

(Re)shaping online narratives: when bots promote the message of President Trump during his first impeachment

Michael C. Galgoczy¹, Atharva Phatak², Danielle Vinson³,
Vijay K. Mago² and Philippe J. Giabbanelli¹

¹ Department of Computer Science & Software Engineering, Miami University, Oxford, OH, United States

² Department of Computer Science, Lakehead University, Thunder Bay, ON, Canada

³ Department of Politics & International Affairs, Furman University, Greenville, SC, United States

ABSTRACT

Influencing and framing debates on Twitter provides power to shape public opinion. Bots have become essential tools of ‘computational propaganda’ on social media such as Twitter, often contributing to a large fraction of the tweets regarding political events such as elections. Although analyses have been conducted regarding the first impeachment of former president Donald Trump, they have been focused on either a manual examination of relatively few tweets to emphasize rhetoric, or the use of Natural Language Processing (NLP) of a much larger *corpus* with respect to common metrics such as sentiment. In this paper, we complement existing analyses by examining the role of bots in the first impeachment with respect to three questions as follows. (Q1) Are bots actively involved in the debate? (Q2) Do bots target one political affiliation more than another? (Q3) Which sources are used by bots to support their arguments? Our methods start with collecting over 13M tweets on six key dates, from October 6th 2019 to January 21st 2020. We used machine learning to evaluate the sentiment of the tweets (*via* BERT) and whether it originates from a bot. We then examined these sentiments with respect to a balanced sample of Democrats and Republicans directly relevant to the impeachment, such as House Speaker Nancy Pelosi, senator Mitch McConnell, and (then former Vice President) Joe Biden. The content of posts from bots was further analyzed with respect to the sources used (with bias ratings from AllSides and Ad Fontes) and themes. Our first finding is that bots have played a significant role in contributing to the overall negative tone of the debate (Q1). Bots were targeting Democrats more than Republicans (Q2), as evidenced both by a difference in ratio (bots had more negative-to-positive tweets on Democrats than Republicans) and in composition (use of derogatory nicknames). Finally, the sources provided by bots were almost twice as likely to be from the right than the left, with a noticeable use of hyper-partisan right and most extreme right sources (Q3). Bots were thus purposely used to promote a misleading version of events. Overall, this suggests an intentional use of bots as part of a strategy, thus providing further confirmation that computational propaganda is involved in defining political events in the United States. As any empirical analysis, our work has several limitations. For example, Trump’s rhetoric on Twitter has previously been characterized by an overly negative tone, thus tweets detected as negative may be echoing his message rather than acting against him. Previous works show that this possibility is limited, and its existence would only strengthen our conclusions. As our analysis is based on NLP, we focus on processing a large volume

Submitted 26 November 2021

Accepted 23 March 2022

Published 15 April 2022

Corresponding author

Philippe J. Giabbanelli,

giabbapj@miamioh.edu

Academic editor

Kathiravan Srinivasan

Additional Information and
Declarations can be found on
page 20

DOI 10.7717/peerj-cs.947

© Copyright

2022 Galgoczy et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

of tweets rather than manually reading all of them, thus future studies may complement our approach by using qualitative methods to assess the specific arguments used by bots.

Subjects Computational Linguistics, Data Science, Natural Language and Speech, Network Science and Online Social Networks

Keywords Bots, Impeachment, Politics, Sentiment, Twitter

INTRODUCTION

The efforts of party leaders and presidents to shape public understanding of issues through mainstream and partisan media are well documented (*Sellers, 2009; Vinson, 2017*). Partisan media intentionally adopt ideological frames and cover stories in ways that favor the politicians in their own party and create negative impressions of those in the other party, contributing to polarization in American politics (*Levendusky, 2013; Forgette, 2018*). At the interface of political communication and computational sciences, the emerging field of computational politics has produced many analyses of polarization over the recent years (*Haq et al., 2020; Pozen, Talley & Nyarko, 2019; Yarchi, Baden & Kligler-Vilenchik, 2021*). Comparatively less computational research has been devoted to how politicians influence debate on social media, but interest in this question has grown during Donald Trump's presidency as he extensively communicated directly to the public *via* Twitter (*Ouyang & Waterman, 2020*).

While mainstream media traditionally favor those with political power, giving them coverage that expands their public reach, social media rewards those who can attract an audience. On Twitter, a large attentive audience can amplify a politician's message and change or mobilize public opinion (*Zhang et al., 2018; Russell, 2021*). Trump's rhetoric features many characteristics of populism, including emotional appeals based on opinion rather than fact and allegations of corruption, that attracts followers; indeed, his messages are often picked up by right-wing populist leaders around the world (*Pérez-Curiel, Rivas-de Roca & García-Gordillo, 2021*). As research on the 2016 presidential campaign illustrates, Trump had an attentive audience, particularly on the far right, that readily amplified his message through retweets (*Zhang et al., 2018*); a similar phenomenon was found in the 2020 election, for which datasets are gradually assembled and examined (*Abilov et al., 2021; Grimminger & Klinger, 2021*). These events repeatedly show that news media responded to Trump's large volume of tweets by giving him substantially more coverage than other candidates, further expanding his reach to influence and frame public debate (*Wells et al., 2020*). Computational analyses further suggest that the use of tweets was a strategic activity, as Trump's Twitter activity increased in response to waning news coverage from Center Right, Mainstream, Left Wing, and Far Left media (*Wells et al., 2020*). Influencing and framing the debate on Twitter thus serves to shape media coverage, public opinion, and ultimately impact the political agenda (*Şahin, Johnson & Korkut, 2021*).

In this sense, shaping public discourse is connected to political power. Studies of Twitter rhetoric indicate that patterns in framing and use of language occur within digital networks that create group identities (Alfonzo, 2021). Critical discourse analysis suggests that examining messaging can identify power asymmetries, revealing the institutions and leaders who control public discourse in particular environments or around specific issues and events (Van Dijk, 2009). Such research can reveal abuse of power by dominant groups and resistance by marginalized groups. Given Trump's populist rhetoric and its potential for spreading disinformation (Pérez-Curiel, Rivas-de Roca & García-Gordillo, 2021), amplification of his messages would have important implications for public discourse. Though critical analysis would require a more detailed and nuanced coding of the tweets than we are able to provide with our methodology, looking at the amplification of political discourse by bots allows us to see the potential manipulation of public discourse, especially if the bots favor one faction, that could affect the power dynamics between parties and have significant negative implications for democracy. This kind of power merits additional study. Our study of the first impeachment of Donald Trump provides such an opportunity.

This paper is part of the growing empirical literature on the social influence carried by Trump's tweets prior to the permanent suspension of his account on 8 January 2021. Such studies are now covered by dedicated volumes (Kamps, 2021; Schier, 2022) as well as a wealth of scholarly works, including large-scale social network analysis and mining (Zheng et al., 2021), analyses of media responses (Christenson, Kreps & Kriner, 2021), the interplay of tweets and media (Morales, Schultz & Landreville, 2021), or the (absence of) impact between his tweets and stock returns (Machus, Mestel & Theissen, 2022). Related studies also include analyses about comments from political figures about Trump (Milford, 2021; Alexandre, Jai-sung Yoo & Murthy, 2021) and vice versa (Brown Crosby, 2022), analyses of his letter during the first impeachment (Reyes & Ross, 2021), and detailed examinations about how Trump's message was perceived by specific groups such as White extremists (Long, 2022). The impeachment and handling of COVID-19 have received particular attention on Twitter (Dejard et al., 2021; Cervi, García & Marín-Lladó, 2021), as key moments of his presidency. While the use of bots to create social media content during presidential elections and COVID-19 have received attention (Xu & Sasahara, 2021), there is still a need to examine how bots were used during the first impeachment.

The specific contribution of our study is to complement existing (predominantly manual) analyses of tweets during the first impeachment, which often focused on tweets from Donald Trump (Gould, 2021; Driver, 2021) or Senators (McKee, Evans & Clark, 2022), by using machine learning to examine the role of bots in a large collection of tweets. This complementary analysis has explicitly been called for in recent publications (Tachaiya et al., 2021), which performed large-scale sentiment analyses during the first impeachment (via Reddit and 4chan), but did not investigate the efforts at social

engineering deployed *via* bots. Our specific contribution focuses on addressing three specific questions:

(Q1) Are bots actively involved in the debate?

(Q2) Do bots target one political affiliation more than another?

(Q3) Which sources are used by bots to support their arguments?

The remainder of this paper is organized as follows. In the next section, we explain how we gathered over 13M tweets based on six key dates of the first impeachment, and analyzed them using BERT (for sentiment) and Botometer (for bot detection) as well as bias ratings (for source evaluation) and Latent Dirichlet Allocation (for theme mining). We then present our results for a balanced sample of key political figures in the impeachment (including Adam Schiff, Joe Biden, Mitch McConnell, Nancy Pelosi, Donald Trump) as well as others relevant to our analysis. For transparency and replicability of research, tweets and detailed results are available on a third-party repository without registration needed, at <https://osf.io/3nsf8/>. Finally, our discussion contextualizes the answer for each of the three questions within the broader literature on computational politics and summarizes some of our limitations.

METHODS

Data collection

For a detailed description of the Ukraine Whistleblower Scandal, we refer the readers to a recent analysis by McKee and colleagues who also contextualize this situation within the broader matter of hyper-polarization (*McKee, Evans & Clark, 2022*). A short overview is as follows. On 9 September 2019, the House Intelligence Committee, headed by Rep. Adam Schiff, was alerted to the existence of a whistleblower complaint that involved the Trump Administration. Over the next 2 weeks, information from media reports indicated that President Trump had withheld aid to Ukraine possibly to pressure the new Ukrainian President to launch a corruption investigation into Democratic presidential hopeful Joe Biden's son Hunter who had been on the board of a Ukrainian company when the elder Biden was Vice President during the Obama Administration (*Trautman, 2019*). By 25th September, House Speaker Nancy Pelosi, a Democrat, had launched an impeachment inquiry, and the Trump Administration had released notes from Trump's call with the Ukrainian President that some saw as evidence that Trump had conditioned aid and a visit to the White House on Ukraine investigating Hunter Biden. Our data collection started soon after, covering six key dates leading to the impeachment process.

