância, amplamente utilizada por muitas pessoas e organizações.

O Twitter tornou-se significativamente mais visível como meio de comunicação política durante as campanhas presidenciais americanas de 2008. Barack Obama e sua equipe foram usuários assíduos do Twitter durante o período eleitoral, e ele continuou a usar o Twitter após eleito (MISSINGHAM, 2010). Segundo Marques et al. (2011), a internet acaba conformando uma alternativa à qual os candidatos recorrem na intenção de estabelecer um contato mais direto com o eleitorado, evitando a intermediação da mídia tradicional, que adota critérios próprios de noticiabilidade e de oferta de visibilidade. No âmbito do olhar do eleitor, os autores ainda acrescentam que a própria disponibilidade do candidato para atualizar as postagens de modo frequente já permite que o eleitor construa uma imagem pública diferente do presidenciável, algo nem sempre alcançado através das modalidades convencionais de comunicação política.

O estudo Twiplomacy (2018) identificou 372 contas pessoais e 579 contas institucionais

pertencentes a chefes de estado e ministros de relações exteriores de 187 países no Twitter. Para efeitos de comparação, o Facebook está presente em 179 países, tornando o Twitter a rede social mais popular entre organizações e chefes políticos das nações. O número de usuários ativos do Twitter é muito menor que os 2,3 bilhões do Facebook. Dados do próprio Twitter mostram que o número de usuários do Twitter é de cerca de 15% dos usuários totais do Facebook. Assim, exalta-se a relevância política desta rede, que mesmo com quase 2 bilhões de usuários a menos, tem a presença de figuras e instituições políticas maior que o Facebook.

Analisar o comportamento de pessoas no Twitter envolve avaliar diferentes tipos de dados: seguidores, curtidas, postagens, respostas, visualizações, entre outros. Tendo como objeto de estudo os perfis dos candidatos a Presidente da República no Brasil nas eleições de 2018 no Twitter, esse estudo busca avaliar alguns dados coletados do Twitter para responder às seguintes perguntas de pesquisa:

- P1 – A popularidade dos candidatos no Twitter reflete o número de votos e intenções de voto nas pesquisas?
- P2 - A popularidade dos candidatos na rede pode apoiar a previsão do resultado eleitoral?

Para responder a estas questões, foi conduzido um experimento de coleta de dados da rede seguido por sua análise.

Este texto está organizado da seguinte forma. O contexto político brasileiro e o funcionamento do Twitter são apresentados na Seção 2, assim como alguns trabalhos relacionados. Na Seção 3, é descrita a metodologia do trabalho. Na Seção 4, são apresentados os resultados do trabalho e é realizada a análise dos dados e discussão. Por fim, a Seção 5 apresenta as conclusões e

limitações deste estudo bem como possíveis trabalhos futuros.

2. Contextualização

2.1 O cenário político brasileiro

Durante a década de 1990, com o Brasil recém-saído do período de ditadura militar, governaram presidentes de 3 partidos diferentes. Após Fernando Henrique Cardoso governar o país por 8 anos, em 2002, os eleitores brasileiros elegeram Luiz Inácio Lula da Silva, do Partido dos Trabalhadores (PT), que governou de 2003 a 2010. Em 2010, Dilma Rousseff, também do Partido dos Trabalhadores, foi eleita e, posteriormente, reeleita em 2014. A Presidente Dilma Rousseff sofreu um processo de impeachment em 2016, sendo obrigada a deixar o governo nas mãos de seu vice, Michel Temer, do Movimento Democrático Brasileiro (MDB), que governou até a posse do candidato escolhido nas eleições presidenciais a que esse artigo se refere.

Nas eleições de 2018, 13 candidatos concorreram à Presidência da República. Entre eles, os cinco que tinham mais intenção de voto na primeira pesquisa do período eleitoral realizada pelo Instituto IBOPE (2018) foram selecionados para este estudo: (1) Ciro Gomes, do Partido Democrático Trabalhista (PDT), já foi Prefeito, Governador, Deputado Estadual, Deputado Federal, Ministro da Fazenda e Ministro da Integração Nacional. Ele concorreu à sua terceira eleição presidencial (1998, 2002, 2018) e nunca foi eleito. (2) Fernando Haddad, do Partido dos Trabalhadores (PT), foi Ministro da Educação de 2005 a 2012 e Prefeito de São Paulo de 2013 a 2017. Em 2018, lançou-se candidato apenas no mês anterior à eleição pois o candidato anunciado pelo partido seria o ex-presidente Lula, que foi condenado e não obteve permissão judicial para seguir com a candidatura. (3)

Geraldo Alckmin, do Partido da Social Democracia Brasileira (PSDB), partido no qual está há 30 anos, já foi Vereador, Prefeito, Deputado Estadual, Deputado Federal, Governador de São Paulo e Presidente do PSDB. Foi sua segunda candidatura à presidência (foi derrotado por Lula no segundo turno em 2006). (4) Jair Bolsonaro, do Partido Social Liberal (PSL), foi Vereador do Rio de Janeiro e se elegeu Deputado Federal pelo Rio de Janeiro por sete mandatos, de 1991 até 2018, variando entre diversos partidos durante tais mandatos. (5) Marina Silva, candidata pela Rede Sustentabilidade (REDE), partido que lidera, tem também longa carreira política, tendo sido Vereadora, Deputada Estadual, Ministra do Meio Ambiente e Senadora por diferentes partidos. Marina concorreu à sua terceira eleição presidencial consecutiva (2010, 2014, 2018) e como Ciro Gomes, nunca foi eleita.

2.2 O uso do Twitter

Os usuários utilizam o Twitter para reportarem as novidades mais recentes ou comentarem sobre eventos atuais. Dada esta natureza dinâmica da rede social, ela tem um grande potencial em explorar e entender as pessoas, suas vidas, preocupações, interesses e opiniões. Em síntese, é uma ótima ferramenta para saber o que as pessoas estão falando espontaneamente sobre tudo o que acontece (AL-KHALIFA, 2012; SAVAGE, 2011).

