

Leveraging the Facebook Ads Platform for Election Polling

Filipe N. Ribeiro
UFOP/UFMG, Brasil
filipe.ribeiro@ufop.edu.br

Daniel Kansaon
UFMG, Brasil
daniel.kansaon@dcc.ufmg.br

Fabício Benevenuto
UFMG, Brasil
fabricao@dcc.ufmg.br

ABSTRACT

Election polls provide valuable insights about voting intention on candidates filtered by demographic characteristics. This information can be used to understand the dynamics of underlying electoral preferences and how it changes in the days or months during the campaign. Despite its importance, election polls are time and money consuming on many occasions. Especially those with face-to-face surveys, that interview a representative population of the voters across the entire country. In this study, we propose a novel approach that explores social media advertising platforms to infer the audience demographics of politicians in the electoral race. We leverage the attribute-based targeting available on the Facebook Advertising platform by using the 'interests' related to candidates to calculate their audience demographics. Then, we compared the online extracted data with election polls taken in the same period. Our findings suggest that the candidate's popularity on Facebook, captured in terms of the number of likes, people talking about him/her, and the number of people interested in the candidate, is a valid indicator of his/her variation on vote intention polls. Additionally, we figured out that the fluctuation in the demographic aspects of supporters detected by election polls are captured by our methodology, obtaining more precision with popular candidates. In particular, we show that sharp variations occasioned by high impact events during the campaign such as protests are well captured by Facebook measurements. Finally, we deployed a system that exposes the audience demographics of Brazilian politicians on Facebook (available at <http://www.audiencia-dos-politicos.dcc.ufmg.br/>) and contributes to the understanding of the political scene in Brazil.

CCS CONCEPTS

• Information systems → Social advertising; • Human-centered computing → Social media; Social network analysis.

KEYWORDS

elections, election polls, advertising, Facebook Ads

ACM Reference Format:

Filipe N. Ribeiro, Daniel Kansaon, and Fabrício Benevenuto. 2019. Leveraging the Facebook Ads Platform for Election Polling. In *Brazilian Symposium on Multimedia and the Web (WebMedia '19)*, October 29–November 1, 2019, Rio de Janeiro, Brazil. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3323503.3349552>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WebMedia '19, October 29–November 1, 2019, Rio de Janeiro, Brazil

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6763-9/19/10...\$15.00

<https://doi.org/10.1145/3323503.3349552>

1 INTRODUÇÃO

O crescimento do número de usuários de Redes Sociais Online nos últimos anos é impressionante. Apenas o Facebook, a rede mais popular, atraiu mais de 500 milhões de novos usuários nos últimos dois anos, alcançando 2,23 bilhões de usuários ativos mensais em 30 de junho de 2018¹. O Facebook, bem como Snapchat, LinkedIn, Twitter e outras redes sociais, representam uma parte importante da vida das pessoas. Uma pesquisa recente indica que dois terços dos adultos dos EUA usam redes sociais, dos quais 42% dizem que seria difícil abandonar a mídia social [24]. A mesma pesquisa mostra que 51% dos usuários do Facebook nos EUA, acessam suas contas várias vezes por dia. Tal cenário impacta diretamente a sociedade em vários aspectos, como por exemplo, na maneira como as pessoas acessam notícias [12]. Os feeds de notícias das redes sociais têm substituído outras fontes de acesso às informações cotidianas como TVs, jornais e revistas. Nos EUA, por exemplo, 62% dos adultos consomem notícias primariamente em sites de mídia social [16].

As Redes Sociais Online permitiram que os usuários participassem ativamente de debates públicos e expusessem suas opiniões publicamente. Não surpreendentemente, vários trabalhos exploraram o conteúdo produzido pela grande massa de usuários online em diferentes cenários, como a detecção de terremotos [21], identificação de áreas com maior incidência de dengue [9] e detecção de epidemias de gripe [14]. Os dados coletados nessas redes online também foram examinados no contexto eleitoral para prever eleições [2, 27], pesquisas eleitorais [28] e para detectar o viés ideológico de usuários [11, 26].

Em particular, o uso de dados das Redes Sociais Online no contexto eleitoral tornou-se popular por muitas razões. Em primeiro lugar, o custo para capturar dados é baixo/ausente quando comparado com as técnicas tradicionais para realizar pesquisas eleitorais ou detectar a inclinação política das pessoas. Além disso, o ambiente de mídia social atrai um grande número de pessoas, o que, em muitos casos, representa uma amostra importante da população offline. É notório que a população online é enviesada para os jovens e a amostra não é bem representativa de todas as categorias demográficas [17]. No entanto, o monitoramento de dados desse nicho de pessoas pode ser altamente elucidativo, uma vez que as pessoas influenciam e são influenciadas pelas discussões e debates nas redes, o que pode impactar nos resultados das eleições.

Neste trabalho, é apresentada uma nova técnica que explora informações obtidas através das plataformas de publicidade das Redes Sociais Online fornecidas aos anunciantes. Mais especificamente a abordagem faz uso da fórmula de segmentação (conjunto de atributos) empregado na plataforma de publicidade do Facebook para definir o público-alvo que um anunciante deseja atingir. Os interesses relacionados aos candidatos são utilizados para selecionar seu público no Facebook e calcular as respectivas características

¹<https://newsroom.fb.com/> Informação da companhia/

demográficas. Em seguida, empregou-se uma metodologia para realizar pesquisas eleitorais com base na audiência dos candidatos no Facebook.

Como forma de validação da abordagem proposta, os resultados foram comparados com pesquisas eleitorais de importantes institutos de pesquisa no Brasil. As descobertas sugerem que a abordagem proposta para pesquisas eleitorais com base na população online do Facebook capta tendências semelhantes às capturadas por pesquisas realizadas na população offline. Adicionalmente, foi construído um sistema que detalha a audiência demográfica dos interessados nos candidatos políticos envolvidos nas eleições brasileiras de 2018.

2 TRABALHOS RELACIONADOS

Os trabalhos relacionados são apresentados em duas etapas: 1 - aqueles que tinham como objetivo prever eleições e pesquisas eleitorais e 2 - estudos que usaram plataformas de publicidade de Redes Sociais Online para inferir dados demográficos.

2.1 Predição de eleições e pesquisas eleitorais

Em um dos trabalhos mais controversos, os pesquisadores argumentaram que o Twitter poderia prever os resultados das eleições [27]. Os autores coletaram *tweets* que mencionavam partidos e/ou candidatos no mês anterior à eleição nacional alemã, inferindo o ranking de candidatos/partidos baseado apenas no volume de *tweets*.