On 6 October 2019, a second whistleblower was identified and we collected 1,968,943 tweets. On 17 October 2019, then Senate Majority Leader Mitch McConnell briefed the Senate on the impeachment process (1,960,808 tweets). On 14 November 2019, the second day of impeachment hearings began in the House Intelligence Committee (1,977,712 tweets). On 5 December 2019, 1 day after the House Judiciary Committee opened impeachment hearings, four constitutional lawyers testified in front of Congress (1,960,813 tweets). On 18 December 2019, the House debated and passed two articles of

Table 1 Distribution of tweets collected.

| Date | Tweets collected |
|------------------|------------------|
| 6 October 2019 | 1,968,943 |
| 17 October 2019 | 1,960,808 |
| 14 November 2019 | 1,977,712 |
| 5 December 2019 | 1,960,813 |
| 19 December 2019 | 2,041,924 |
| 21 January 2020 | 3,360,434 |

impeachment. We collected 2,041,924 to gather responses on the following day. Finally, on 21 January 2020, the Senate impeachment trial began and we ended the data collection with 3,360,434 tweets. In total, 13,272,577 tweets were collected (Table 1).

Throughout the process, both parties and the President tried to control the public narrative over impeachment. Early in the process, Democrats began to talk about abuse of power or corruption by Trump and eventually shifted to the more specific language of bribery and extortion. Meanwhile, Republican elected officials seemed to be trying to frame this as a coup by the Democrats. Republicans directed their ire at House Intelligence Committee Chairman Democrat Adam Schiff who led the impeachment inquiry. The President and some of his Republican allies, including Senator Lindsey Graham and aide Stephen Miller, tried to divert attention from the President by alleging misconduct by Joe and Hunter Biden. Our key words for data collection were thus chosen to reflect the messaging of the two parties in what was shaping up as a largely partisan process: Trump, Impeachment, Coup, Abuse of Power, Schiff (as in Adam), and Biden. We did not explicitly include words like disinformation, though the messaging may have included unfounded allegations, because such words were not part of either party's framing.

Tweets were gathered using our Twitter mining system (Mendhe et al., 2020), which allows for the collection of social media data and the application of many filters for cleaning and further use for machine learning techniques.

Preprocessing

This study and the methods employed only target texts in English, hence we removed tweets in other languages as well as tweets only containing hyperlinks. Language detection of Twitter data is more difficult than that of standard texts due to the unusual language used within tweets. We used the `langdetect` library (Table 2), which utilizes a naive Bayes classifier and achieves strong performances on tweets (Lui & Baldwin, 2014). Tweets in English may have to undergo extensive cleaning through a variety of filters, depending on the model used for analysis. As detailed in the next section, we use the BERT model, or Bidirectional Encoder Representations from Transformers. BERT is pre-trained on *unprocessed* English text (Devlin et al., 2018) hence the only processing performed was to remove tweets that consisted of only hyperlinks. A more detailed discussion of our data collection platform and its combination with BERT can be found in (Qudar & Mago, 2020),

Table 2 Language detection using langdetect.

| Tweet | Output |
|---|--------|
| tüm diŞ politikayı tek bir adama yani trump'a bağlamak tüm | tr |
| I'm sure Trump will tweet about rep Elijah Cummings passing | en |

Table 3 Nicknames of key actors.

| Actor | Nicknames |
|-----------------|------------------------|
| Nancy Pelosi | Nervous Nancy |
| Adam Schiff | Shifty Schiff |
| Joe Biden | Sleepy Joe, Creepy Joe |
| Mitch McConnell | Midnight Mitch |

while an abundance of studies show how BERT is used for classification of tweets ([Singh, Jakhar & Pandey, 2021](#); [Yenduri et al., 2021](#); [Sadia & Basak, 2021](#)).

Analysis

In textual data analysis, we can apply information extraction and text mining to parse information from a large data set. In order to understand how each side put their messages out, data was broken down by mention of specific politicians—those who were the subject of impeachment (Trump) and countermessaging by Republicans (Biden), those who were important to the impeachment process and messaging for their party (Pelosi, McConnell, and Schiff), and two Republicans (Miller and Graham) “who represent accounts closely associated with President Trump” per previous Twitter analyses ([Boucher & Thies, 2019](#); [Hawkins & Kaltwasser, 2018](#)). This sample of actors is evenly divided among Democrats and Republicans. After assigning tweets to each of the actors, we then used sentiment analysis and bot detection. The analysis was further nuanced by accounting for the use of a derogatory nickname ([Table 3](#)) in lieu of the official actor’s name. In addition, we performed complementary examinations of the content of tweets by bots, by examining their themes (in contrast with human tweets) and the bias exhibited in the websites used as sources.

In summary, the data associated with each tweet (content, time, author) is automatically expanded with the list of actors involved, the type of author (bot, human, unknown), the sentiment (negative, neutral, positive), and the list of websites used.

Sentiment analysis

Sentiment analysis is a very important technique to analyze textual data and is frequently used in natural language processing, text analysis, and computational linguistics. Our model classifies sentiment into three categories: negative, neutral, and positive. As mentioned above, we use the natural language processing model BERT, designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers ([Devlin et al., 2018](#)). As its name indicates, BERT is a

pre-trained deep bidirectional transformer, whose architecture consists of multiple encoders, each composed of two types of layers (multi-head self-attention layers, feed forward layers). To appreciate the number of parameters, consider that the text first goes through an embedding process (two to three dozen million parameters depending on the model), followed by transformers (each of which adds 7 or 12.5 million parameters depending on the model), ending with a pooling layer (0.5 or 1 million more parameters depending on the model). All of these parameters are trainable. Although the model has already been pre-trained, it is common practice to give it additional training based on data specific to the context of interest.

In order to train the model, a set of 3,000 unprocessed tweets was randomly selected, distributed evenly across each of the six collection dates. The tweets were then manually labelled for overall sentiment, and fed into the model for training. This approach is similar to other works, such as [Grimminger & Klinger \(2021\)](#), in which 3,000 tweets were manually annotated for a BERT classification. While there is no universal number for how many tweets should be labeled when using BERT, we observe a range across studies from under 2,000 labeled tweets ([Şaşmaz & Tek, 2021](#); [Alomari, ElSherif & Shaalan, 2017](#); [Peisenieks & Skadiņš, 2014](#)) up to several thousands ([Golubev & Loukachevitch, 2020](#); [Rustam et al., 2021](#); [Nabil, Aly & Atiya, 2015](#); [Zhang et al., 2020](#)). In line with previous studies on sentiment analysis in tweets ([Rustam et al., 2021](#); [Khan et al., 2021](#)), we compared BERT and several other supervised machine learning models from the `textblob` library (*i.e.*, decision tree, Naive Bayes) as well as `scikit-learn` (logistic regression, support vector machine). Our comparison in [Table 4](#) confirms that BERT leads to satisfactory and more stable performances than alternatives, thus validating the choice of BERT in our *corpus*.

Bot detection

We use the `botometer` library (formerly know as `BotOrNot`) to detect the presence of bots in Twitter. `botometer` is a paid service that utilizes the Twitter API to compile over 1,000 features of a given Twitter account, including how frequently the account tweets, how similar each tweet is to previous tweets from the account, and the account's ratio of followers to followees ([Davis et al., 2016](#)). Features are examined both within their respective categories and collectively. The resulting series of probabilities is returned ([Table 5](#)). To provide a conservative estimate, we scanned almost half of the different accounts associated with tweets (1,101,023 out of 2,438,343 unique accounts). Since some accounts were deleted or suspended ($n = 113,037$), our analysis of bots is based on a sample of 987,987 active accounts. We note that the percentage of accounts unavailable is similar to [Le et al. \(2019\)](#), in which 9.5% of Twitter accounts involved in the 2016 election were suspended.

In order to classify Twitter accounts as human or bot, a classifier needs to be properly calibrated on annotated `botometer` data ([Spagnuolo, 2019](#)). We considered three supervised machine learning techniques, which are frequently used alongside BERT to process tweets: decision trees, extreme gradient boosted trees (XGBoost), and random forests ([Kumar et al., 2021](#); [Rustam et al., 2021](#)). All three methods were implemented in

Table 4 Comparison of classification accuracy for BERT (which we use in this study), TextBlob library, and two scikit-learn algorithms.

| Model name | Class | Metric | Value | |
|--------------------|-----------|-----------|----------|--------|
| Decision Tree | Overall | Accuracy | 0.62 | |
| | | Precision | 0.64 | |
| | | Recall | 0.8816 | |
| | Negative | F1 | 0.744 | |
| | | Precision | 0.537 | |
| | | Recall | 0.2562 | |
| | Neutral | F1 | 0.3465 | |
| | | Precision | 0.478 | |
| | | Recall | 0.2104 | |
| Naive Bayes | Positive | F1 | 0.2922 | |
| | | Overall | Accuracy | 0.658 |
| | | Precision | 0.664 | |
| | Negative | Recall | 0.918 | |
| | | F1 | 0.7707 | |
| | | Precision | 0.638 | |
| | Neutral | Recall | 0.3607 | |
| | | F1 | 0.4608 | |
| | | Precision | 0.617 | |
| BERT | Positive | Recall | 0.0979 | |
| | | F1 | 0.1686 | |
| | | Overall | Accuracy | 0.7398 |
| | Negative | Precision | 0.821 | |
| | | Recall | 0.806 | |
| | | F1 | 0.813 | |
| | Neutral | Precision | 0.6246 | |
| | | Recall | 0.666 | |
| | | F1 | 0.6348 | |
| Logistic Regressor | Positive | Precision | 0.612 | |
| | | Recall | 0.584 | |
| | | F1 | 0.598 | |
| | Overall | Accuracy | 0.635 | |
| | | Precision | 0.652 | |
| | | Recall | 0.9 | |
| | Negative | F1 | 0.7535 | |
| | | Precision | 0.543 | |
| | | Recall | 0.303 | |
| Neutral | F1 | 0.389 | | |
| | Precision | 0.72 | | |
| | Recall | 0.13125 | | |
| Positive | F1 | 0.217 | | |

Table 4 (continued)

| Model name | Class | Metric | Value |
|------------------------|----------|-----------|--------|
| Support Vector Machine | Overall | Accuracy | 0.64 |
| | | Precision | 0.639 |
| | | Recall | 0.94 |
| | Negative | F1 | 0.762 |
| | | Precision | 0.637 |
| | | Recall | 0.238 |
| | Neutral | F1 | 0.3465 |
| | | Precision | 0.665 |
| | | Recall | 0.128 |
| | Positive | F1 | 0.199 |

Table 5 Botometer results.

| Twitter account ID | English | Content | ... | Universal |
|--------------------|----------|----------|-----|-----------|
| 999800336750530560 | 0.067063 | 0.938607 | ... | 0.536176 |
| 306127388 | 0.046774 | 0.833483 | ... | 0.207906 |

scikit-learn, and our scripts are provided publicly on our repository. Training data, consisting of manually labelled Twitter accounts, was retrieved from <https://botometer.osome.iu.edu/bot-repository/datasets.html> and compiled, thus providing 3,000 accounts. For each of the three supervised machine learning training, we optimized a classifier using 10-fold cross validation and a grid search over commonly used hyper-parameters. The hyper-parameters considered and their values are listed in Table 6, while each model was exported from scikit-learn and saved on our open repository for inspection and reuse by the research community. As performances are comparable across the three techniques, we used the decision tree (which can more readily be interpreted by experts) to classify accounts from our collected data.