De acordo com Bode et al. (2016), ao contrário dos esforços tradicionais de campanha que utilizam anúncios de TV, e-mail e *telemarketing* para influenciar eleitores, o Twitter apresenta uma oportunidade para expandir e diversificar o discurso político indistintamente. Impulsionados pela ausência de custo, todos os candidatos, inclusive os que não podem ser ouvidos na mídia tradicional, são capazes de envolver o público.

O Twitter tem como principal unidade o *tweet*, uma mensagem curta de no máximo 280 caracteres. A ele podem ser anexados vídeos, imagens, arquivos do tipo *Graphic Interchange Format* (GIF) ou um outro *tweet*.

O *tweet* possui um usuário autor. O usuário autenticado possui uma página inicial (*home*). Nesta página, são apresentados *tweets* mais recentes de contas que tal usuário segue. Da mesma forma, as contas que seguem o usuário (seguidores) terão os *tweets* recentes do mesmo em suas respectivas páginas iniciais.

Por um usuário, o *tweet* pode ser curtido ou *retweetado*. Quando curtido, o *tweet* vai para o perfil do usuário na aba Curtidas. Quando *retweetado*, o *tweet* vai para o perfil do usuário na aba Linha do tempo (*timeline*) e é propagado para todos os seguidores do usuário em suas páginas iniciais (da mesma forma que um *tweet* autoral). Em outras palavras, cada perfil de usuário possui, além de suas informações básicas, um conjunto de *tweets* dividido em duas diferentes abas: Na linha do tempo ficam os *tweets* criados por ele e *tweets retweetados* por ele, enquanto em curtidas ficam os *tweets* curtidos por ele.

2.3 Estudos relacionados

Alguns trabalhos utilizaram dados do Twitter para realizar análises em contextos políticos. Nesta seção, apresentamos alguns destes trabalhos. Marques et al. (2011) realizaram uma análise da atividade do candidato José Serra no Twitter durante os 15 dias anteriores ao segundo turno nas eleições de 2010 e classificaram a amostra de 221 *tweets* do candidato em 5 categorias: promoção de ideias, campanha negativa, mobilização e engajamento, promoção de eventos e temas alheios. O estudo de caso indicou que o Twitter é um elemento importante na estratégia de comunicação política.

Sang e Bos (2012) coletaram, na semana anterior às eleições ao senado holandês em

2011, uma amostra de 64.395 *tweets* que mencionavam partidos. Foram eliminados *tweets* que continham mais de um partido e *tweets* do mesmo usuário (apenas um *tweet* por usuário era relevante). Nos 28.704 *tweets* restantes, foram contadas as menções a cada partido para tentar prever os resultados da eleição. A previsão foi próxima do resultado real. De 74 assentos no senado, a previsão pelo Twitter errou apenas 18, enquanto as maiores pesquisas eleitorais erraram 14.

Gaurav et al. (2013) realizaram também um estudo com uma grande amostra de *tweets* para tentar prever as eleições presidenciais da Venezuela, do Paraguai e do Equador. A coleta continha *tweets* da semana anterior às eleições, e contavam menções ao nome dos candidatos. As previsões foram próximas dos resultados e mostraram que utilizar pseudônimos e palavras chaves pode melhorar a acurácia da previsão.

Bode et al. (2016) realizaram uma comparação entre a campanha de candidatos ao senado americano em 2010 no Twitter e a campanha tradicional, em propagandas na televisão. Enquanto a campanha televisiva demanda mais recursos e mais tempo de produção para alcançar o eleitor, a campanha no Twitter não demanda recursos financeiros e nem requer um longo tempo para o mesmo fim. O que ambos os meios de comunicação têm em comum, como apontam os autores, é que ambos *tweets* e propagandas são divulgados várias vezes por dia, e como mostrou o estudo, tendem a aumentar a frequência conforme a campanha avança, e diminuir nas proximidades do dia da eleição.

Gayo-Avello (2012) citou como falha em muitas pesquisas que utilizam Twitter para previsões eleitorais o fato de que não são previsões de forma alguma, e sim estudos que comprovam que uma previsão poderia ter sido feita, ou seja, a análise ocorre após as eleições, e, claro, resultados negativos são raros de encontrar.

Poucos estudos que trabalhavam atributos quantitativos do Twitter dos candidatos foram encontrados. Ambler e Roberts (2006) e Peters et al. (2013) afirmaram que não há métrica “bala de prata” no contexto de social media. Por fim, sabe-se que há pouco consenso sobre como o engajamento pode ser medido nesse contexto (SCHULTZ e PELTIER, 2013).

3. Metodologia

Diferente dos estudos relacionados, que utilizaram menções em *tweets* de usuários aleatórios para realizar previsão, nesse estudo foram trabalhados os perfis dos candidatos e seus *tweets* do período eleitoral. O estudo foi realizado desta forma, pois entende-se que a intolerância política existente no país no contexto das eleições presidenciais poderia influenciar negativamente a pesquisa, levando a falsos resultados. Foram fatores levados em conta: (1) foi suposto que a intolerância faria pessoas deixarem de expressar seu apoio político explicitamente (falando, em *tweets*) e passarem a expressá-lo implicitamente (seguindo, curtindo e *retweetando*); (2) foi suposto que a análise de menções levaria a muitos *tweets* que não expressam apoio, e sim depreciação do candidato.

Os candidatos selecionados para este estudo foram os 5 melhores colocados na pesquisa de intenção de votos do instituto Ibope realizada em 20 de agosto de 2018 (substituindo posteriormente Lula por Fernando Haddad em função da proibição de Lula ser candidato). Com o objetivo de determinar a popularidade dos candidatos à Presidência, foram realizadas as etapas de coleta de dados e preparação de dados. A coleta, que se baseou nos scripts da obra de Bonzanini (2016), ocorreu entre 7 de setembro e 28 de outubro de 2018, e a preparação dos dados foi feita concomitantemente e finalizada logo após o fim do experimento.