Jungherr *et al.* [10] refutou os resultados que indicavam a correlação entre o total de *tweets* e os resultados da eleição. Os autores apontaram uma série de falhas metodológicas, incluindo a falta de informações sobre a coleta e outros detalhes do método que impedem outros pesquisadores de reproduzir o experimento. Além disso, realizaram uma coleta semelhante e concluíram que o verdadeiro vencedor da eleição seria o Partido Pirata (Piraten), cujos *tweets* foram excluídos da análise original sem explicação efetiva do motivo. Outros pesquisadores tentaram reproduzir o método que contava os *tweets* para prever eleições, e isso funcionou parcialmente para a eleição de Cingapura [23]. No geral, o método provou ser falho, como na eleição especial do Senado dos EUA em 2010 em Massachusetts [3] e nas eleições holandesas de 2011 [22].

Williams e Gulati [29] analisaram a eleição presidencial dos EUA em 2008 e descobriram que o número de apoiadores no Facebook, em geral, indica sucesso eleitoral. Por outro lado, Giglietto [8] tentou prever os resultados das eleições em cidades italianas usando o número de curtidas dos candidatos e não obteve bons resultados.

Wang *et al.* [28] conduziu uma série de pesquisas diárias de intenção de votos através da plataforma de jogos do Xbox, audiência formada predominantemente por homens jovens. Eles então ajustaram as estimativas usando regressão linear, particionando a população em células que consideravam as regiões com maior prevalência de cada candidato e regiões com mais jogadores.

Gayo-Avello levantou muitas preocupações na predição de eleições baseada em dados online e elicitou um conjunto de recomendações para direcionar pesquisas desse tipo [6, 7]. Entre as preocupações, que também podem ser aplicadas a outros estudos que usam as redes sociais como fonte de dados, destaca-se a questão do viés de autosseleção. O autor enfatiza ainda que o viés demográfico nas mídias sociais, também inspecionado por Mislove *et al.* [15], é

frequentemente ignorado. Além disso, a população online normalmente não representa a população do mundo real e isso deve ser considerado nas análises.

Por outro lado, muitos estudos mostraram que alguns dados confiáveis sobre política podem ser extraídos das Redes Sociais Online. DiGrazia *et al.* [4] apontou que a distribuição de votos por candidatos teve uma alta correlação com a popularidade dos candidatos no Twitter em muitas eleições distritais. Birmingham e Smeaton [2] usaram regressão linear considerando o volume de *tweets* juntamente com um *score* da análise de sentimento para prever a eleição irlandesa de 2011. Outro estudo detectou que, nas eleições suecas de 2010, o debate online foi dominado por alguns grupos específicos de usuários, sendo os usuários mais ativos identificados principalmente como jornalistas, blogueiros consolidados ou políticos, embora os autores também tenham detectado alguns anônimos influentes [13].

Diferentemente da maioria dos estudos acima relacionados, que visam prever resultados eleitorais, o foco neste estudo é capturar pesquisas eleitorais através de dados provenientes das redes sociais. Definir a popularidade e a demografia do público dos candidatos nas redes sociais durante a campanha eleitoral é bastante elucidativo, pois, além da alta presença da população nessas plataformas, é também um lugar onde as pessoas se influenciam. Deve-se enfatizar que o objetivo não consiste em substituir as pesquisas eleitorais tradicionais. Pelo contrário, espera-se fornecer informações úteis, complementares às pesquisas offline, que certamente serão úteis para usuários online, entidades políticas, bem como pesquisadores, que podem explorar os dados para diversos propósitos.

2.2 Estudos que exploraram a plataforma de publicidade das Redes Sociais Online

Araújo *et al.* [1] usou a infraestrutura de publicidade do Facebook para identificar a ocorrência de doenças relacionadas ao estilo de vida em diferentes regiões. A ideia básica é verificar se a quantidade de pessoas online interessadas em doenças como diabetes e obesidade se correlaciona com a prevalência real das doenças no mundo offline.

Zagheni *et al.* [30] analisou o movimento de imigrantes nos EUA contando o número de residentes nascidos em 52 países segundo o Facebook e comparando com dados reais fornecido pelo Census Bureau. Eles encontraram uma alta correlação, mesmo considerando diferentes intervalos de idade e sexo.

Alguns estudos investigaram a relação entre a diferença de gênero verificada no Facebook e quatro índices de igualdade de gênero medidos pelo Fórum Econômico Mundial (World Economic Forum - WEF), que incluem oportunidades econômicas, participação política e educação. Foi observado que quanto maior a proporção de homens em relação às mulheres no Facebook, menores as oportunidades para mulheres de acordo com as métricas observadas pelo WEF para aquele país [5]. Por fim, a mesma metodologia foi empregada para inferir o viés político de milhares de meios de comunicação nos EUA [19] e detectar qual o público dos anúncios veiculados pela Agência de Pesquisa de Inteligência da Rússia (IRA) antes das eleições de 2016 nos EUA [20].

Até o momento não há nenhum estudo que explore este novo método para extrair informações agregadas sobre usuários de Redes Sociais Online e realizar pesquisas eleitorais. Ao explorar esta técnica, pretende-se inferir a demografia das audiências dos políticos

nas semanas anteriores à corrida eleitoral, verificando se a flutuação de popularidade verificada nas pesquisas eleitorais antes das eleições, é capturada pela perspectiva online através dos dados da plataforma de publicidade do Facebook.

3 METODOLOGIA

Uma das principais vantagens das plataformas de publicidade das Redes Sociais Online são as vastas possibilidades de alcançar usuários. Técnicas disponíveis para a seleção da audiência alvo passam pela utilização de uma lista de informações pessoalmente identificáveis (nome, número de telefone, e-mail, etc.) e pela utilização de atributos refinados, como raça, nível de renda, interesses e comportamentos [25].

A maioria dos estudos que exploram plataformas de publicidade, realizam uma simulação de um anúncio real em que se seleciona uma fórmula de segmentação (conjunto de atributos), podendo assim obter o número de usuários que se encaixam com a fórmula selecionada. Como exemplo de segmentação por atributos, um anunciante pode selecionar usuários que moram em um estado brasileiro qualquer e que estejam interessados em um candidato específico nas eleições presidenciais brasileiras. Antes de disparar o anúncio e da cobrança pela exibição da propaganda, a plataforma de publicidade exibe o alcance potencial, ou seja, o número de usuários que correspondem às características definidas pela fórmula, indicando o número de pessoas interessadas no candidato no estado em questão. Fórmulas semelhantes em que se varia o estado alvo podem ser criadas, permitindo o cálculo da distribuição do público do candidato em todos os estados brasileiros. Este mesmo método permite extrair muitos outros atributos demográficos da audiência do candidato tais como intervalo de idade, região geográfica e sexo.