Content of tweets by Bots

Once we detect the posts emanating from bots, we can further examine their content. In particular, we use a frequency counts approach to measure the number of times that each website is cited in the bots' tweets. Specifically, we retrieve all URLs from these tweets and map them to a common form such that 'www.foxnews.com', 'media2.foxnews.com' or 'radio.foxnews.com' are all counted as foxnews. Then, we examine the political leanings of the websites by cross-referencing them with sources for bias ratings. We accomplish this in the same manner as in the recent work of *Huszár et al. (2022)*, which used two sources for bias ratings: AllSides (<https://www.allsides.com/media-bias/ratings>) and Ad Fontes (<https://adfontesmedia.com/interactive-media-bias-chart>). In line with their work, we do not claim that either source provides objectively better ratings (*Huszár et al., 2022*), hence we report both. If a website is rated by neither source, then we access it to read its content and evaluate it manually; such evaluations are disclosed explicitly.

Table 6 Classification performances on bot detection.

| Approach | Hyper-parameters | | Performances | | | | |
|---|--|-------------|--------------|-------|---------|-----------|--------|
| | Values explored <i>via</i> grid search | Best values | Acc. | F1 | ROC-AUC | Precision | Recall |
| Decision Tree | Criterion (gini/entropy), | entropy | 0.885 | 0.903 | 0.890 | 0.938 | 0.871 |
| | max_depth (1, 2, ..., 10), | 5 | | | | | |
| | min samples leaf (2, 3, ..., 20), | 14 | | | | | |
| | max leaf nodes (1, 2, ..., 20) | 17 | | | | | |
| eXtreme Gradient Boosting (XGBoost) for trees | Max depth (5, 6, ..., 10), | 5 | 0.892 | 0.909 | 0.893 | 0.933 | 0.887 |
| | alpha (0.1, 0.3, 0.5), | 0.5 | | | | | |
| | learning rate (0.01, 0.02, ..., 0.05), | 0.02 | | | | | |
| | estimators (100, 200, 300) | 200 | | | | | |
| Random Forests | Max depth (5, 6, ..., 10), | 8 | 0.898 | 0.915 | 0.899 | 0.934 | 0.897 |
| | criterion (gini/entropy), | entropy | | | | | |
| | estimators (100, 200, 300) | 200 | | | | | |

In addition to sources, we examined themes. In line with the typical computational approach used in similar studies on Twitter (*Jelodar et al., 2019*; *Ostrowski, 2015*; *Negara, Triadi & Andryani, 2019*), we use Latent Dirichlet allocation (LDA) to extract topics. To allow for fine-grained comparison, we perform the extraction on every day of the data collection and separate themes of human posts from bots.

RESULTS

The size of the dataset after preprocessing is provided in the first subsection. For transparency and replicability of research, tweets are available at <https://osf.io/3nsf8/>; they are organized by key actors and labeled with the sentiment expressed or whether they originate from a bot.

Preprocessing

In Twitter research, pre-processing often leads to removing most of the data. For example, our previous research on Twitter regarding the Supreme Court (*Sandhu et al., 2019*) discarded 87–89% of the data, while our examination of Twitter and obesity discarded 73% of the data (*Sandhu, Giabbanelli & Mago, 2019*). The reason is that pre-processing has historically involved a series of filters (*e.g.*, removing words that are not deemed informative in English, removing hashtags and emojis), which were necessary as the analysis model could not satisfactorily cope with raw data. In contrast, BERT can directly take tweets as input. Our pre-processing thus only eliminated non-English tweets and tweets with no substantive information that consisted of only hyperlinks. As a result, most of the data is preserved for analysis: we kept 11,007,403 of the original 13,568,750 tweets (*Table 7*).

Table 7 Distribution of tweets.

| Date | Tweets collected | Proportion remaining (%) |
|------------------|------------------|--------------------------|
| 6 October 2019 | 1,877,693 | 95.3 |
| 17 October 2019 | 1,830,552 | 93.4 |
| 14 November 2019 | 1,828,402 | 92.5 |
| 5 December 2019 | 1,855,771 | 94.6 |
| 19 December 2019 | 1,998,802 | 97.9 |
| 21 January 2020 | 3,230,843 | 96.1 |

| Actor | Name | Post type | | | Sentiment | | |
|-----------------|----------|-----------|-------|---------|-----------|---------|----------|
| | | Bot | Human | Unknown | Negative | Neutral | Positive |
| Adam Schiff | Any | 24.10 | 40.90 | 35.00 | 74.10 | 16.70 | 9.20 |
| Adam Schiff | Nickname | 26.80 | 36.10 | 37.10 | 96.80 | 0.20 | 3.10 |
| | Normal | 24.10 | 41.00 | 34.90 | 73.50 | 17.10 | 9.30 |
| Joe Biden | Any | 22.70 | 41.00 | 36.30 | 76.30 | 21.50 | 2.20 |
| | Nickname | 28.10 | 34.50 | 37.30 | 97.80 | 1.90 | 0.40 |
| | Normal | 22.40 | 41.40 | 36.30 | 75.10 | 22.60 | 2.30 |
| Mitch McConnell | Any | 22.80 | 49.60 | 27.60 | 64.20 | 33.00 | 2.80 |
| Mitch McConnell | Nickname | 16.60 | 40.90 | 42.50 | 94.40 | 4.70 | 1.00 |
| | Normal | 22.80 | 49.60 | 27.60 | 64.10 | 33.10 | 2.80 |
| Nancy Pelosi | Any | 24.20 | 41.20 | 34.70 | 67.00 | 22.60 | 6.50 |
| Nancy Pelosi | Nickname | 26.80 | 36.80 | 36.50 | 99.40 | 0.50 | 0.10 |
| | Normal | 24.10 | 41.30 | 34.60 | 65.90 | 27.40 | 6.70 |

Figure 1 Classification results (sentiments and bots) for each actor. The high-resolution figure can be zoomed in, using the digital version of this article. [Full-size](#) DOI: 10.7717/peerj-cs.947/fig-1

Sentiment analysis

Analyses shows significant negative sentiment for all key actors (Fig. 1). As expected, when referred to by their nicknames, tweets about the actors were almost unanimously negative. Of the four key actors, Adam Schiff has the most polarizing results, with the highest proportion of positive tweets (9.20%) and the second highest proportion of negative tweets (74.10%). This matches observations during the trial, suggesting that Republicans strongly disliked Schiff for his leading role in the Impeachment Trial, while Democrats supported his efforts. Although the proportions of sentiments expressed vary over time (Fig. 2), we note that negative sentiments dominate on every one of the six dates for data sampling.

Bot detection

Results from bot detection show that at least 22–24% of tweets were sent by bot controlled accounts, with a few examples provided in Table 8. Deleted accounts as well as accounts for which no information was obtained are grouped together in the ‘Unknown’ category. As Twitter will remove bots that manipulate conversations, it is possible that a significant share of the ‘Unknown’ category is also composed of bots. We note a

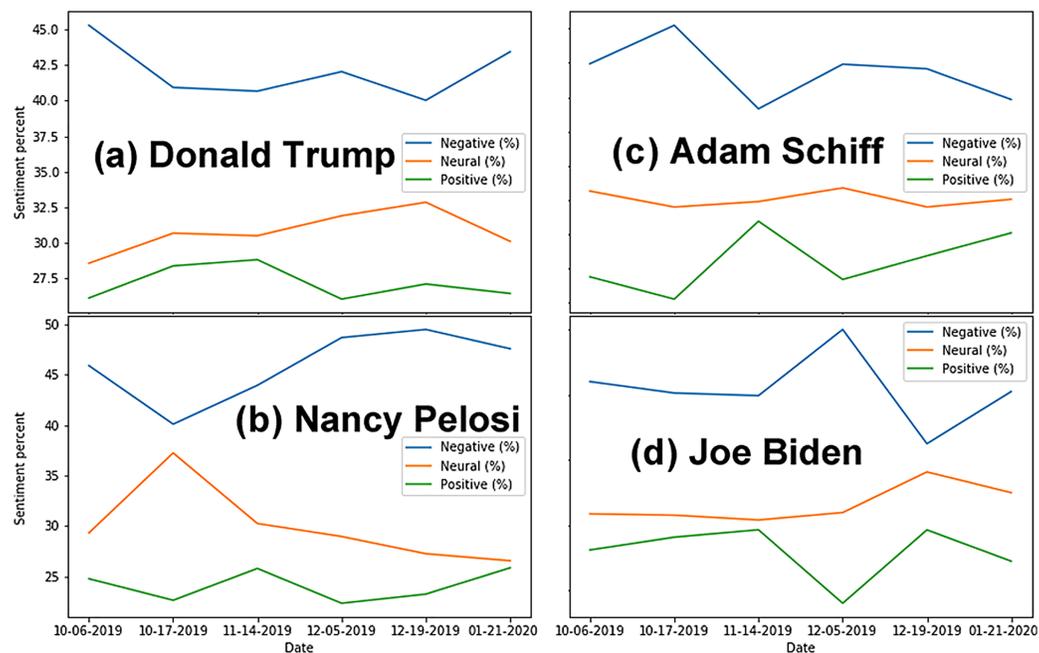


Figure 2 Percentage of tweets per category of sentiments over time for each actor. (A) Donald Trump. (B) Nancy Pelosi. (C) Adam Schiff. (D) Joe Biden. Note that using a *percentage* instead of the absolute number of tweets allows to compare results across actors, since each one is associated with a different volume of tweets. [Full-size !\[\]\(a6fe094a331555a6fd67d72f7f1bf63e_img.jpg\) DOI: 10.7717/peerj-cs.947/fig-2](https://doi.org/10.7717/peerj-cs.947/fig-2)

difference between results for the Republican Mitch McConnell, on the one hand, and the three Democrats (Joe Biden, Nancy Pelosi, and Adam Schiff) on the other hand. For Democrats, bots created at least 26–28% of the more derogatory tweets (*i.e.*, based on nicknames), whereas this percentage falls to 16% when referring to Mitch McConnell by his nickname. Overall, a greater proportion of tweets on Mitch McConnell (by almost 10 percentage points) were made by humans, compared to Democrats. Further analyses on other political figures confirmed that bots targeted Democrats (Jerry Nadler 25.83%, Chuck Schumer 24.40%, Hakeem Jeffries 23.55%) more often than Republicans (Lindsey Graham 22.08%, Pat Cipollone 22.46%, Josh Hawley 15.93%, Stephen Miller 22.14%), with an average difference of four percentage points.