Na primeira etapa do trabalho, a coleta de dados, foram elaborados 2 scripts em Python que, através da biblioteca Tweepy, se comunicam com a API do Twitter, como mostra a Figura 1. O script 1 obtém o perfil do candidato e o script 2 obtém a *timeline* do candidato. Cada um destes 2 scripts, quando executado, salva no disco local 5 arquivos JSON (1 para cada candidato).

A estrutura do JSON, apresentada na Figura 1, é similar à de dicionários da linguagem Python: todo o corpo do documento fica entre duas chaves e todos os dados dentro destes símbolos são pares *key-value* (chave-valor), sendo que as chaves são Strings e o valor correspondente pode ser de diversos tipos: inteiro, *float*, *booleano*, *String* e até vetor.

Do JSON perfil, foi relevante para nossa análise o atributo *followers_count*, que representa o número de seguidores no momento da execução. A execução ocorreu durante todos os dias do experimento às 00:00, para uma medida precisa de qual foi o saldo de seguidores no dia anterior. Então cada um dos cinco candidatos teve 52 arquivos JSON que representavam seu perfil no Twitter, totalizando 260 JSONs.

No script que obtém a *timeline* (conjunto de *tweets* do dia), foi necessário inserir o período desejado, que no caso deste estudo, era o período de apenas um dia. O JSON recebido da *Application Programming Interface* (API) continha todos os *tweets* do dia, cada um como um objeto dentro do JSON. Desses arquivos de *timeline* foram relevantes para a análise os atributos *favorite_count*, *retweet_count* e *metadata*, que ficavam dentro de cada *tweet*. *Metadata* foi importante para calcular o número de *tweets* do dia por que todos os *tweets* possuem apenas um atributo com este nome.