A chave da metodologia proposta consiste nos elementos que representam as entidades nas quais os usuários estão interessados. Conhecidos como 'interesses', esses elementos são fornecidos pelas plataformas de publicidade das redes sociais para facilitar a criação de públicos diversificados pelos anunciantes. Os interesses representam entidades que vão desde alimentos (macarrão, pizza, etc.) a temas relacionados a religião (protestantismo, islamismo, etc.) e figuras públicas (artistas, mártires e políticos). O método aqui proposto utiliza o interesse definido para cada candidato para obter os usuários interessados nele/nela, criando cortes demográficos sobre suas respectivas audiências. Esses cortes demográficos são criados definindo várias fórmulas de segmentação específicas, exemplo: usuários masculinos que moram no Brasil, que possuem mais do que 18 anos e têm interesse no candidato A, ou, usuárias entre 18 e 35 anos, com interesse no mesmo candidato.

A plataforma de publicidade do Facebook foi escolhida como a fonte de dados por dois motivos principais: 1 - o número significativo de usuários ativos em todo o mundo, contando com mais de 130 milhões de usuários apenas no Brasil; 2 - a API (*Application Programming Interface*) bem documentada, permitindo aos desenvolvedores fazerem requisições automatizadas.

A fim de comparar a flutuação demográfica do público das mídias sociais com a intenção de voto capturada por pesquisas eleitorais semanas antes da eleição, foi coletada a distribuição de dois atributos demográficos da audiência de cada um dos candidatos: região geográfica e gênero. Como base de comparação, foram utilizadas pesquisas eleitorais de dois institutos de pesquisa pública de opinião

popular, o Datafolha² e o IBOPE³, ambos com resultados amplamente utilizados em grande parte da mídia para cobertura jornalística no Brasil. As pesquisas eleitorais fornecidas por esses institutos, estratificam as intenções de voto de cada candidato em segmentos demográficos, fornecendo dados semelhantes aos coletados da plataforma de anúncios.

As eleições presidenciais brasileiras de 2018 ocorreram em 7 de outubro (primeiro turno) e 28 de outubro (segundo turno), com 13 candidatos, dos quais selecionou-se os cinco candidatos mais bem posicionados nas pesquisas iniciais: Ciro Gomes, Fernando Haddad, Geraldo Alckmin, Marina Silva e Jair Bolsonaro. Vale destacar que o candidato Ciro Gomes, que manteve a terceira posição nas pesquisas durante quase todo o período anterior ao primeiro turno, não possuía interesse relacionado ao seu nome na plataforma de publicidade do Facebook na primeira coleta realizada em novembro de 2017.

A fim de fornecer uma visão geral da evolução da demografia da audiência dos candidatos no Facebook, realizou-se a primeira coleta em novembro de 2017, considerando a lista não oficial de candidatos para a disputa presidencial. Em julho e agosto de 2018, os dados foram coletados na segunda terça-feira do mês. A partir de setembro, foram conduzidas coletas semanais às segundas-feiras até o dia seguinte ao segundo turno. Também foram realizadas coletas nas sextas-feiras antes do primeiro e segundo turno.

Como base de comparação para o estudo foram utilizadas as pesquisas eleitorais publicadas pelos institutos de pesquisa de opinião nas semanas que antecederam as pesquisas eleitorais e em datas próximas às coletas nos meses que antecederam as eleições. Deve-se enfatizar que os institutos de pesquisa não publicam um cronograma de pesquisa. Esta é a razão pela qual as coletas do Facebook não foram conduzidas exatamente no mesmo dia da divulgação das pesquisas. Assim, os resultados são apresentados utilizando as datas de coleta do Facebook como base, agrupando as pesquisas eleitorais com a data de coleta mais próxima. Com exceção das duas primeiras datas, os resultados das pesquisas eleitorais utilizadas foram publicados com no máximo quatro dias de diferença dos dados extraídos online. As duas comparações iniciais contêm uma diferença próxima a vinte dias, o que não representa um grande impacto na análise, uma vez que, devido à distância das eleições a variação nesse período não foi significativa.

4 ELEIÇÕES SOB A ÓTICA DO FACEBOOK

Antes de apresentar os resultados das pesquisas medidos com a metodologia desenvolvida, é necessário (i) contextualizar o cenário político no Brasil, (ii) entender os eventos notáveis que ocorreram ao longo das eleições presidenciais, que possivelmente tiveram impacto na decisão dos eleitores, e (iii) discutir o envolvimento e penetração dos candidatos nas mídias sociais, especialmente no Facebook. Esta Seção é dedicada a cobrir esses três tópicos.

A eleição presidencial de 2018 no Brasil foi cercada por diversas discussões polêmicas e acaloradas, que sugeriam manipulação da mídia tradicional e perseguição aos candidatos. Houve até um ataque à vida de um candidato durante a corrida eleitoral. As datas desses eventos foram incluídas nas figuras de validação, a fim de exibir

²<http://datafolha.folha.uol.com.br/>

³<http://www.ibope.com.br>

seu impacto na campanha. Em primeiro lugar, o ex-presidente Lula, um dos presidentes mais populares da curta história da democracia no Brasil, foi preso após ser condenado por corrupção. Seguindo a lei eleitoral brasileira, ele foi julgado e impedido de prosseguir como candidato à presidência, embora fosse o candidato mais bem posicionado nas pesquisas eleitorais naquele momento. O candidato escolhido para substituir Lula nas eleições foi Fernando Haddad, ex-prefeito da maior cidade do Brasil, mas desconhecido pela população brasileira em regiões como norte e nordeste. A expectativa do Partido dos Trabalhadores (partido de Lula) era que os votos do ex-presidente preso fossem herdados por seu substituto.