Although the overall tone of the tweets is negative (Fig. 3), the *ratio of negatives-to-positives* shows a clear differentiation in the use of bots compared to humans. For example, consider Adam Schiff (Fig. 3, top-left). Based on tweets by humans, he has a ratio of $\frac{583616}{70666} \approx 8.3$ negatives-to-positives. In comparison, this ratio is $\frac{354241}{36542} \approx 9.7$ for tweets authored by bots. Overall, we observe that the only cases in which bots had a greater negative-to-positive ratio of posts than humans was for Democrats: Adam Schiff (9.7 negative tweets by bot for each positive *vs.* 8.3 for humans), Chuck Schumer (40.2 *vs.* 33.4), Joe Biden (39.1 *vs.* 33.8), and Nancy Pelosi (10.9 *vs.* 10.2). In contrast, bots were always less negative than humans for Republicans: Donald Trump (6.1 for bots *vs.* 7.9 for humans), Lindsey Graham (8.1 *vs.* 9.0), Mitch McConnell (19.7 *vs.* 23.6), and Stephen Miller (59.6 *vs.* 103.8).

Table 8 Two sample tweets for each day of data collection.

| Date | Tweets |
|------------------|--|
| 5 December 2019 | RT @usa4_trump: Ukraine Pulls Back Curtain On Biden – Claims Burisma Paid The Vice Presi... https://t.co/lhGlqOaNhn via @YouTube HERE IT CO... @GOP @RepMikeJohnson Too much evidence to ignore. Trump won't even let staff testify. No transparency, no innocence! |
| 19 December 2019 | Merry Christmas Eve Patriots! I have a grateful heart for President Donald Trump, a man who loves our God and our country, and defends our people. Keep the faith Patriots, and know that we are on the right path. God Bless you all. #TRUMP2020Landside https://t.co/y5duaB6wrR @RyanHillMI @realDonaldTrump You're right. It wasn't a trial. It was a perfect example of tyranny. But, tweets can only be so long and impeachment didn't fit. So, trial. |
| 6 October 2019 | @AndyOstroy @realDonaldTrump What Law has been broken? What rule? Why are they not taking a formal vote to begin impeachment? What crime or misdemeanors took place? Pelosi cant articulate the actual crime! Schiff is 4–50 pinnochios deep in 3 years! Jim Jordan Doesn't 'Think' Trump Did Something He Actually Did On Camera https://t.co/TjXU4chJ5T |
| 17 October 2019 | Survey: 54 percent of Americans support Trump impeachment inquiry https://t.co/0eYzwSn03l : Hillary beat Trump by a majority of votes. Polls show, that disgruntled Hillary voters, want the W.H. back, there will always be a majority against Trump ??? RT @weavejenn: Cause he's a spineless coward #TrumpIsADisgrace https://t.co/6RhP0Ufld9 |
| 14 November 2019 | @CumeTalitha trump is our god and lord and savior. dont say gods name in vein. i will report you. @realDonaldTrump Adam Schiff is the fake whistle blower! No one exists all made up! And what are they whistle blowing we have the transcripts end of story. |
| 21 January 2020 | @CHHR01 Pence also tried to oust Trump from the ticket in 2016. They even created a website, but then abandoned the effort and threatened the life of the web designer. I will find that story for you. @realDonaldTrump FATHER TODAY I ASK FOR A MIRACLE FOR PRWSIDENT TRUMP LET HIM SEE ALL HIS ENEMYS HANDCUFFED ESCORTED TO GITMO4 TREASON IN JESUSNAME.REMOVE SOROS ALL THE DEMOCRATES WHO WE SEE DESTROYING CALI NEW YORK CHICAGO ALL STATES/CITYS(ISA41:11-16)JNA |

Note:

Some links may have broken since the study was conducted, which is one indicator that they were propaganda posts.

We did not find evidence of a clear temporal trend (*e.g.*, monotonic increase or decrease) of bots over time (Fig. 4), which may suggest that they were employed based on specific events.

Top sources used by bots

A total of 1,171 distinct websites were used as sources across 54,899 tweets classified as written by bot accounts. The complete list is provided as part of our supplementary online materials (<https://osf.io/3nsf8/>). The distribution of sources is heavily imbalanced. The vast majority ($n = 909$, 77.62%) of these websites were used less than 10 times each, while the top 10 websites were used 32,050 times (*i.e.*, in 58.38% of tweets with URLs). The most cited website is Twitter (19,258 tweets, 35.08%), as expected by the practice of retweets. Several other websites are social media (YouTube in 1,884 tweets or 3.43%, Facebook in 266 tweets or 0.48%, Periscope in 215 tweets or 0.39%), web hosting solutions (WordPress in 111 tweets or 0.20%, Change.org in 106 tweets or 0.19%, Blogspot in 101 tweets or 0.18%), or search engines (Civiqs in 65 tweets or 0.12%, Google in 65 tweets or 0.12%). The remainder are primarily websites that propose news, hence they were evaluated with respect to the bias ratings from AllSides and Ad Fontes.

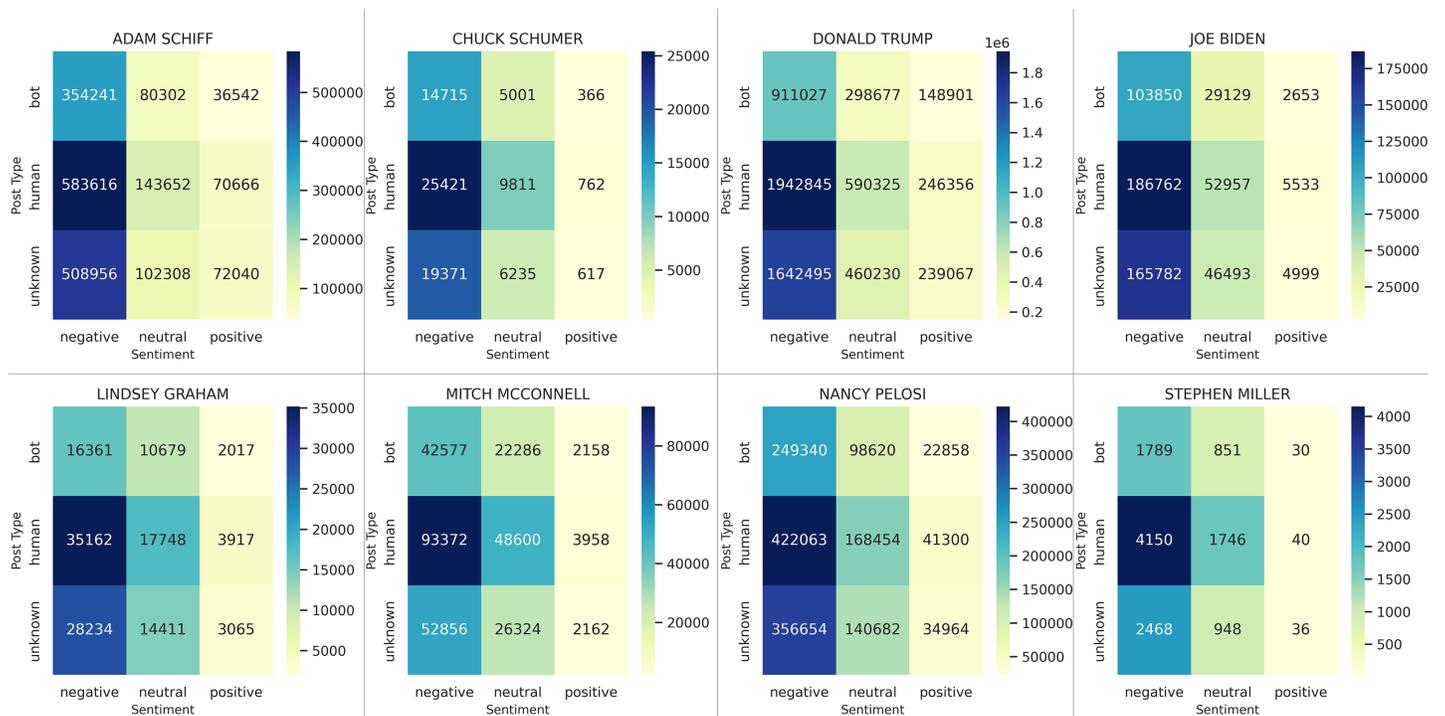


Figure 3 Joint examination of sentiments and bots for each major actor, with additional political figures included. The high-resolution figure can be zoomed in, using the digital version of this article. [Full-size !\[\]\(365da8a2cbf7f1d19047f927ee6f7f2e_img.jpg\) DOI: 10.7717/peerj-cs.947/fig-3](https://doi.org/10.7717/peerj-cs.947/fig-3)

Given the large number of distinct websites, we focused on those used at least 50 times, excluding the categories aforementioned (social media, web hosting, search engines). When using AllSides, the sample *appears* relatively balanced (Table 9), because a large number of websites are not rated. Ad Fontes provides more evaluations, which starts to show that websites lean to the right. We complemented the ratings of Ad Fontes with manual evaluations (Table 9), showing that (i) there are almost twice as many right-leaning sources than left-leaning sources and (ii) the right-leaning sources are much more extreme. For example, 41 websites are more right-leaning than FoxNews while only 16 were more left-leaning than CNN. The imbalance is even more marked when we consider the volume of tweets. Indeed, out of the 10 most used websites, eight lean to the right and two to the left (italicized): FoxNews (3.73% of tweets with sources), Trending Politics (2.80%), Breitbart (2.60%), The Right Scoop (2.49%), The Gateway Pundit (2.48%), Sara A Carter (2.19%), Wayne Dupree (1.86%), CNN (1.69%), *Raw Story* (1.57%), Kerry Picket (1.41%). Tweets associated with some of these websites are provided in Table 10 as examples. In sum, there is ample evidence that the bots tend to promote a Republican viewpoint, often based on misleading sources.