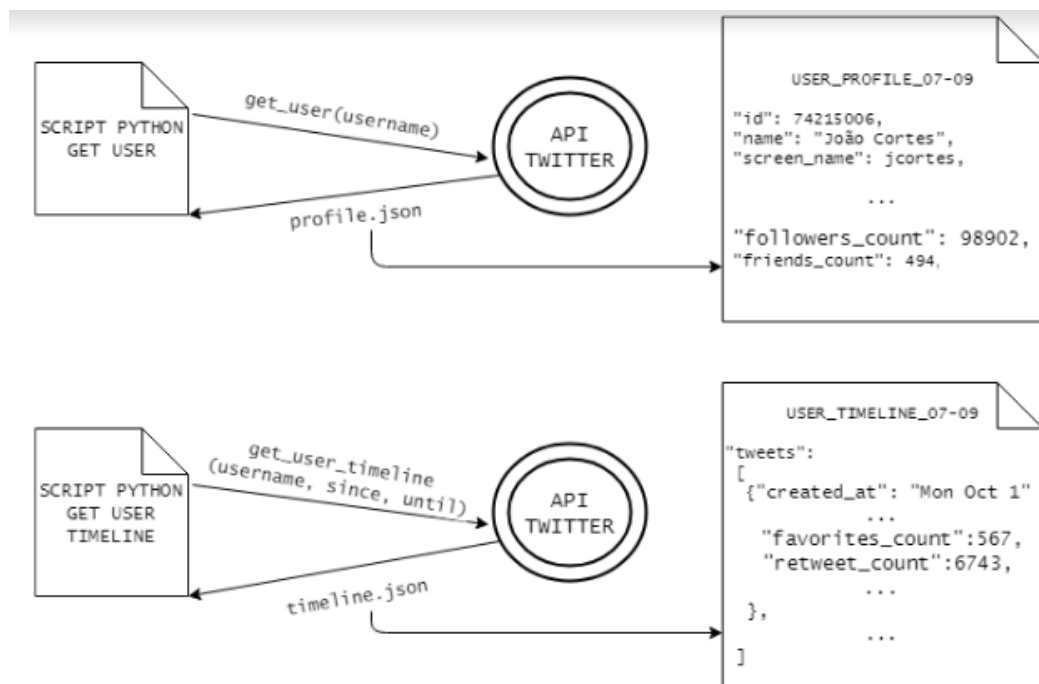


Figura 1. Coleta de dados da API

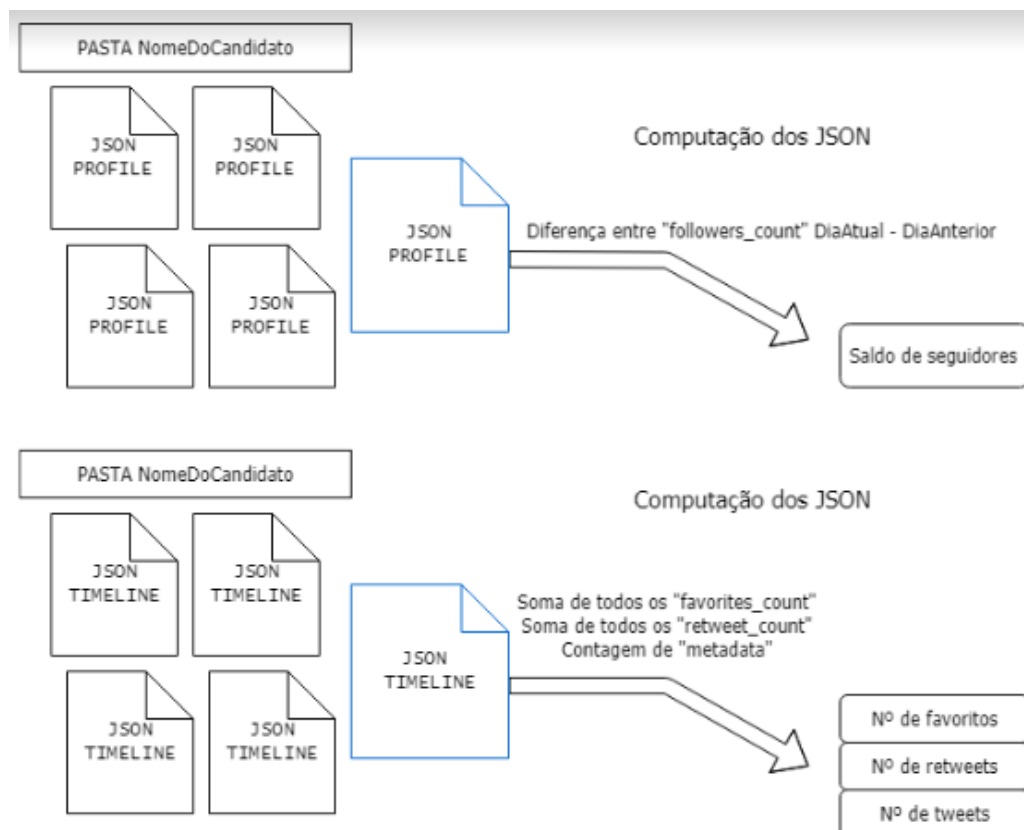


Figura 2. Computação de dados coletados

Para essa coleta, foi considerado que um *tweet* precisa de algum tempo para ter sua devida repercussão. Por essa razão, a coleta de *timeline* ocorreu 5 dias após a publicação da mesma. Por exemplo, os *tweets* de 7 de setembro foram coletados no dia 12 de setembro, para que no dia 12 já tivessem um certo nível de repercussão e o número de curtidas e *retweets* já houvesse parado de crescer significativamente.

A segunda etapa do trabalho envolveu a computação e preparação dos dados coletados (JSONs) e está representada na Figura 2. A tarefa também foi dividida em 2 scripts Python, um para tratar os JSONs perfil e outro para tratar os JSONs *timeline*. O primeiro calculou o saldo de *tweets* de cada dia de coleta e forneceu como saída uma lista contendo o saldo de seguidores por dia, por candidato. Um exemplo de saída para um candidato no quinto dia de pesquisa poderia ser:

candidato: [100, 54, 67, 131, 108]

o que indicaria que o saldo foi de 100 seguidores no primeiro dia, 54 no segundo e assim por diante. O segundo script calculou o número total de *tweets*, curtidas e *retweets* por dia, por candidato. Por exemplo, um candidato poderia ter a seguinte saída no terceiro dia de pesquisa:

candidato: [21, 43, 78]

[321, 432, 239]

[543, 678, 431]

Neste exemplo, as listas apresentam para cada dia, respectivamente, número de *tweets*, curtidas totais e *retweets* totais.

4. Resultados

4.1 Análise de dados do primeiro turno

Esta análise considerou os dados coletados de 7 de setembro às 00:00 até 7 de outubro às 00:00. Nas Tabelas 1, 2 e 3 as respectivas previsões (coluna “Twitter”) são comparadas

com a pesquisa Boca de urna (7 de outubro) do Instituto Ibope e com o resultado, divulgado no mesmo dia. Há um fator a considerar nesta comparação, como mostra a última linha das tabelas: o total de votos válidos na pesquisa boca de urna e no resultado final mostrados aqui não levam em conta os outros 8 candidatos, por isso destes valores foram normalizados para somarem 100%. Desta forma, torna-se possível a comparação entre todos estes contextos.

Observa-se na Tabela 1 que Marina Silva e Geraldo Alckmin tiveram seu número de seguidores aumentado em menos de 20 mil ao longo do período da coleta no 1º turno, sendo 13 mil para Geraldo Alckmin e 17 mil para Marina Silva. Já Fernando Haddad aumentou seu número em 71 mil, ficando atrás de Ciro Gomes, que aumentou 149 mil, seguido por Jair Bolsonaro, com quase 259 mil seguidores a mais.

Tabela 1: Previsão do 1º turno em relação ao saldo de seguidores

Usuário	Aumento de seguidores	Contexto		
		Twitter	Boca de urna	Resultado
JairBolsonaro	258.845	50,8%	48,4%	49,2%
Cirogomes	149.061	29,2%	15,1%	13,4%
Haddad_fernando	71.168	14%	30,1%	31,3%
Marinasilva	17.382	3,4%	2,1%	1,1%
Geraldoalckmin	13.353	2,6%	4,3%	5,0%
Total		100%	100%	100%

Na Tabela 2, é possível ver que a colocação da previsão anterior se manteve: Jair Bolsonaro teve cerca de 6.2 milhões de curtidas em seus *tweets*, dando novamente o segundo lugar a Ciro Gomes, que teve 2,5 milhões no mesmo critério. Em terceiro lugar ficou Fernando Haddad, com 1 milhão de curtidas, e nos últimos lugares, Marina Silva, com 457 mil e Geraldo Alckmin com 190 mil.

Tabela 2: Previsão do 1º turno em relação ao total de curtidas

Usuário	Total de curtidas	Contexto		
		Twitter	Boca de urna	Resultado
Jairbolsonaro	6.259.803	59,7%	48,4%	49,2%
Ciorgomes	2.582.958	24,6%	30,1%	31,3%
Haddad_fernando	1.001.204	9,6%	15,1%	13,4%
Marinasilva	457.074	4,3%	2,1%	1,1%
Geraldoalckmin	190.221	1,8%	4,3%	5,0%
Total		100%	100%	100%

Na Tabela 3, que representa a última previsão, em relação ao total de *retweets*, mais uma vez a colocação se manteve a mesma. Jair Bolsonaro ficou à frente com 1.5 milhões de *retweets*, seguido por Ciro Gomes, com 558 mil, Fernando Haddad, com 305 mil, Marina Silva, com quase 111 mil e Geraldo Alckmin, com 57 mil.

Tabela 3: Previsão do 1º turno em relação ao total de retweets

Usuário	Total de retweets	Contexto		
		Twitter	Boca de urna	Resultado
Jairbolsonaro	1.584.842	60,6%	48,4%	49,2%
Ciorgomes	557.776	21,3%	30,1%	31,3%
Haddad_fernando	305.507	11,7%	15,1%	13,4%
Marinasilva	110.702	4,2%	2,1%	1,1%
Geraldoalckmin	57.384	2,2%	4,3%	5,0%
Total		100%	100%	100%

4.2. Análise de dados do segundo turno

As previsões do segundo turno consideraram os dados coletados de 8 de outubro às 00:00 até 28 de outubro às 00:00. Nas Tabelas 4, 5 e 6, as respectivas previsões (coluna “Twitter”) são comparadas com a pesquisa Boca de urna (28 de outubro) do Instituto

Ibope e com o resultado, divulgado também no dia 28 de outubro.