Por outro lado, o candidato de direita Jair Bolsonaro, se apresentou como a alternativa ao corrompido sistema político brasileiro. Polêmico e com discursos inflamados, expandiu sua rede de apoio após intensificar sua presença nas redes sociais e com grupos organizados para difundir mensagens de apoio [18]. Acusado de ser misógino, ele foi o principal alvo de uma série de protestos em 29 de setembro, ocorridos em todo o país. Estes protestos nomeados como 'EleNao', foram propostos e coordenados por mulheres insatisfeitas com a sua crescente popularidade. Os dias subsequentes foram seguidos por protestos generalizados em resposta ao 'EleNao', organizado por partidários de Bolsonaro. Na semana que antecedeu o segundo turno, ocorreu outra onda de protestos em muitas cidades em todo o país. Em 20 de outubro, milhares de pessoas contra Bolsonaro tomaram as ruas para protestar. Mais uma vez, seus partidários também foram às ruas para apoiá-lo no dia seguinte. No fim, Jair Bolsonaro foi o vencedor da eleição presidencial de 2018.

Outros três candidatos completam a lista dos cinco mais populares conforme pesquisas eleitorais preliminares. Marina Silva vem concorrendo nas últimas três eleições presidenciais (desde 2010), e em todos os anos figurou nas pesquisas eleitorais como candidata proeminente. No entanto, quando as eleições se aproximam, os eleitores parecem migrar para outros candidatos.

Ciro Gomes parecia herdar os eleitores do ex-candidato Lula nas primeiras pesquisas de intenções de voto, no entanto, ele acabou na terceira posição e não foi classificado para o segundo turno. Finalmente, esperava-se que Geraldo Alckmin, ex-governador do estado mais populoso do Brasil, tivesse participação expressiva na corrida de 2018, uma vez que seu partido fez alianças com muitos partidos e obteve a maior parte do tempo eleitoral gratuito na TV e no Rádio. Esse tempo livre é garantido por lei no Brasil nas semanas que antecedem as eleições, e historicamente representa um recurso importante para os candidatos vencedores.

A seguir, as flutuações de popularidade dos candidatos nas eleições são analisadas considerando-se os eventos descritos anteriormente e comparadas com três diferentes métricas derivadas do Facebook. Duas delas foram obtidas através da *Facebook Graph API* e indicam respectivamente o número total de *likes* (*fan_count*) da página oficial do candidato e o número total de pessoas "falando" sobre a página (*talking_about_count*), esta última capturando as interações com a página em questão ⁴. Já a terceira métrica identifica o número total de pessoas interessadas em cada um dos candidatos e foi obtida com a API de Marketing do Facebook (descrita na metodologia).

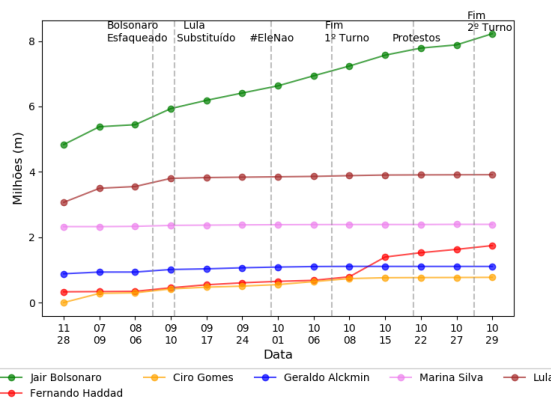


Figure 1: Número de curtidas por candidato

A Figura 1 mostra a variação no número de curtidas dos cinco principais candidatos e Lula. Observe que o vencedor da eleição, Bolsonaro, teve um grande aumento no número de curtidas em seu perfil público, aumentando de 4.8 milhões para 8.2 milhões. Aparentemente, sua popularidade no Facebook apresentou um crescimento acentuado logo após seu esfaqueamento em 6 de setembro.

Da mesma forma, os perfis públicos de Ciro Gomes e Fernando Haddad apresentaram um comportamento semelhante em termos de curtidas, tornando-se amplamente populares à medida que a campanha se passava. Ciro Gomes tinha cerca de 250 mil em sua página pública em novembro de 2017 e terminou a eleição com quase 800 mil, enquanto o crescimento dos seguidores de Haddad variou entre 330 e 1.7 milhão. Observe que após o final do primeiro turno, quando Haddad e Bolsonaro já estavam classificados para o segundo turno, Haddad se distanciou de Gomes no número de curtidas. Isso pode ser explicado pela união dos seguidores de Gomes à campanha de Haddad, ambos com alinhamentos ideológicos semelhantes, visando derrotar o candidato de direita.

Por outro lado, Geraldo Alckmin e Marina Silva eram de longe mais populares que Gomes e Haddad antes da campanha eleitoral, com 880 mil e 2.3 milhões, respectivamente. Enquanto o perfil público de Marina atraiu apenas 70 mil novas curtidas, Geraldo Alckmin teve um crescimento de apenas 25%, atingindo 1.1 milhão. Finalmente, pode-se observar que o número de curtidas de Lula cresceu até próximo à data do seu julgamento, estabilizando-se à medida que a campanha prosseguia.

Esses resultados indicam uma correlação entre o aumento no número de curtidas no Facebook e o aumento nas intenções de voto. A Figura 2 apresenta a variação dos cinco principais candidatos em termos de intenção de voto, de acordo com as pesquisas eleitorais do IBOPE. Nota-se que Bolsonaro, Haddad e Gomes têm as variações de intenções de voto correlacionadas com o aumento de curtidas. Após calcular a correlação de Spearman, observa-se valores altos, 0,97, 0,98, 0,80, respectivamente. Para Marina Silva e Alckmin a correlação foi negativa (-0,98 e -0,18).

A Figura 3 a) retrata a variação do número de pessoas falando sobre um candidato, indicando o número de pessoas interagindo com a página do Facebook (comentando, compartilhando o conteúdo publicado, mencionando a página, etc.). Já a Figura 3 b) representa o número total de pessoas interessadas em cada candidato. Observe que a quantidade de pessoas interessadas é às vezes, bem superior

⁴<https://developers.facebook.com/docs/graph-api/reference/page/>

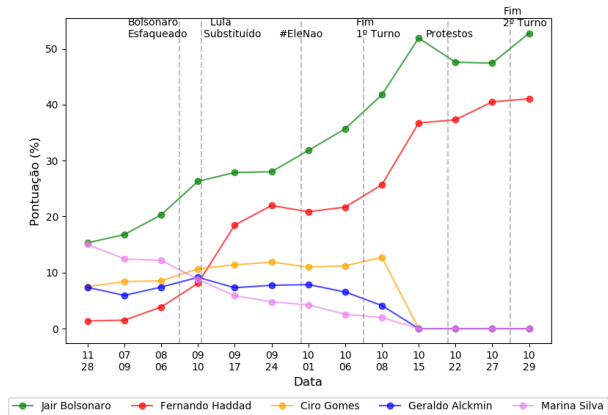


Figure 2: Intenção de voto por candidato (IBOPE)

ao número de curtidas da página oficial. O número de pessoas interessadas em Bolsonaro foi superior a 15 milhões no final da eleição. Isso pode ser explicado pelo fato de o Facebook incluir na contagem, pessoas que postam algo sobre o candidato ou interações com outras páginas relacionadas. Especialmente no período eleitoral, muitas páginas de apoio se popularizaram tais como *Bolsonaro Opressor e Bolsonaro 2.0*.