Our findings are similar to *Tripodi & Ma (2022)*, who noted that almost two thirds of the sources in official communication from the White House under the Trump Administration relied on “RWME, a network fueled by conspiracy theories and fringe personalities who reject normative journalistic practices.” Although other news outlets

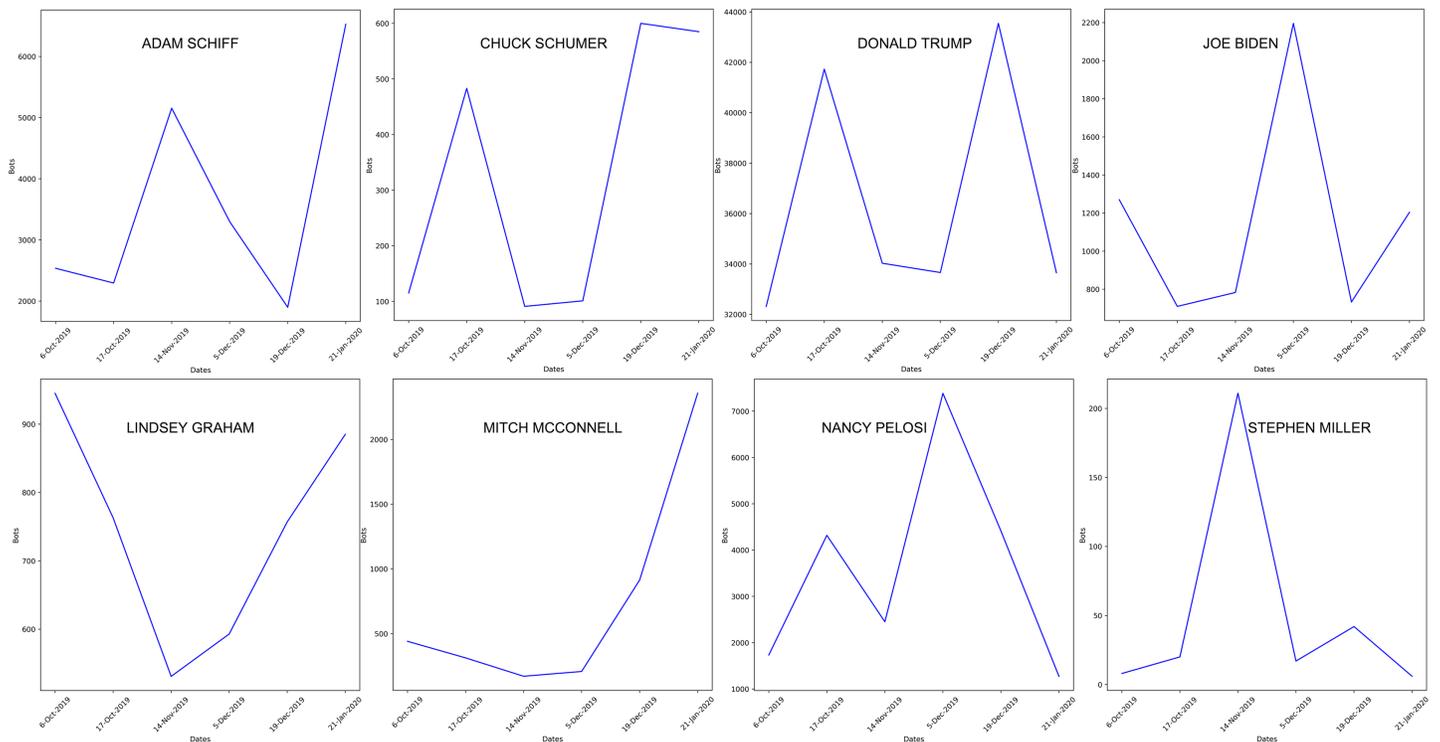


Figure 4 Temporal trend in the number of posts by bots for each major actor, with additional political figures included. The high-resolution figure can be zoomed in, using the digital version of this article. [Full-size !\[\]\(ebfe6d37ad86655679811e032f633da4_img.jpg\) DOI: 10.7717/peerj-cs.947/fig-4](https://doi.org/10.7717/peerj-cs.947/fig-4)

Table 9 Website categories per source for bias rating (AllSides, Ad Fontes), using their nomenclature (e.g., ‘very left’ in AllSides, ‘hyper-partisan left’ in Ad Fontes). *Manual screening was applied to websites not rated by Ad Fontes.

| Leaning | AllSides | Ad Fontes | Manual* | Ad Fontes + Manual | Overall |
|---------------------------------|----------|-----------|---------|--------------------|-----------|
| Not rated | 40 | 16 | | | |
| Very left/hyper-partisan left | 9 | 6 | | 6 | Left: 25 |
| Left | 12 | 19 | | 19 | |
| Center/balanced | 8 | 13 | 1 | 14 | |
| Right | 5 | 16 | 1 | 17 | Right: 46 |
| Very right/hyper-partisan right | 14 | 16 | | 16 | |
| Most extreme right | | 2 | 11 | 13 | |
| No longer accessible | | | 3 | 3 | |

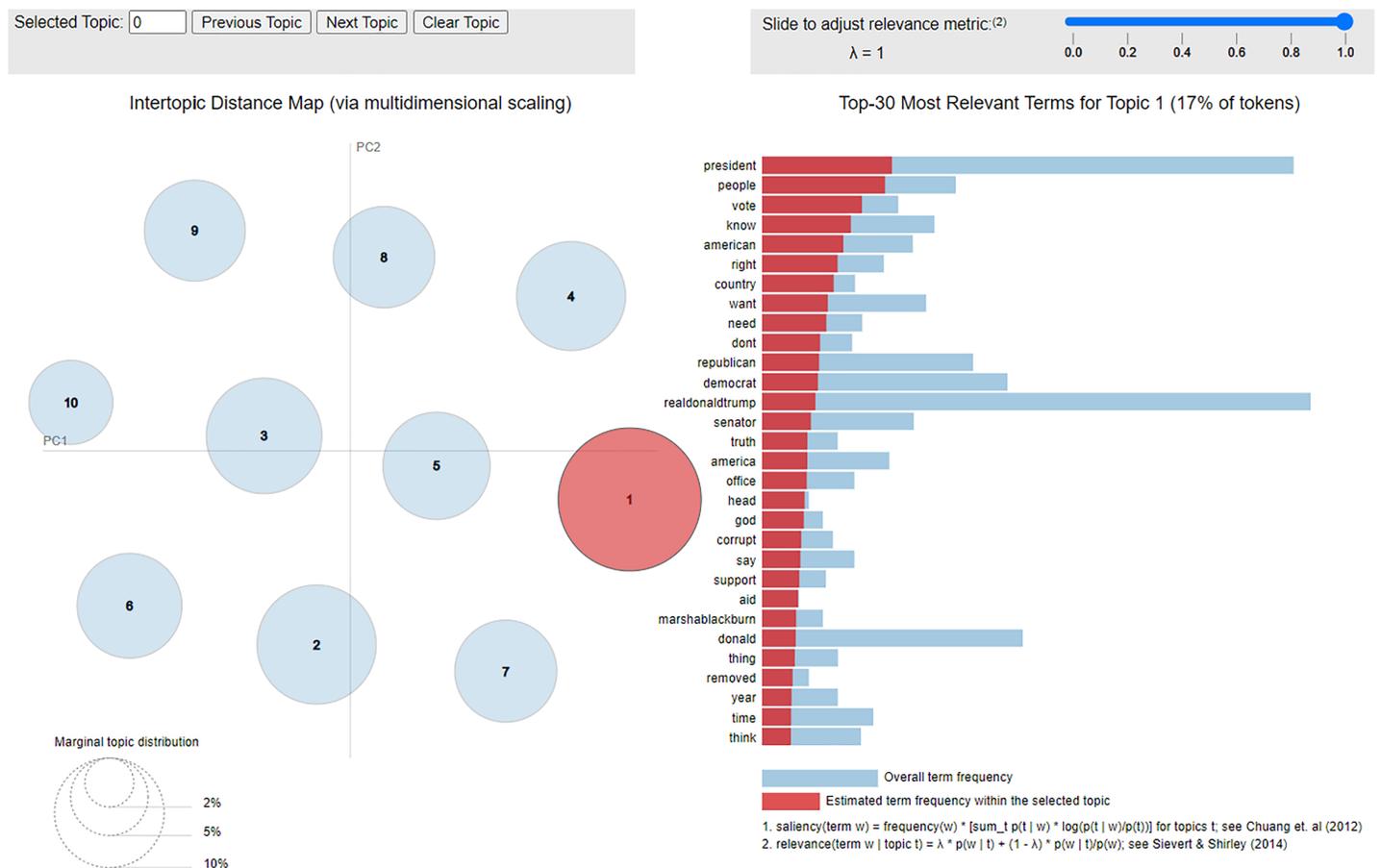
were cited, such as the Washington Post and CNN, this was often in the context of characterizing them as ‘fake news’. Other recent analyses have also noted that the use of popular far-right websites was more prevalent in counties that voted for Trump in 2020 (*Chen et al., 2022*). The possibility for websites that fuel partisanship and anger to be highly profitable may have contributed to their proliferation (*Vorhaben, 2022*).

Table 10 Sample tweets pointing to hyper-partisan right or extreme right websites.

| Website | Tweet |
|----------------------|---|
| rightwingtribune.com | Hes one to ca'll the kettle black, Where's the whistle blower there Schifty?? You said there was, then he wasn't, you werent clear either, but Pelosi had your backside, didn't she? You, all the other DemocrATS. @realDonaldTrump @GodLovesUSA1 @FogCityMidge https://t.co/6jihkrJrXM |
| twitchy.com | Straightforward from here! Vox sounds pretty convinced that impeachment could ultimately lead to President Nancy Pelosi https://t.co/qFUY9aPyKX via @twitchyteam |
| thegatewaypundit.com | RT @gatewaypundit: REMINDER: PELOSI AND SCHIFF BOTH CONNECTED TO UKRAINIAN ARMS DEALER @JoeHoft https://t.co/bTlxwGZ4gy via @gatewaypundit |
| therightscoop.com | RT @marklevinshow: Fascistic Nancy Pelosi, also the dumbest speaker https://t.co/PbCWnWGeHs |
| saraacarter.com | RT @SaraCarterDC: Trump Calls Speaker Pelosi 'A Third-Rate Politician' https://t.co/9jd3MzOuH9 |
| bongino.com | This Says it All: Watch Chairman Nadler Fall ASLEEP During Impeachment Hearing Dan Bongino https://t.co/LySASgD8vF |

Note:

Some links may have broken since the study was conducted, which is one indicator that they were propaganda posts.