Na Tabela 4, que representa a previsão em relação ao aumento de seguidores, pode ser observado que Jair Bolsonaro teve o número de seguidores aumentado em 349 mil, ou 35% a mais que no primeiro turno. Fernando Haddad teve o número de seguidores aumentado em 237 mil, ou 233% a mais que no primeiro turno. O erro da previsão (Twitter) foi de 4,9% em relação ao resultado oficial.

Tabela 4: Previsão do 2º turno em relação ao saldo de seguidores

Usuário	Aumento de seguidores	Contexto		
		Twitter	Boca de urna	Resultado
Jairbolsonaro	349.597	60%	56%	55,1%
Haddad_fernando	237.060	40%	44%	44,9%
Total		100%	100%	100%

Na Tabela 5, que representa a previsão em relação ao total de curtidas, é possível ver que a colocação se manteve a mesma, com Jair Bolsonaro à frente. Em relação ao 1º turno, Jair Bolsonaro teve 91% curtidas a mais, enquanto Fernando Haddad teve 714% curtidas a mais. O erro da previsão (Twitter) foi de 4,4% em relação ao resultado oficial.

Tabela 5: Previsão do 2º turno em relação ao total de curtidas

Usuário	Total de curtidas	Colocação		
		Twitter	Boca de urna	Resultado
Jairbolsonaro	11.964.327	59,5%	56%	55,1%
Haddad_fernando	8.156.316	40,5%	44%	44,9%
Total		100%	100%	100%

Na Tabela 6, que representa a previsão em relação ao total de *retweets*, é possível ver que novamente Fernando Haddad ficou em segundo lugar. Em relação ao 1º turno, ele

teve 633% *retweets* a mais, enquanto Bolsonaro teve 104% *retweets* a mais. O erro da previsão (Twitter) foi de 4% em relação ao resultado oficial.

Tabela 6: Previsão do 2º turno em relação ao total de *retweets*

Usuário	Total de <i>retweets</i>	Colocação		
		Twitter	Boca de urna	Resultado
Jairbolsonaro	3.240.323	59,1%	56%	55,1%
Haddad_fernando	2.242.332	40,9%	44%	44,9%
Total		100%	100%	100%

4.3 Visualização dos atributos ao longo do período de coleta

A seguir são apresentados 3 gráficos que permitem visualizar o comportamento dos atributos (saldo de seguidores por dia, total de curtidas por dia e total de *retweets* por dia) ao longo de toda a pesquisa, de 7 de setembro a 28 de outubro.

O Gráfico 1 mostra o comportamento do atributo *followers_count*, ou contagem de seguidores. Esse atributo é independente da atividade do usuário, isto é, alguns usuários ficam sem postar *tweets* por longos períodos e ainda assim podem ter variação no número de seguidores.

O Gráfico 2 representa o comportamento do atributo *favorite_count*, ou contador de curtidas. Esse atributo, ao contrário do analisado anteriormente, depende, sim, da atividade do candidato. Se ele não tiver postado nenhum *tweet* no dia, seu contador de curtidas fica em 0.

Similarmente ao Gráfico 2, o Gráfico 3 representa um atributo dependente de atividade do usuário: *retweet_count*, o contador de *retweets*. Ao contrário dos gráficos anteriores, nos quais os picos foram atribuídos a Jair Bolsonaro, em contagem de *retweets* o ponto mais alto foi de Fernando Haddad, passando de 300 mil *retweets* em um só dia em 27/10 (dia anterior às eleições do segundo turno). Outros dados interessantes podem ser vistos com esta análise:

- Jair Bolsonaro liderou nos 3 atributos (seguidores, *retweets* e curtidas) na maioria dos dias cobertos pela pesquisa.
- Ciro Gomes ocupou o segundo lugar em praticamente todo o primeiro turno, mas após o resultado que o tirou da disputa, sua popularidade caiu, enquanto a de Fernando Haddad aumentou. Nos 3 gráficos observa-se nitidamente a inversão entre os dois candidatos, que ocorreu exatamente no dia 8 de outubro (dia seguinte à votação e ao resultado).
- Nos dias imediatamente após o resultado do primeiro turno, os dois candidatos selecionados tiveram picos de popularidade, ganhando mais seguidores, curtidas e *retweets*.
- Os números de Fernando Haddad no primeiro turno eram muito baixos em relação a Jair Bolsonaro, mas no segundo turno, Haddad cresceu muito, chegando a alcançar Bolsonaro em alguns dias e superá-lo em outros poucos.
- Marina Silva e Geraldo Alckmin ficaram com as pontuações mais baixas, ocupando sempre as duas últimas colocações.