Ao analisar essas duas métricas, observa-se uma clara distinção na popularidade de Bolsonaro, Haddad e Gomes em comparação com Alckmin e Silva. Os candidatos mais votados atraíram muito mais atenção no Facebook do que candidatos menos votados. Em particular, nota-se que o número de pessoas falando sobre Bolsonaro, Haddad e Ciro Gomes aumentou consistentemente após os protestos 'EleNao' e logo após o primeiro turno. Depois disso, Gomes, não classificado para o segundo turno, perdeu a audiência, enquanto Bolsonaro e Haddad mantiveram o alto número de pessoas falando sobre eles. O quadro de interesse mostra um comportamento muito semelhante para as audiências de Haddad e Lula, cujas linhas seguem o mesmo padrão durante todo o período. De fato, boa parte das postagens do ex-candidato Lula faziam referência a Fernando Haddad. Isso indica claramente que o candidato substituído se aproveitou da popularidade de Lula, consistentemente consideravelmente.

É importante mencionar também que o número total de curtidas por si só não representa uma boa fonte de informação. Apesar da segunda posição no número de curtidas entre os candidatos, com pouco mais de 2 milhões de curtidas, Marina Silva obteve menos de 1.1 milhão de votos na eleição. Em uma análise de alto nível, o ato de curtir uma página indica que, em algum momento o usuário gostou do conteúdo ou considerou aquela página alinhada com seus interesses. Por outro lado, o aumento acentuado de curtidas em um curto período representa um aumento na popularidade e pode ser uma indicação mais forte de um aumento nas intenções de voto.

5 DADOS DEMOGRÁFICOS DO FACEBOOK VS. PESQUISAS ELEITORAIS

Na análise a seguir, pretende-se verificar até que ponto a variação na demografia da audiência de políticos no Facebook reflete a demografia das pesquisas eleitorais nos quesitos gênero e geográfico (nível estadual).

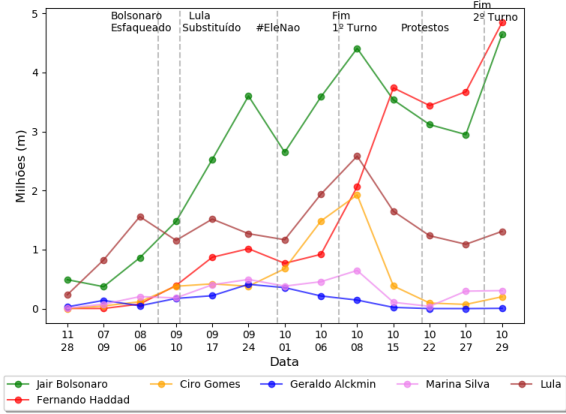
5.1 Gênero

Inicialmente, explorou-se a variação da distribuição por gênero da audiência dos candidatos. A Figura 4 exibe a variação de gênero na audiência dos 3 candidatos mais bem colocados no primeiro turno. Em uma análise de alto nível, a variação apresentou uma tendência similar entre o Facebook e os resultados das pesquisas eleitorais. Observe que a proporção de mulheres no público de Bolsonaro era muito menor do que a porcentagem de homens nas primeiras medições, 72% contra 28% de acordo com o Facebook. No entanto, a presença de mulheres cresceu no decorrer da campanha. O esfaqueamento do candidato teve uma ligação com o aumento inicial na popularidade geral do candidato. O percentual de mulheres também aumentou consideravelmente nas pesquisas eleitorais e no Facebook após a tentativa de assassinato.

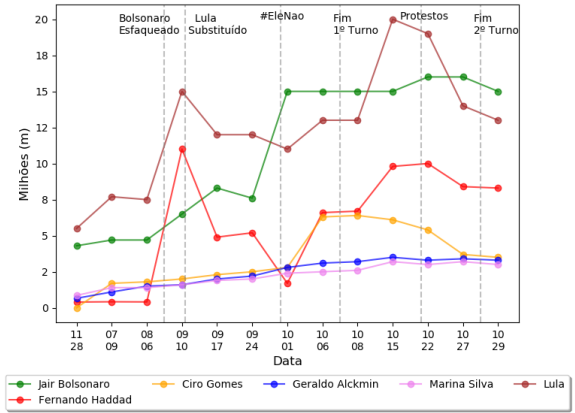
Logo após os protestos 'EleNao' e o fim do segundo turno, são detectados picos de audiência, nos quais o candidato vencedor atinge 52% e 50% no percentual de mulheres no Facebook. A variação do percentual nessas datas ao comparar com os valores da semana seguinte, foi de 12% e 8%, respectivamente. Os protestos também parecem influenciar o percentual de mulheres presentes no público do Fernando Haddad e Ciro Gomes. Em ambos os casos, a participação das mulheres cresceu, indicando que os protestos geraram um impacto no cenário geral. Esta tendência também é detectada por pesquisas eleitorais e pelos dados do Facebook. A variação de gênero nas audiências de Geraldo Alckmin e Marina Silva também apresentaram uma tendência semelhante, no entanto, as mudanças são mais suaves em comparação com outros candidatos. Por fim, observa-se que a substituição oficial do candidato Lula por Haddad foi seguida por um pico de mulheres na audiência do candidato Fernando Haddad.

Os dados do Facebook apresentam algumas mudanças repentinas na proporção de gênero quando comparados com as pesquisas eleitorais. Este aumento acentuado detectado por nossa metodologia pode ser explicado pelo enorme aumento no número de pessoas interessadas em Bolsonaro, conforme ilustrado na Figura 3 b), de 7,6 milhões para 15 milhões no período de protestos. Conforme detalhado anteriormente, as pessoas interessadas em determinada entidade (o candidato neste caso), não são apenas as pessoas que curtiram seu perfil público, mas também as pessoas que interagem com a página, que mencionam o candidato além de outros tipos de interação. Bolsonaro foi o centro das discussões nos dias anteriores aos protestos 'EleNao', o que provavelmente atraiu mais interações das mulheres e, no final, causou uma mudança significativa em seu público. As pesquisas eleitorais publicadas após os protestos, confirmaram que a popularidade de Bolsonaro aumentou entre as mulheres, no entanto, o Facebook parece superdimensionar essas mudanças, o que pode ser explicado pelo ambiente altamente polarizado durante o período eleitoral brasileiro. O mesmo ocorreu após a substituição do candidato Lula por Haddad, em que o interesse total e o número de pessoas falando sobre o substituto de Lula cresceram acentuadamente. Por outro lado, as menores variações detectadas na audiência de Marina Silva e Geraldo Alckmin podem ser explicadas pelo baixo engajamento da audiência desses candidatos (ver Figura 3).