**Figure 5** Interactive visualization for topics from bots on 21 January 2020.

Full-size DOI: 10.7717/peerj-cs.947/fig-5

Themes for humans vs. bots

Interactive visualizations for themes are provided as part of our supplementary online materials (<https://osf.io/3nsf8/>), with one example shown in Fig. 5. Themes should be

interpreted *with caution*. For example, consider posts on 6 December 2019. For bots, a major theme links Senator Nancy Pelosi with ‘hate’, whereas for humans she is first associated with impeachment, house, president, and then only with hate. However, bots are not actually conveying images of hatred *towards* Pelosi. Rather, the connection is related to a news story that occurred the day before, in which a reporter asked Pelosi if she hated Trump, and she responded that she did not hate anyone; the reaction went viral. Although bots have picked up on it more than the humans, it is only tangentially related to impeachment and does not indicate that bots are hating Pelosi.

A more robust and relevant pattern, across the entire time period, is that Biden is a much more salient term for the humans than the bots. Based on associated terms, it appears to be connected to Trump and other Republicans (particularly Lindsey Graham and outlets such as Fox News) trying to distract from impeachment by making allegations that Biden was corrupt in his dealings with Ukraine. It thus appears that bots and humans occasionally focus on different stories.

DISCUSSION

Throughout the process, both parties and the President tried to control the public narrative over impeachment. Early in the process, Democrats began to talk about abuse of power or corruption by Trump and eventually shifted to the more specific language of bribery and extortion. Meanwhile, Republican elected officials seemed to be trying to frame this as a coup by the Democrats. Republicans directed their ire at House Intelligence Committee Chairman Democrat Adam Schiff who led the impeachment inquiry. As summarized by Barberio in the context of power scandals, “Trump’s approach to scandal management included the contemplation of the use of backfires and their full deployment” ([Barberio, 2020](#)). The President and some of his Republican allies thus tried to divert attention from the President by alleging misconduct by Joe and Hunter Biden. With the 2020 Democratic presidential primary campaigns well underway, Trump saw an opportunity to undermine Biden’s presidential campaign by raising doubts about Biden’s integrity. In early October, the Trump campaign demonstrated a willingness to influence perceptions *via* social media, by spending “as much as \$21,000” on a Facebook ad entitled “Biden Corruption”, which included debunked claims that then Vice President Joe Biden had tied Ukrainian aid to dropping an investigation of his son ([Grynbaum & Hsu, 2019](#)).

Throughout the process, the President denigrated Democrats as a group and individually in his public comments and on Twitter. Strategies have included the use of nicknames, which contributes to painting a gallery of ‘known villains’ ([Montgomery, 2020](#)) and fits within the broader pattern of the far right in online harassment of political enemies ([Bambenek et al., 2022](#)). This use of nicknames also echoes Trump’s strategy in employing derogatory language when depicting other groups, such as the frequent use of terms such as ‘animal’ or ‘killer’ when referring to immigrants ([Bilewicz & Soral, 2022](#)). Such strategies are part of a broader trend of incivility by American politicians on Twitter, with uncivil tweets serving as powerful means of political mobilization and fundraising ([Ballard et al., 2022](#)) by drawing on emotions such as anger ([Joosse & Zelinsky, 2022](#); [Bernecker, Wenzler & Sassenberg, 2019](#)); the ensuing attention may even lead politicians to engage

into greater incivility (Frimer et al., 2022). In the case of Trump, the violence-inducing rhetoric resulted in permanent suspension of his Twitter account, days before being charged by the House of Representatives for “incitement of insurrection” (Wheeler & Muwanguzi, 2022).

Previous research on Trump and social media mining concluded that “political troll groups recently gained spotlight because they were considered central in helping Donald Trump win the 2016 US presidential election” (Flores-Saviaga, Keegan & Savage, 2018). Far from isolated individuals with their own motives, research has shown that ‘trolls’ could be involved in coordinated campaigns to manipulate public opinion on the Web (Zannettou et al., 2019). Given the polarization of public opinions and the “hyper-fragmentation of the mediascape”, such campaigns frequently tailor their messages to appeal to very specific segments of the electorate (Raynauld & Turcotte, 2022). A growing literature also highlights that the groups involved in these campaigns are occasionally linked to state-funded media (Sloss, 2022), as seen by the strategic use of Western platforms by Chinese state media to advance ‘alternative news’ (Liang, 2021) with a marked preference for positives about China (Wu, 2022) or the disproportionate presence of bots among followers of Russia Today (now RT) (Crilley et al., 2022).

In this paper, we pursued three questions about bots: their overall involvement (Q1), their targets (Q2), and the sources used (Q3). We found that bots were actively used to influence public perceptions on online social media (Q1). Posts created by bots were almost exclusively negative and they targeted Democrat figures more than Republicans (Q2), as evidenced both by a difference in ratio (bots were more negative on Democrats than Republicans) and in composition (use of nicknames). While the effect of a presidential tweet has been debated (Miles & Haider-Markel, 2020), there is evidence that such an aggressive use of Twitter bots enhances political polarization (Gorodnichenko, Pham & Talavera, 2021). Consequently, the online political debate was not so much framed as an exchange of ‘arguments’ or support of various actors, but rather as a torrent of negative posts, heightened by the presence of bots. The sources used by bots confirm that we are not in the presence of arguments supported by factual references; rather, we witnessed a large use of heavily biased websites, disproportionately promoting the views of partisan or extreme right groups (Q3). Since social media platforms are key drivers of traffic towards far-right websites (together with search engines) (Chen et al., 2022), it is important to note that many of the links towards such websites are made available to the public through bots. Although our findings may point to the automatic removal of bots from political debates as a possible intervention (Cantini et al., 2022), there is evidence that attempts at moderating content can lead to even greater polarization (Trujillo & Cresci, 2022), hence the need for caution when intervening.

While previous studies found a mostly anti-Trump sentiment in human responses to tweets (Roca-Cuberes & Young, 2020), the picture can become different when we account for the sheer volume of bots and the marked preference in their targets. This use of bots was not an isolated incident, as the higher use of bots favoring Trump compared to

Clinton was already noted during the second U.S. Presidential Debate ([Howard, Kollanyi & Woolley, 2016](#)), with estimates as high as three pro-Trump bots for every pro-Clinton bot ([Ferrara, 2020](#)). Our analysis thus provides further evidence for the presence of ‘computational propaganda’ in the US.

There are three limitations to this work. First, it is possible that tweets *support* an individual despite being *negative*. However, this effect would be more likely for Republicans than Democrats, which does not alter our conclusions about the targeted use of bots. Indeed, Trump’s rhetoric has been found to be “unprecedentedly divisive and uncivil” ([Brewer & Egan, 2021](#)), which can be characterized by a definite and negative tone ([Savoy, 2021](#)). Consequently, some of the posts categorized as negatives may not be *against* Trump, but rather amplify his message and rhetoric to *support him*. The prevalence of such cases may be limited, as a recent analysis (over the last two US presidential elections) showed that the prevalence of negative posts was associated with lower popular support ([Shah, Li & Hadaegh, 2021](#)).

Second, although we observe an overwhelmingly negative tone in posts from bots, there is still a *mixture* of positive and negative tweets. This mixture may reflect that the use of bots in the political sphere is partly a state-sponsored activity. That is, various states intervene and their different preferences produce a heterogeneous set of tweets ([Kießling et al., 2019](#); [Stukal et al., 2019](#)). Although studies on centrally coordinated disinformation campaigns often focus on how one specific entity (attempts to) shape the public opinion ([Keller et al., 2020](#); [Im et al., 2020](#)), complex events involve multiple entities promoting different messages. For example, analysis of the 2016 election showed that IRA trolls generated messages in favor of Trump, whereas Iranian trolls were against Trump ([Zannettou et al., 2019](#)). It is not currently possible to know with certainty which organizations were orchestrating which bot accounts, hence our study cannot disentangle bot posts to ascribe their mixed messages to different organizations.

Finally, in line with previous large-scale studies in natural language processing (NLP) regarding Trump, we *automatically* categorized each post with respect to sentiment ([Tachaiya et al., 2021](#)). Although this approach may involve manual reading for a small fraction of tweets (to create an annotated dataset that then trains a machine learning model), most of the *corpus* is then processed automatically. Such an automatic categorization allows analysts to cope with a vast amount of data, thus favoring breadth in the pursuit of processing millions of tweets. In contrast, studies favoring depth have relied on a qualitative approach underpinned by a *manual* analysis of the material, which allows examination of the ‘storytelling’ aspect through changes in rhetoric over time ([Phelan, 2021](#)). As our study is rooted in NLP and performed over millions of tweets, we did not manually read them to construct and follow narratives. This may be of interest for future studies, who could use a complementary qualitative approach to contrast the rhetoric of posts from bots and humans, or how bots may shape the nature of the arguments rather than only their polarity. Such mixed methods studies can combine statistical context analysis and qualitative textual analysis ([O’Boyle & Haq, 2022](#)) to offer valuable insights.

CONCLUSIONS

We found that bots have played a significant role in contributing to the overall negative tone of the debate during the first impeachment. Most interestingly, we presented evidence that bots were targeting Democrats more than Republicans. The sources promoted by bots were twice as likely to espouse Republican views than Democrats, with a noticeable presence of highly partisan or even extreme right websites. Together, these findings suggest an intentional use of bots as part of a larger strategy rooted in computational propaganda.

ACKNOWLEDGEMENTS

We thank Z. Li, B. Lee, J. Holt, and D. Smith, for their assistance with this project. We also thank Dr. J. Mueller for supporting our use of the Redhawk Computing Cluster.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Michael C. Galgoczy performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, and approved the final draft.
- Atharva Phatak performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, and approved the final draft.
- Danielle Vinson conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Vijay K. Mago conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Philippe J. Giabbanelli conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The code and data are available at the Open Science Framework (OSF):

Giabbanelli, Philippe J. 2022. "Trump Twitter." OSF. February 17. osf.io/3nsf8.

REFERENCES

- Abilov A, Hua Y, Matatov H, Amir O, Naaman M. 2021.** Voterfraud2020: a multi-modal dataset of election fraud claims on Twitter. *ArXiv preprint*. DOI [10.48550/arXiv.2101.08210](https://doi.org/10.48550/arXiv.2101.08210).
- Alexandre I, Jai-sung Yoo J, Murthy D. 2021.** Make tweets great again: who are opinion leaders, and what did they tweet about Donald Trump? *Social Science Computer Review* **11(6)**:089443932110088 DOI [10.1177/08944393211008859](https://doi.org/10.1177/08944393211008859).