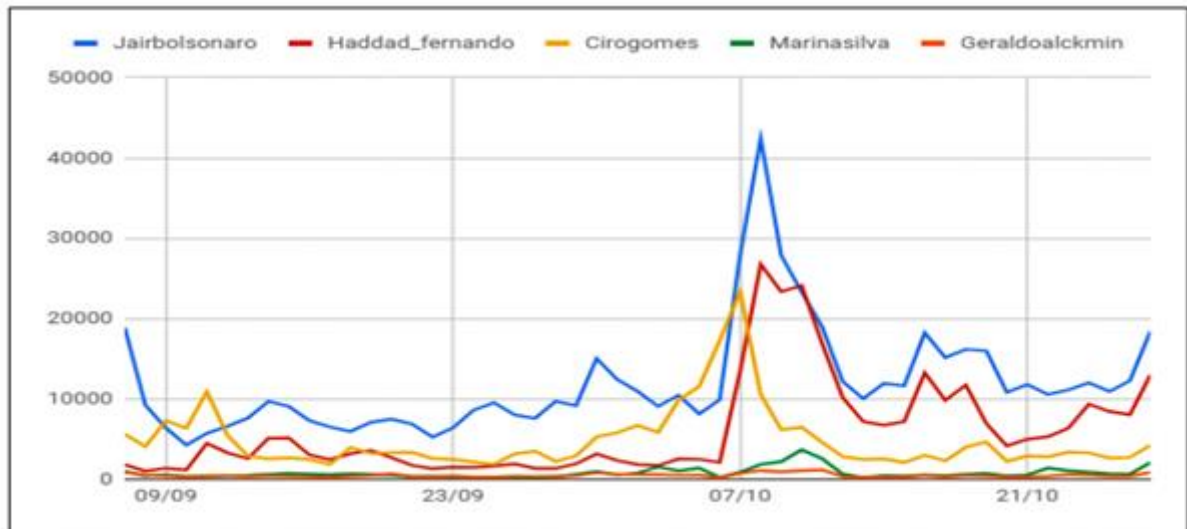


Gráfico 1: Aumento (saldo) de seguidores por dia

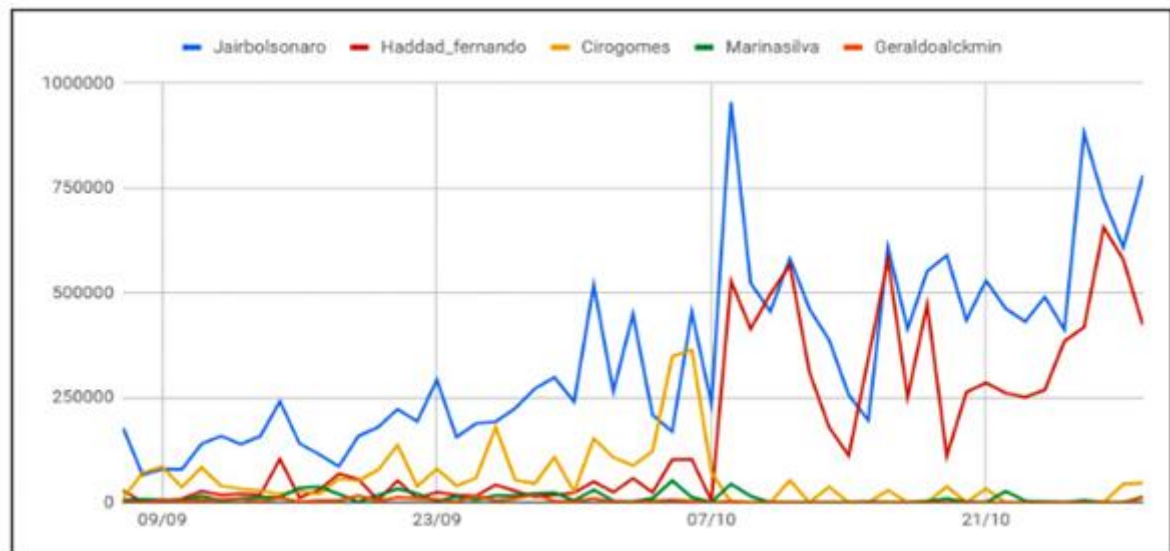


Gráfico 2: Total de curtidas em tweets autorais por dia

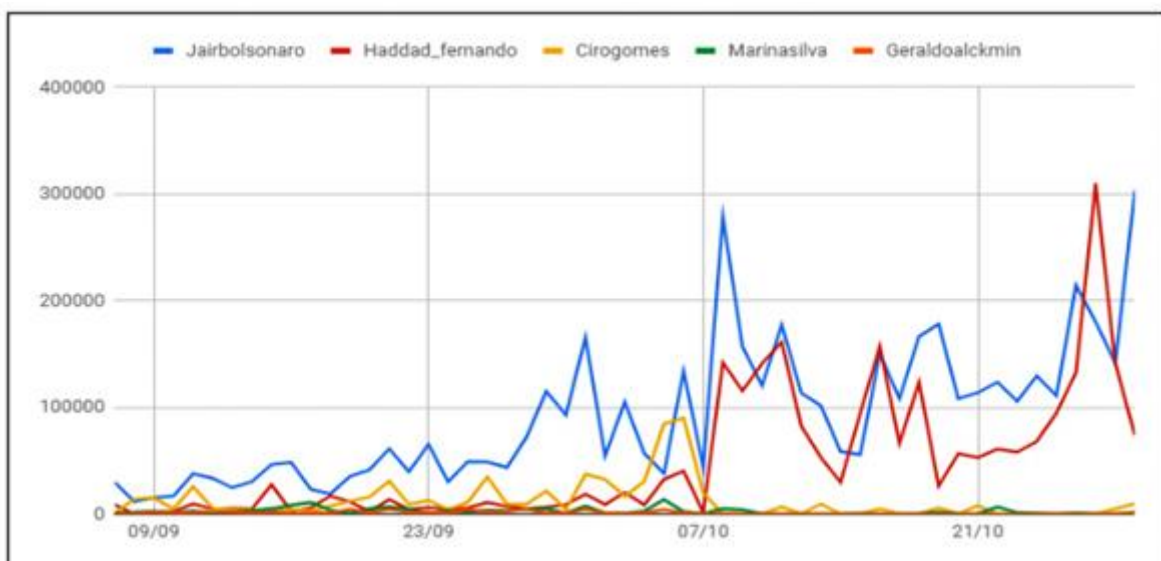


Gráfico 3: Total de retweets em tweets autorais por dia

Os resultados da coleta de dados mostram a vantagem de Jair Bolsonaro sobre os outros candidatos em relação à popularidade no Twitter. Esse fato corresponde à realidade eleitoral, pois a maioria dos brasileiros o elegeu.

Um resultado interessante foi a inversão entre a popularidade de Ciro Gomes e Fernando Haddad. Os usuários do Twitter que contribuíram para a pesquisa tinham uma preferência clara pelo candidato Ciro Gomes no 1º turno, como ficou evidenciado nas previsões da seção anterior e nos gráficos. No entanto, quando se depararam com o resultado do primeiro turno, no qual seu candidato ficou para trás, passaram a dar muita atenção a Fernando Haddad.