A Figura 4 também descreve o percentual de homens e mulheres no primeiro e no segundo turno, conforme apontados por pesquisas



(a) Falando Sobre um Candidato



(b) Pessoas Interessadas

Figure 3: Número de pessoas “falando sobre um candidato” e total de interesse por candidato.

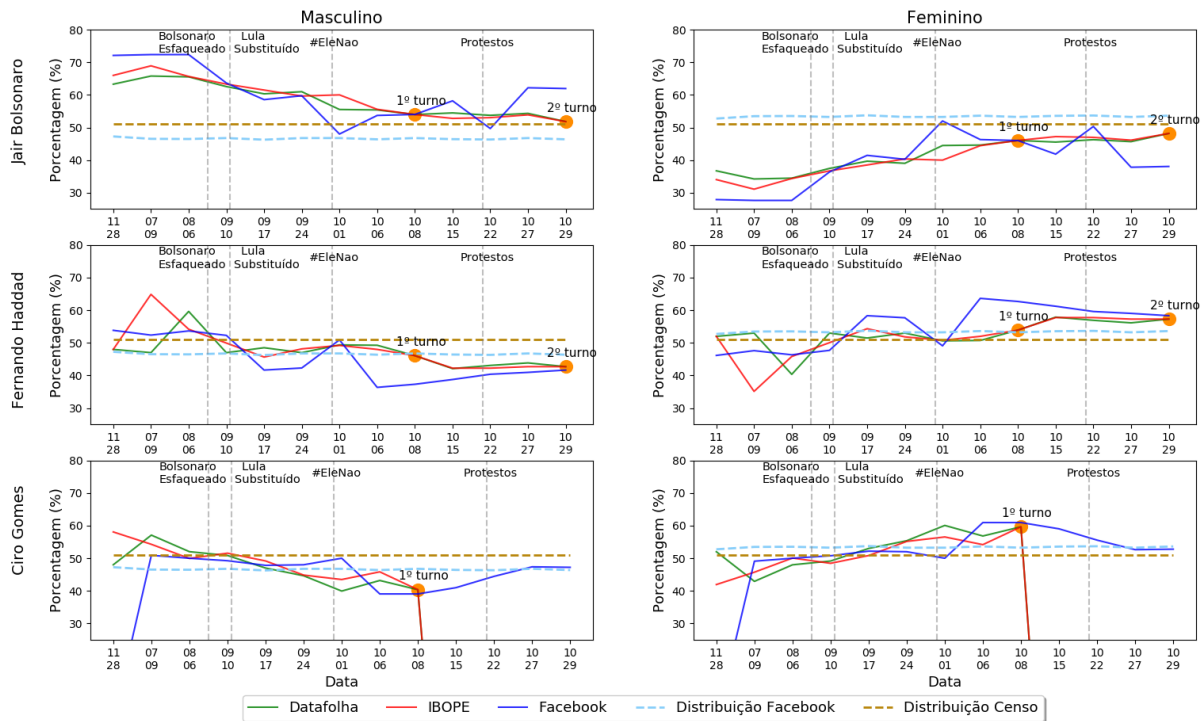


Figure 4: Interesses de gênero por candidatos (Top 3).

de boca de urna (destacados pelos pontos em laranja). Ao analisar os pontos em cor laranja, observa-se, por exemplo, que o percentual de homens e mulheres eleitores de Ciro Gomes está mais próxima da medida do Facebook do que das pesquisas eleitorais anteriores, ou seja, o erro das porcentagens do Facebook é inferior aos das pesquisas eleitorais. Para expandir a análise dos erros, considerou-se os valores da boca de urna para calcular a taxa de erro por gênero (e) de cada candidato. Utilizou-se fórmula a seguir, em que (i) é cada uma das dimensões demográficas (masculino e feminino), (pE) é a porcentagem calculada pela boca de urna e (pV) é a porcentagem

obtida com o Facebook ou com a pesquisa eleitoral que precedeu a eleição.

$$e = \sum_{i=1}^n \frac{pVi - pEi}{pEi}$$

De maneira geral, o erro representa a distância da última medição realizada antes do dia das eleições (primeiro e segundo turno) para o valor da pesquisa de boca de urna, incluindo o percentual para homens e mulheres. A Tabela 1 exibe o erro total na audiência de cada candidato considerando-se ambos os gêneros e os dois turnos. Nota-se que o erro medido no Facebook é menor do que das

pesquisas para Bolsonaro, Ciro Gomes e Marina Silva no primeiro turno. Em uma análise geral, o Facebook obteve taxas de erro menores em três dos sete casos, enquanto o IBOPE obteve menores erros para as outras quatro medições.

Table 1: Erro total (%) por gênero.

Primeiro Turno	Facebook	IBOPE	DataFolha
Jair Bolsonaro	0.06%	3.21%	2.85%
Fernando Haddad	17.65%	3.86%	6.41%
Ciro Gomes	2.71%	11.35%	5.85%
Marina Silva	4.01%	19.84%	34.62%
Geraldo Alckmin	13.7%	7.67%	16.14%
Segundo Turno	Facebook	IBOPE	DataFolha
Jair Bolsonaro	20.75%	4.12%	5.03%
Fernando Haddad	3.58%	0%	2.33%

5.2 Região

A análise por região fornece resultados semelhantes aos da análise anterior, sendo possível observar tendências similares de variação.

Uma das variações mais evidentes de público por regiões (os valores do centro-oeste e norte foram agrupados pois o Datafolha utiliza essa abordagem), sugere a transferência de votos de Lula para Haddad. Como mostrado na Figura 5, a porcentagem de apoiadores do Haddad na região sudeste cai drasticamente a partir de 9 de setembro e aumenta acentuadamente no nordeste e norte/centro-oeste. A variação ocorre juntamente com o julgamento de Lula, no qual ele foi impedido de disputar a eleição presidencial (30 de agosto). O candidato Fernando Haddad tornou-se oficialmente substituto de Lula em 10 de setembro. Observe que a tendência de diminuição no sudeste e aumento em outras regiões foi detectada em todas as pesquisas.