- Alfonzo P. 2021.** A topology of twitter tactics: tracing the rhetorical dimensions and digital labor of networked publics. *Social Media+ Society* 7(2):20563051211025514 DOI 10.1177/20563051211025514.
- Alomari KM, ElSherif HM, Shaalan K. 2017.** Arabic tweets sentimental analysis using machine learning. In: *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Berlin: Springer, 602–610.
- Ballard AO, DeTamble R, Dorsey S, Heseltine M, Johnson M. 2022.** Dynamics of polarizing rhetoric in congressional tweets. *Legislative Studies Quarterly* 41(1):12 DOI 10.1111/lsq.12374.
- Bambenek J, Fodor J, Hausserman S, Hoffman M, Loadenthal M, Thierry M. 2022.** *Cyberhate: the far right in the digital age*. Lanham: Rowman & Littlefield.
- Barberio R. 2020.** President Trump: the conquest of backfire scandal management? In: *Presidents and Political Scandal*. Berlin: Springer, 97–124.
- Bernecker K, Wenzler M, Sassenberg K. 2019.** Tweeted anger predicts county-level results of the 2016 United States Presidential Election. *International Review of Social Psychology* 32(1):6 DOI 10.5334/irsp.256.
- Bilewicz M, Soral W. 2022.** The politics of hate: derogatory language in politics and intergroup relations. In: Osborne D, Sibley CG, eds. *The Cambridge Handbook of Political Psychology*. Cambridge UK, 429–441.
- Boucher J-C, Thies CG. 2019.** I am a tariff man: the power of populist foreign policy rhetoric under president Trump. *The Journal of Politics* 81(2):712–722 DOI 10.1086/702229.
- Brewer EC, Egan C. 2021.** Clear and present danger standard 100th anniversary: examining Donald J. Trump’s “presidential” rhetoric as a clear and present danger. *First Amendment Studies* 55(1):1–15 DOI 10.1080/21689725.2021.1886967.
- Brown Crosby D. 2022.** In the spirit of queen Araweelo: an analysis of congresswoman Ilhan Omar’s disruption of nativism and white supremacy. In: *Get Your Knee Off Our Necks*. Berlin: Springer, 229–251.
- Cantini R, Marozzo F, Talia D, Trunfio P. 2022.** Analyzing political polarization on social media by deleting bot spamming. *Big Data and Cognitive Computing* 6(1):3 DOI 10.3390/bdcc6010003.
- Cervi L, García F, Marín-Lladó C. 2021.** Populism, Twitter, and covid-19: narrative, fantasies, and desires. *Social Sciences* 10(8):294 DOI 10.3390/socsci10080294.
- Chen Z, Chen H, Freire J, Nagler J, Tucker JA. 2022.** Understanding how people consume low quality and extreme news using web traffic data. *ArXiv preprint*. DOI 10.48550/arXiv.2201.04226.
- Christenson D, Kreps S, Kriner D. 2021.** Contemporary presidency: going public in an era of social media: Tweets, corrections, and public opinion. *Presidential Studies Quarterly* 51(1):151–165 DOI 10.1111/psq.12687.
- Crilley R, Gillespie M, Vidgen B, Willis A. 2022.** Understanding RT’s audiences: exposure not endorsement for twitter followers of Russian state-sponsored media. *The International Journal of Press/Politics* 27(1):220–242 DOI 10.1177/1940161220980692.
- Davis C, Varol O, Ferrara E, Flammini A, Menczer F. 2016.** Botornot: a system to evaluate social bots. In: *Proceedings of the 25th International Conference Companion on World Wide Web*. 273–274.
- Dejard L, dos Santos AM, de Mello Junior HD, de Cássia Romeiro Paulino R, Figueiredo K, Costa FAR, Vidal DA, Pires YP, Seruffo M, da César R. 2021.** Social and institutional presence of the presidents of the Americas on social media: an analysis of the communication on twitter about covid-19. *International Journal of Advanced Engineering Research and Science* 8(8):512–526 DOI 10.22161/ijaers.88.58.

- Devlin J, Chang M-W, Lee K, Toutanova K. 2018.** Bert: pre-training of deep bidirectional transformers for language understanding. *ArXiv preprint*. DOI 10.48550/arXiv.1810.04805.
- Driver L. 2021.** Trump's brand is crisis: using situational crisis communication theory and framing theory to examine Trump's twitter posts during four crisis situations. Master's thesis, Department of Communication and Journalism, University of Wyoming, Laramie, Wyoming, USA. Available at <https://search.proquest.com/openview/2531a4387826e31c291ffc17314bbd80/1.pdf?pq-origsite=gscholar&cbl=18750&diss=y>.
- Ferrara E. 2020.** Bots, elections, and social media: a brief overview. In: Shu K, Wang S, Lee D, Liu H, eds. *Disinformation, Misinformation, and Fake News in Social Media*. Berlin: Springer, 95–114.
- Flores-Saviaga C, Keegan B, Savage S. 2018.** Mobilizing the trump train: Understanding collective action in a political trolling community. In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 12.
- Forgette R. 2018.** *News grazers: media, politics, and trust in an information age*. Washington, D.C.: CQ Press.
- Frimer JA, Aujla H, Feinberg M, Skitka L, Aquino K, Eichstaed JC, Willer R. 2022.** Incivility is rising among American politicians on Twitter. *Social Psychology and Personality Science* DOI 10.31234/osf.io/2hku3.
- Golubev A, Loukachevitch N. 2020.** Improving results on Russian sentiment datasets. In: *Conference on Artificial Intelligence and Natural Language*. Berlin: Springer, 109–121.
- Gorodnichenko Y, Pham T, Talavera O. 2021.** Social media, sentiment and public opinions: evidence from# brexit and# uselection. *European Economic Review* 136:103772 DOI 10.1016/j.eurocorev.2021.103772.
- Gould N. 2021.** Donald Trump v. the electorate: a Twitter feud for the ages. Available at https://scholarship.claremont.edu/cmc_theses/2690.
- Grimminger L, Klinger R. 2021.** Hate towards the political opponent: a Twitter corpus study of the 2020 US elections on the basis of offensive speech and stance detection. *ArXiv preprint*. DOI 10.48550/arXiv.2103.01664.
- Grynbaum M, Hsu T. 2019.** *CNN rejects 2 Trump campaign ads, citing inaccuracies*. New York: New York Times.
- Haq EU, Braud T, Kwon YD, Hui P. 2020.** A survey on computational politics. *IEEE Access* 8:197379–197406 DOI 10.1109/ACCESS.2020.3034983.
- Hawkins KA, Kaltwasser CR. 2018.** Measuring populist discourse in the united states and beyond. *Nature Human Behaviour* 2(4):241–242 DOI 10.1038/s41562-018-0339-y.
- Howard P, Kollanyi B, Woolley SC. 2016.** Bots and automation over twitter during the second us presidential debate. Available at <https://blogs.oii.ox.ac.uk/politicalbots/wp-content/uploads/sites/89/2016/10/Data-Memo-Second-Presidential-Debate.pdf>.
- Huszár F, Ktena SI, O'Brien C, Belli L, Schlaikjer A, Hardt M. 2022.** Algorithmic amplification of politics on twitter. *Proceedings of the National Academy of Sciences* 119(1):e202533411 DOI 10.1073/pnas.2025334119.
- Im J, Chandrasekharan E, Sargent J, Lighthammer P, Denby T, Bhargava A, Hemphill L, Jurgens D, Gilbert E. 2020.** Still out there: modeling and identifying Russian troll accounts on twitter. In: *12th ACM Conference on Web Science*. 1–10.
- Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, Zhao L. 2019.** Latent Dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications* 78(11):15169–15211 DOI 10.1007/s11042-018-6894-4.

- Joose P, Zelinsky D. 2022.** Berserk! Anger and the charismatic populism of Donald Trump. *Critical Sociology*. Epub ahead of print 6 January 2022 DOI [10.1177/08969205211063131](https://doi.org/10.1177/08969205211063131).
- Kamps K. 2021.** *Commander-in-Tweet: Donald Trump and the deformed presidency*. Berlin: Springer Nature.
- Keller FB, Schoch D, Stier S, Yang J. 2020.** Political astroturfing on twitter: how to coordinate a disinformation campaign. *Political Communication* **37(2)**:256–280 DOI [10.1080/10584609.2019.1661888](https://doi.org/10.1080/10584609.2019.1661888).
- Khan R, Rustam F, Kanwal K, Mehmood A, Choi GS. 2021.** US based covid-19 tweets sentiment analysis using textblob and supervised machine learning algorithms. In: *2021 International Conference on Artificial Intelligence (ICAI)*. Piscataway: IEEE, 1–8.
- Kießling B, Homburg J, Drozdzyński T, Burkhardt S. 2019.** State propaganda on twitter. In: *Multidisciplinary International Symposium on Disinformation in Open Online Media*. Berlin: Springer, 182–197.
- Kumar TS, Pandian KA, Aara ST, Pandian KN. 2021.** A reliable technique for sentiment analysis on tweets via machine learning and bert. In: *2021 Asian Conference on Innovation in Technology (ASIANCON)*. Piscataway: IEEE, 1–5.
- Le H, Boynton GR, Shafiq Z, Srinivasan P. 2019.** A postmortem of suspended twitter accounts in the 2016 US presidential election. In: *International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. Piscataway: IEEE, 258–265.
- Levendusky M. 2013.** *How partisan media polarize America*. Chicago: University of Chicago Press.
- Liang F. 2021.** Globalizing propaganda: examining the practice and regulation of China’s rise on Facebook and Twitter. PhD thesis, University of Michigan.
- Long S. 2022.** White identity, Donald Trump, and the mobilization of extremism. *Politics, Groups, and Identities* **9**:1–29 DOI [10.1080/21565503.2022.2025868](https://doi.org/10.1080/21565503.2022.2025868).
- Lui M, Baldwin T. 2014.** Accurate language identification of twitter messages. In: *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)*. 17–25.
- Machus T, Mestel R, Theissen E. 2022.** Heroes, just for one day: the impact of Donald Trump’s tweets on stock prices. *Journal of Behavioral and Experimental Finance* **33**:100594 DOI [10.1016/j.jbef.2021.100594](https://doi.org/10.1016/j.jbef.2021.100594).
- McKee SC, Evans HK, Clark JH. 2022.** The “perfect” call: congressional representation by tweet in the midst of the Ukraine whistleblower scandal. *American Politics Research* **50(1)**:30–44 DOI [10.1177/1532673X211041922](https://doi.org/10.1177/1532673X211041922).
- Mendhe C, Henderson N, Srivastava G, Mago V. 2020.** A scalable platform to collect, store, visualize, and analyze big data in real time. *IEEE Transactions on Computational Social Systems* **8(4)**:260–269 DOI [10.1109/TCSS.2020.2995497](https://doi.org/10.1109/TCSS.2020.2995497).
- Miles M, Haider-Markel D. 2020.** Polls and elections Trump, twitter, and public dissuasion: a natural experiment in presidential rhetoric. *Presidential Studies Quarterly* **50(2)**:436–450 DOI [10.1111/psq.12615](https://doi.org/10.1111/psq.12615).
- Milford M. 2021.** Trump’s impeachment: cheap grace and the comic frame. *Southern Communication Journal* **86(4)**:1–13 DOI [10.1080/1041794X.2021.1925952](https://doi.org/10.1080/1041794X.2021.1925952).
- Montgomery M. 2020.** Populism in performance? Trump on the stump and his audience. *Journal of Language and Politics* **19(5)**:733–765 DOI [10.1075/jlp](https://doi.org/10.1075/jlp).
- Morales EA, Schultz CJP, Landreville KD. 2021.** The impact of 280 characters: an analysis of Trump’s tweets and television news through the lens of agenda building. *Electronic News* **15(1–2)**:21–37 DOI [10.1177/19312431211028610](https://doi.org/10.1177/19312431211028610).