Marina Silva, tanto nas previsões como nos gráficos, superou a popularidade de Geraldo Alckmin em praticamente todos os dias de pesquisa. Isso não é explicado pelo resultado eleitoral, no qual ela obteve 1% e Alckmin, quase 5%. Este ocorrido pode estar relacionado ao fato de Marina Silva já ter consolidada uma maior popularidade no período anterior às eleições: em seguidores, no dia 7 de setembro Marina tinha 1,88 milhão, Bolsonaro tinha 1,35 milhão, Alckmin 978 mil, Haddad 641 mil e Ciro tinha 236 mil.

Druckman et al. (2009) declaram que à medida que a corrida eleitoral se acirra, os candidatos ficam mais inclinados a melhorar seus sites com características como informações atualizadas, que tornam seus sites mais interessantes. Um comportamento similar à essa afirmação foi observado na medida de assiduidade ou engajamento do candidato no Twitter. A atividade dos candidatos, ou quantidade de Tweets por dia (Gráfico 4), mostra que os mais assíduos do primeiro turno, ou os que mais postaram mensagens aos seguidores, foram Ciro Gomes e Geraldo Alckmin. O único a passar de 100 tweets em um dia foi Ciro, que no dia

5 de outubro, inclusive, chegou a publicar 165 mensagens.

Os dois líderes em campanha no Twitter, entretanto, “desapareceram” da rede após o resultado do primeiro turno, que mostrou que ambos estavam fora da disputa presidencial, juntamente com Marina Silva. Apenas Haddad e Bolsonaro continuaram a ter atividade significativa na rede social. Essa ausência dos outros candidatos deixa a impressão de que sua comunicação e sua tentativa de chegar aos eleitores tinha um interesse bem delineado, e uma vez que não seria mais possível alcançá-lo, não mais era necessário estar tão perto do eleitor. Isso nem sempre ocorre, como Marques et al. (2011) constatou: nas eleições de 2010, mesmo fora do segundo turno, Marina Silva utilizou mais o Twitter do que a candidata Dilma Rousseff, que continuava na disputa e acabou por vencer.

A Tabela 7 mostra que os candidatos tiveram campanhas ativas no Twitter, utilizando intensamente a rede durante o período eleitoral. Os usuários do eleitorado responderam também ativamente a essa campanha, mostrando interesse pelas declarações de um candidato ao segui-lo, e mostrando preferência ou, no mínimo, concordância com um candidato ao curtir ou *retweetar*.

Mesmo a coleta do segundo turno tendo 10 dias a menos que a do primeiro, quando comparados os dois períodos, os números de Haddad e Bolsonaro foram muito maiores no segundo turno. Supõe-se que devido à disputa mais direta e mais acirrada entre candidatos de frentes opostas no segundo turno, os apoiadores de cada um tentaram com mais afinco demonstrar apoio e propagar as ideias de seu escolhido.

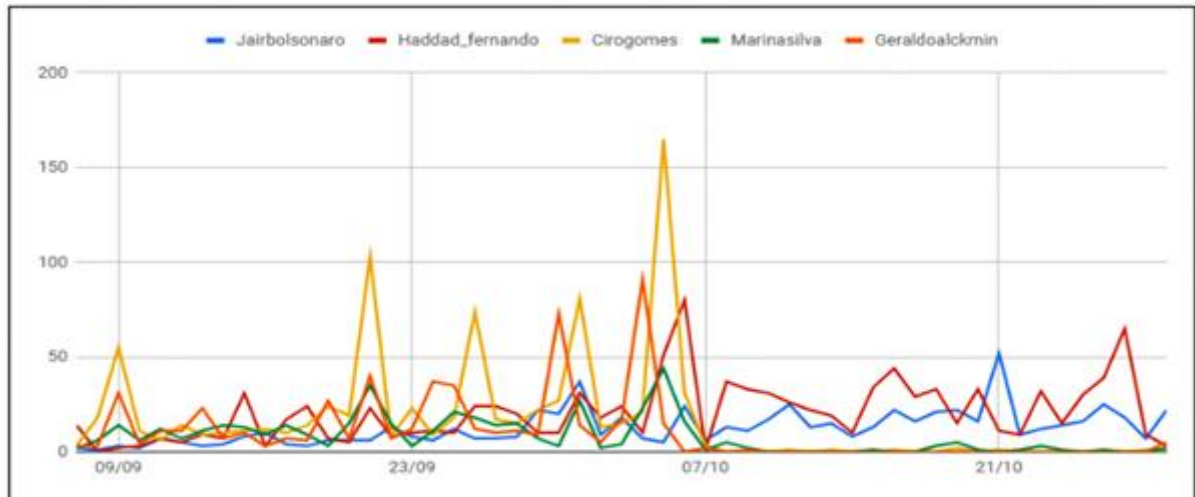


Gráfico 4: Total de tweets por dia

Tabela 7: Comparação de dados entre primeiro e segundo turnos

Conta/ Atributo	1º turno (1º ao 31º dia de pesquisa)			2º turno (32º ao 52º dia de pesquisa)		
	Saldo de seguidores	Total de curtidas	Total de <i>retweets</i>	Saldo de seguidores	Total de curtidas	Total de <i>retweets</i>
Jairbolsonaro	258.845	6.259.803	1.584.842	349.597	11.964.327	3.240.323
Cirogomes	149.061	2.582.958	557.776	99.142	351.426	69.493
Haddad_fernando	71.168	1.001.204	305.507	237.060	8.156.316	2.242.332
Marinasilva	17.382	457.074	110.702	22.733	125.870	22.578
Geraldoalckmin	13.353	190.221	57.384	11.672	15.433	1.388

5. Conclusões e trabalhos futuros

Embora alguns pesquisadores tenham questionado a validade da massa de dados de redes sociais, sugerindo que o conteúdo produzido nessas aplicações são majoritariamente focados em entretenimento e expressão emocional (JAVA et al., 2007; NAAMAN et al., 2010), e de acordo com DIGRAZIA et al. (2013) muitos estudos sugerem que as redes sociais fornecem uma amostra tendenciosa e não-representativa da população, esse estudo mostrou um grande potencial do Twitter em prever as intenções de voto dos brasileiros.