Também se calculou o erro nos resultados das eleições para a categoria de regiões, empregando a mesma fórmula utilizada na análise por gênero. Neste caso, foram adicionadas quatro diferentes dimensões demográficas, norte/centro-oeste, nordeste, sul e sudeste. No entanto, utilizou-se o resultado final da eleição para o cálculo. Diferente do percentual de votos por gênero, que não é revelado em virtude do voto secreto, o número de votos por estado é divulgado oficialmente, logo, não havia razão para utilizar as pesquisas de boca de urna. Os valores calculados com a nossa metodologia apresentam uma pequena taxa de erro para a Marina Silva e Ciro Gomes, ambos no primeiro turno. O erro do IBOPE foi menor em três casos, enquanto o Datafolha foi inferior em dois casos.

6 SISTEMA DESENVOLVIDO

A fim de fornecer um serviço útil como resultado deste trabalho, foi desenvolvido um sistema real que calcula e exibe a demografia da audiência dos políticos de acordo com a metodologia proposta (disponível em <http://www.audiencia-dos-politicos.dcc.ufmg.br/>). O sistema foi implantado meses antes da eleição brasileira e contribuiu para melhor compreensão do cenário político nacional no Brasil, na tentativa de elucidar os aspectos demográficos do público de cada candidato. O sistema apresenta a distribuição demográfica do público dos candidatos no Facebook, permitindo buscar os públicos

enviados e classificar por dimensões demográficas específicas (gênero, região, etc.). Observa-se, por exemplo, que 61% da audiência do candidato Jair Bolsonaro é masculina⁵ e 37% do público de Álvaro Dias é casado, contra apenas 23% de Audiência de Haddad⁶. Esses recursos atraíram atenção da imprensa pelo trabalho⁷, o que possivelmente influenciou o Facebook a adicionar mais interesses relacionados aos políticos brasileiros em sua plataforma de publicidade. Inicialmente, o Ciro Gomes não estava presente na plataforma de anúncios do Facebook como interesse, o que significa que não era possível direcionar anúncios políticos aos interessados no candidato. Logo após algumas notícias questionarem publicamente o Facebook se tal situação seria justa, o candidato foi incluído como um interesse na plataforma de publicidade⁸.

7 CONCLUSÃO

Este artigo apresenta uma nova abordagem que explora dados da plataforma de publicidade do Facebook para reproduzir as pesquisas eleitorais, cuja metodologia foi aplicada durante a campanha presidencial de 2018 no Brasil. Entre as principais conclusões, destaca-se que a popularidade dos candidatos no Facebook, capturada por meio três métricas diferentes (número de curtidas, pessoas falando sobre o candidato e número de pessoas interessadas) pode ser um bom indicador das variações de voto nas pesquisas de intenção de voto.

Concluiu-se também que a variação na audiência dos candidatos no Facebook quanto ao gênero e região geográfica segue tendências similares às detectadas pelas pesquisas eleitorais. Em particular, as mudanças repentinas ocasionadas por eventos relevantes durante a campanha, como protestos e candidato impedido de concorrer nas eleições, são bem capturados pelas medições do Facebook. No entanto, os picos parecem ser superdimensionados na perspectiva do Facebook, o que pode ser explicado pelo ambiente altamente polarizado e tenso verificado nas Redes Sociais Online.

Por fim, foi desenvolvido um sistema que detalha a audiência demográfica dos políticos brasileiros no Facebook (disponível em <http://www.audiencia-dos-politicos.dcc.ufmg.br/>). Sistemas como o apresentado neste artigo são úteis não apenas para usuários de mídias sociais, mas também para jornalistas e pesquisadores de mídias sociais que desejam entender o ecossistema eleitoral nas redes sociais. O sistema foi intensamente usado por jornalistas brasileiros para discutir o público online dos políticos. Como trabalho futuro, pretende-se expandir o sistema para outros países, em particular aqueles com eleições próximas.

8 ACKNOWLEDGMENTS

Este trabalho foi financiado por CNPq, CAPES e Fapemig.

REFERENCES

- [1] Matheus Araujo, Yelena Mejova, Ingmar Weber, and Fabricio Benevenuto. 2017. Using Facebook Ads Audiences for Global Lifestyle Disease Surveillance: Promises and Limitations. In *Proc. of the WebSci'17*.
- [2] Adam Bermingham and Alan F. Smeaton. 2011. On Using Twitter to Monitor Political Sentiment and Predict Election Results. In *Proc. of the SAAIP'11*.
- [3] Jessica Chung and Eni Mustafaraj. 2011. Can Collective Sentiment Expressed on Twitter Predict Political Elections?. In *Proc. of the AAAI'11*.

⁵<https://bit.ly/2De2Xa1>

⁶<https://bit.ly/2UIRTIV>

⁷<https://www.bbc.com/portuguese/salasocial-43824463>

⁸<https://bit.ly/2LqHwN>

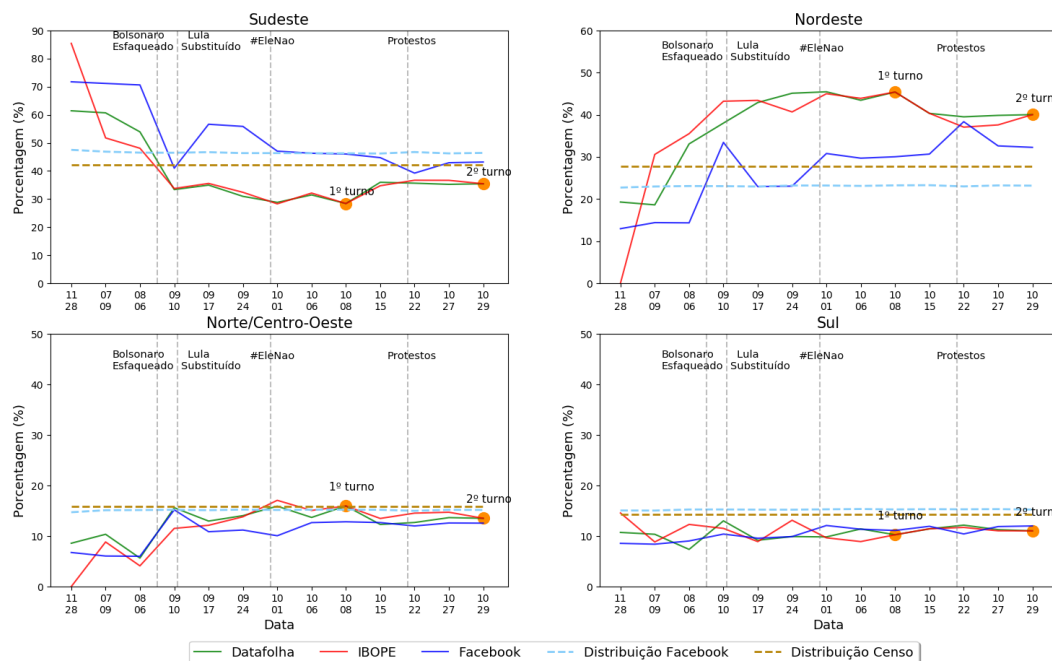


Figure 5: Distribuição por região - Fernando Haddad.