- Nabil M, Aly M, Atiya A. 2015.** ASTD: Arabic sentiment tweets dataset. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 2515–2519.
- Negara ES, Triadi D, Andryani R. 2019.** Topic modelling twitter data with latent Dirichlet allocation method. In: *2019 International Conference on Electrical Engineering and Computer Science (ICECOS)*. Piscataway: IEEE, 386–390.
- O’Boyle J, Haq S. 2022.** Twitter images across boundaries: comparing the use of images in political posts from six nations. *The Agenda Setting Journal* 5(2):219–242 DOI 10.1075/asj.20004.oba.
- Ostrowski DA. 2015.** Using latent Dirichlet allocation for topic modelling in twitter. In: *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)*. Piscataway: IEEE, 493–497.
- Ouyang Y, Waterman R. 2020.** *Trump, Twitter, and the American Democracy*. Berlin: Springer.
- Peisenieks J, Skadiņš R. 2014.** Uses of machine translation in the sentiment analysis of tweets. Amsterdam: IOS Press, 126–131.
- Pérez-Curiel C, Rivas-de Roca R, García-Gordillo M. 2021.** Impact of Trump’s digital rhetoric on the US elections: a view from worldwide far-right populism. *Social Sciences* 10(5):152 DOI 10.3390/socsci10050152.
- Phelan J. 2021.** Donald J. Trump’s storytelling, May 12–June 7, 2020; or, can his saying make things so? *Narrative* 29(2):275–295 DOI 10.1353/nar.2021.0015.
- Pozen DE, Talley EL, Nyarko J. 2019.** A computational analysis of constitutional polarization. *Cornell Law Review* 105:1.
- Qudar M, Mago V. 2020.** Tweetbert: a pretrained language representation model for twitter text analysis. *ArXiv preprint*. DOI 10.48550/arXiv.2010.11091.
- Raynauld V, Turcotte A. 2022.** Replicating the 2016 “lightning in a bottle” political moment: Biden, Trump, and winning the US presidency. In: *Political Marketing in the 2020 US Presidential Election*. Berlin: Springer, 21–40.
- Reyes A, Ross A. 2021.** From the white house with anger: conversational features in president Trump’s official communication. *Language & Communication* 77(4):46–55 DOI 10.1016/j.langcom.2020.12.003.
- Roca-Cuberes C, Young A. 2020.** Vox Populi? Trump’s twitter page as public forum. *Social Sciences* 9(12):226 DOI 10.3390/socsci9120226.
- Russell A. 2021.** *Tweeting is leading: how senators communicate and represent in the age of Twitter*. Oxford: Oxford University Press.
- Rustam F, Khalid M, Aslam W, Rupapara V, Mehmood A, Choi GS. 2021.** A performance comparison of supervised machine learning models for covid-19 tweets sentiment analysis. *PLOS ONE* 16(2):e0245909 DOI 10.1371/journal.pone.0245909.
- Sadia K, Basak S. 2021.** Sentiment analysis of covid-19 tweets: how does bert perform? In: *Proceedings of International Joint Conference on Advances in Computational Intelligence*. Berlin: Springer, 407–416.
- Şahin O, Johnson R, Korkut U. 2021.** Policy-making by tweets: discursive governance, populism, and Trump presidency. *Contemporary Politics* 27(5):591–610 DOI 10.1080/13569775.2021.1935009.
- Sandhu M, Danielle Vinson C, Mago VK, Giabbanelli PJ. 2019.** From associations to sarcasm: mining the shift of opinions regarding the supreme court on twitter. *Online Social Networks and Media* 14:100054 DOI 10.1016/j.osnem.2019.100054.

- Sandhu M, Giabbanelli P, Mago V. 2019.** From social media to expert reports: the impact of source selection on automatically validating complex conceptual models of obesity. In: *International Conference on Human-Computer Interaction*. Berlin: Springer, 434–452.
- Şaşmaz E, Tek FB. 2021.** Tweet sentiment analysis for cryptocurrencies. In: *2021 6th International Conference on Computer Science and Engineering (UBMK)*. Piscataway: IEEE, 613–618.
- Savoy J. 2021.** Stylometric analysis of Trump’s tweets. *Digital Scholarship in the Humanities*. Epub ahead of print 19 May 2021 DOI 10.1093/lc/fqab048.
- Schier SE. 2022.** *The Trump effect: disruption and its consequences in US politics and government*. Lanham: Rowman and Littlefield Publishers.
- Sellers P. 2009.** *Cycles of spin: strategic communication in the US congress*. Cambridge: Cambridge University Press.
- Shah D, Li Y, Hadaegh A. 2021.** Twitter based sentiment analysis of each presidential candidate using long short-term memory. *International Journal of Computer Science and Security* 15:87–96.
- Singh M, Jakhar AK, Pandey S. 2021.** Sentiment analysis on the impact of coronavirus in social life using the bert model. *Social Network Analysis and Mining* 11(1):1–11 DOI 10.1007/s13278-021-00737-z.
- Sloss DL. 2022.** *Tyrants on twitter: protecting democracies from information warfare*. Stanford: Stanford University Press.
- Spagnuolo I. 2019.** Analyzing bot activity on social media over political discourse. the case of Brexit. Available at <http://hdl.handle.net/10589/150484>.
- Stukal D, Sanovich S, Tucker JA, Bonneau R. 2019.** For whom the bot tolls: a neural networks approach to measuring political orientation of twitter bots in Russia. *Sage Open* 9(2):2158244019827715 DOI 10.1177/2158244019827715.
- Tachaiya J, Gharibshah J, Esterling KE, Faloutsos M. 2021.** Raffman: measuring and analyzing sentiment in online political forum discussions with an application to the Trump impeachment. *Proceedings of the International AAAI Conference on Web and Social Media* 15(1):703–713.
- Trautman L. 2019.** Impeachment, Donald Trump and the attempted extortion of Ukraine. *Pace Law Review* 40:141.
- Tripodi F, Ma Y. 2022.** You’ve got mail: how the Trump administration used legislative communication to frame his last year in office. *Information, Communication & Society* 7(4):1–21 DOI 10.1080/1369118X.2021.2020873.
- Trujillo A, Cresci S. 2022.** Make reddit great again: assessing community effects of moderation interventions on r/the_donald. *ArXiv preprint*. DOI 10.48550/arXiv.2201.06455.
- Van Dijk TA. 2009.** Critical discourse studies: a sociocognitive approach. *Methods of Critical Discourse Analysis* 2(1):62–86.
- Vinson C. 2017.** *Congress and the media: beyond institutional power*. Oxford: Oxford University Press.
- Vorhaben E. 2022.** Technology amplified disinformation, now it must demonetize it. *Chicago Policy Review*. Available at <https://chicagopolicyreview.org/2022/01/18/technology-amplified-disinformation-now-it-must-demonetize-it/>.
- Wells C, Shah D, Lukito J, Pelled A, Pevehouse JCW, Yang JH. 2020.** Trump, twitter, and news media responsiveness: a media systems approach. *New Media & Society* 22(4):659–682 DOI 10.1177/1461444819893987.
- Wheeler MB, Muwanguzi S. 2022.** Mapping intersections of politics and information in twenty-first-century cyberculture. *The Library Quarterly* 92(1):24–38 DOI 10.1086/717231.

- Wu P. 2022.** China's propaganda strategy: evidence from state-owned news networks' English-language twitter posts. Available at <https://psyarxiv.com/78239/>.
- Xu W, Sasahara K. 2021.** Characterizing the roles of bots on twitter during the covid-19 infodemic. *Journal of Computational Social Science*. Epub ahead of print 30 August 2021 DOI [10.1007/s42001-021-00139-3](https://doi.org/10.1007/s42001-021-00139-3).
- Yarchi M, Baden C, Kligler-Vilenchik N. 2021.** Political polarization on the digital sphere: a cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication* **38**(1–2):98–139 DOI [10.1080/10584609.2020.1785067](https://doi.org/10.1080/10584609.2020.1785067).
- Yenduri G, Rajakumar B, Praghash K, Binu D. 2021.** Heuristic-assisted bert for twitter sentiment analysis. *International Journal of Computational Intelligence and Applications* **20**(3):2150015 DOI [10.1142/S1469026821500152](https://doi.org/10.1142/S1469026821500152).
- Zannettou S, Caulfield T, Setzer W, Sirivianos M, Stringhini G, Blackburn J. 2019.** Who let the trolls out? Towards understanding state-sponsored trolls. In: *Proceedings of the 10th ACM Conference on Web Science*. New York: ACM, 353–362.
- Zhang L, Fan H, Peng C, Rao G, Cong Q. 2020.** Sentiment analysis methods for HPV vaccines related tweets based on transfer learning. In: *Healthcare*. Vol. 8. Basel: Multidisciplinary Digital Publishing Institute, 307.
- Zhang Y, Wells C, Wang S, Rohe K. 2018.** Attention and amplification in the hybrid media system: the composition and activity of Donald Trump's twitter following during the 2016 presidential election. *New Media & Society* **20**(9):3161–3182 DOI [10.1177/1461444817744390](https://doi.org/10.1177/1461444817744390).
- Zheng X, Wang X, Li Z, Jing R, Xu S, Wang T, Li L, Zhang Z, Zhang Q, Jiang H, Guo Z, Zhang X, Wang FY. 2021.** Donald J. Trump's presidency in cyberspace: a case study of social perception and social influence in digital oligarchy era. *IEEE Transactions on Computational Social Systems* **8**(2):279–293 DOI [10.1109/TCSS.2021.3063167](https://doi.org/10.1109/TCSS.2021.3063167).