Com a campanha mais barata entre os cinco candidatos segundo dados divulgados pelo Tribunal Superior Eleitoral (TSE) em 2018, Jair Bolsonaro conquistou no segundo turno a vitória eleitoral. Tendo tido como marco de sua campanha uma utilização intensa das redes sociais, o candidato mostrou o grande potencial das mesmas, em contraposição ao uso da mídia tradicional.

O Twitter proporcionou uma boa detecção da popularidade dos usuários estudados. Assim, mostrou um alto poder preditivo em relação ao resultado eleitoral. Nas previsões do primeiro turno, embora o Twitter tenha acertado a colocação de apenas um candidato, gerou um esboço aproximado do cenário real: O primeiro candidato seria Bolsonaro, seguido por Ciro e Haddad, independente da ordem, seguidos por Alckmin e Marina, independente da ordem.

Quanto ao segundo turno, os dados usados do Twitter não apenas acertaram a colocação, como também se aproximaram satisfatoriamente da porcentagem real, com a taxa de erro variando entre 4% e 5%, conforme visto nas previsões do 2º turno. A pesquisa Boca de urna teve a taxa de erro de 1%.

Os resultados deste estudo indicam que os candidatos mais relevantes no Twitter podem ter maior intenção de voto e a popularidade

dos candidatos na rede pode gerar uma boa previsão sobre o resultado eleitoral. Jair Bolsonaro manteve absoluta liderança, ficando em segundo lugar muito eventualmente, e foi o candidato mais votado tanto no primeiro quanto no segundo turno. Ciro e Haddad, que alternaram entre a vice-liderança durante a pesquisa, acabaram em segundo e terceiro lugar. Por outro lado, Marina e Alckmin, que foram relativamente pouco populares, acabaram de fato tendo menor intenção de voto. Marina foi mais relevante que Geraldo no Twitter, mas acabou atrás dele no resultado eleitoral. Isso pode ser explicado pelo fato de a candidata ter um perfil mais consolidado no Twitter, por que no início da pesquisa ela era a candidata com mais seguidores entre os 5.

Para futuras pesquisas, poderia ser realizada uma análise de conteúdo dos *tweets*, na tentativa de relacionar popularidade/repercussão do *tweet* a palavras específicas, ou analisar do que cada candidato mais fala. Também poderiam ser trabalhadas as menções aos candidatos, para saber a que palavras os usuários mais relacionam cada um deles.

REFERÊNCIAS

- AL-KHALIFA, H. (2012), "A first step towards understanding Saudi political activities on Twitter". *International Journal of Web Information Systems*, Vol. 8 Issue: 4, pp.390-400.
- AMBLER, T. e ROBERTS, J. (2006), "Beware the silver metrics: marketing performance measurement has to be multidimensional". *Marketing Science Institute*.
- BODE, L. et al. (2016), "Coherent campaigns? Campaign broadcast and social messaging". *Online Information Review*, Vol. 40, n.5, pp.580-594.

- BONZANINI, Marco. *Mastering Social Media Mining with Python*. 1. ed. Packt Publishing, 2016.
- DIGRAZIA, J. et al. (2013), “More Tweets, More Votes: Social Media as a Quantitative Indicator of Political Behavior”. *PLoS One* v.8, n.11.
- DRUCKMAN, J. et al. (2009), “The Technological Development of Candidate Web Sites: How and Why Candidates Use Web Innovations”. *Social Science Computer Review*, v.25 n.4, pp. 425-442.
- G1. Pesquisa Ibope de 20 de agosto para presidente por sexo, idade, escolaridade, renda, região, religião e raça. Disponível em <<https://g1.globo.com/politica/eleicoes/2018/eleicao-em-numeros/noticia/2018/08/21/pesquisa-ibope-de-20-de-agosto-para-presidente-por-sexo-idade-escolaridade-renda-regiao-religiao-e-raca.ghtml>>. Acesso em 1 de novembro de 2018.
- GAURAV, M. et al. (2013), “Leveraging Candidate Popularity On Twitter To Predict Election Outcome”. In: *SNAKDD '13 Proceedings of the 7th Workshop on Social Network Mining and Analysis*. Chicago, Illinois.
- GAYO-AVELLO, D. (2012), “A Balanced Survey on Election Prediction using Twitter Data”. *CoRR*, abs/1204.6441.
- JAVA, A. et al. (2007), “Why we twitter: understanding microblogging usage and communities”. In: *WebKDD/SNA-KDD '07 Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pp.56-65.
- MARQUES, F. et al. (2011), “Estratégias de comunicação política online: uma análise do perfil de José Serra no Twitter”. *Contemporanea – Revista de Comunicação e Cultura*, v. 9, n.3.
- MISSINGHAM, R. (2010), “The Australian Parliament in the Twittersverse”. *Australasian Parliamentary Review*, v. 25(1), pp. 3–16.
- NAAMAN, M. et al. (2010), “Is it really about me? Message content in social awareness streams”. In: *CSCW '10 Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pp. 189-192
- PETERS, K. et al. (2013), “Social media metrics—a framework and guidelines for managing social media”. *Journal of Interactive Marketing*, v. 27, n.4, pp. 281-298.
- SANG, E. e BOS, J. (2012), “Predicting the 2011 Dutch Senate Election Results with Twitter”. In: *EACL '12 Proceedings of the Workshop on Semantic Analysis in Social Media*, pp. 53-60.
- SAVAGE, N. (2011), “Twitter as medium and message”. *Communications of the ACM*, v.54, n.3, pp.18-20
- SCHULTZ, D.E. e PELTIER, J. (2013), “Social media’s slippery slope: challenges, opportunities and future research directions”. *Journal of Research in Interactive Marketing*, v. 7, n. 2, pp.86-99.
- STATISTA (2019). Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019 (in millions). Disponível em <<https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users>>. Acesso em 23 de maio de 2019.
- STATISTA (2016). Number of social media users worldwide from 2010 to 2021 (billions). Disponível em <<https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>>. Acesso em 23 de maio de 2019.
- TWIPLOMACY 2018. Disponível em <<https://twiplomacy.com/blog/twiplomacy-study-2018/>>.