- [4] Joseph DiGrazia, Karissa McKelvey, Johan Bollen, and Fabio Rojas. 2013. More Tweets, More Votes: Social Media as a Quantitative Indicator of Political Behavior. *PLOS ONE* 8 (11 2013), 1–5.
- [5] David Garcia, Yonas Mitike Kassa, Angel Cuevas, Manuel Cebrian, Esteban Moro, Iyad Rahwan, and Ruben Cuevas. 2018. Analyzing gender inequality through large-scale Facebook advertising data. *Proc. of the PNAS* 115, 27 (2018), 6958–6963.
- [6] Daniel Gayo-Avello. 2012. “I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper” – A Balanced Survey on Election Prediction using Twitter Data. *CoRR* abs/1204.6441 (2012).
- [7] Daniel Gayo-Avello, Panagiotis Takis Metaxas, and Eni Mustafaraj. 2011. Limits of Electoral Predictions Using Twitter. In *Proc. of the ICWSM’11*.
- [8] Fabio Giglietto. 2012. If Likes Were Votes: An Empirical Study on the 2011 Italian Administrative Elections. In *Proc. of the ICWSM’12*.
- [9] J. Gomide, A. Veloso, W. Meira Jr., V. Almeida, F. Benevenuto, F. Ferraz, and M. Teixeira. 2011. Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. In *Proc. of the WebSci’11*.
- [10] Andreas Jungherr, Pascal Jürgens, and Harald Schoen. 2012. Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. “Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment”. *Social Science Computer Review* 30, 2 (2012), 229–234.
- [11] Juhi Kulshrestha, Motahhare Eslami, Johnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P. Gummadi, and Karrie Karahalios. 2017. Quantifying Search Bias: Investigating Sources of Bias for Political Searches in Social Media. In *Proc. of the CSCW’17*.
- [12] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. 2010. What is Twitter, a social network or a news media?. In *Proc. of the WWW’10*.
- [13] Anders Olof Larsson and Hallvard Moe. 2012. Studying political microblogging: Twitter users in the 2010 Swedish election campaign. *New Media & Society* 14, 5 (2012), 729–747.
- [14] Kathy Lee, Ankit Agrawal, and Alok Choudhary. 2013. Real-time Disease Surveillance Using Twitter Data: Demonstration on Flu and Cancer. In *Proc. of the KDD’13*.
- [15] Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J. Niels Rosenquist. 2011. Understanding the Demographics of Twitter Users. In *Proc. of the ICWSM’11*.
- [16] Amy Mitchell. 2016. Key findings on the traits and habits of the modern news consumer. <http://www.pewresearch.org/fact-tank/2016/07/07/modern-news-consumer/>. Pew Research Center (July 2016).
- [17] Fred Morstatter, Jürgen Pfeffer, Huan Liu, and Kathleen M. Carley. 2013. Is the Sample Good Enough? Comparing Data from Twitter’s Streaming API with Twitter’s Firehose. *CoRR* abs/1306.5204 (2013).
- [18] Gustavo Resende, Philippe Melo, Hugo Sousa, Johnatan Messias, Marisa Vasconcelos, Jussara Almeida, and Fabrício Benevenuto. 2019. (Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures. In *Proc. of the WWW’19*.
- [19] Filipe N. Ribeiro, Lucas Henrique, Fabrício Benevenuto, Abhijnan Chakraborty, Juhi Kulshrestha, Mahmoudreza Babaei, and Krishna P. Gummadi. 2018. Media bias monitor: Quantifying biases of social media news outlets at large-scale. In *Proc. of the ICWSM’18*.
- [20] Filipe N. Ribeiro, Koustuv Saha, Mahmoudreza Babaei, Lucas Henrique, Johnatan Messias, Fabrício Benevenuto Oana Goga, Krishna P. Gummadi, and Elissa M. Redmiles. 2019. On Microtargeting Socially Divisive Ads: A Case Study of Russia-Linked Ad Campaigns on Facebook. In *Proc. of the FAT’19*.
- [21] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. In *Proc. of the WWW’10*.
- [22] Erik Tjong Kim Sang and Johan Bos. 2012. Predicting the 2011 Dutch Senate Election Results with Twitter. In *Proc. of the EACL’12*.
- [23] M. Skoric, N. Poor, P. Achananuparp, E. Lim, and J. Jiang. 2012. Tweets and Votes: A Study of the 2011 Singapore General Election. In *Proc. of the HICSS’12*.
- [24] Aaron Smith and Monica Anderson. 2018. Social Media Use in 2018. <http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/>. Pew Research Center (July 2018).
- [25] Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe N. Ribeiro, George Arvanitakis, Fabrício Benevenuto, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. 2018. On the Potential for Discrimination in Online Targeted Advertising. In *Proc. of the FAT’18*.
- [26] Karolina Sylwester and Matthew Purver. 2015. Twitter Language Use Reflects Psychological Differences between Democrats and Republicans. *PLOS ONE* 10, 9 (09 2015), 1–18.
- [27] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. 2010. Predicting elections with Twitter: what 140 characters reveal about political sentiment. In *Proc. of the ICWSM’10*.
- [28] Wei Wang, David Rothschild, Sharad Goel, and Andrew Gelman. 2014. Forecasting elections with non-representative polls. *International Journal of Forecasting* (September 2014).
- [29] Christine Williams and Girish Gulati. 2009. What is a Social Network Worth? Facebook and Vote Share in the 2008 Presidential Primaries. In *Proc. of the APSA’09*.
- [30] Emilio Zagheni, Kivan Polimis, Monica Alexander, Ingmar Weber, and Francesco C. Billari. 2018. Combining Social Media Data and Traditional Surveys to Nowcast Migration Stocks. In *Proc. of the PAA’18*.