TrollHunter2020: Real-time Detection of Trolling Narratives on Twitter During the 2020 U.S. Elections

Peter Jachim
DePaul University
Chicago, IL
pjachim@depaul.edu

Filipo Sharevski
DePaul University
Chicago, IL
fsharevs@cdm.depaul.edu

Emma Pieroni
DePaul University
Chicago, IL
epieroni@depaul.edu

ABSTRACT

This paper presents *TrollHunter2020*, a real-time detection mechanism we used to hunt for trolling narratives on Twitter during and in the aftermath of the 2020 U.S. elections. Trolling narratives form on Twitter as alternative explanations of polarizing events with the goal of conducting information operations or provoking emotional responses. Detecting trolling narratives thus is an imperative step to preserve constructive discourse on Twitter and remove the influx of misinformation. Using existing techniques, the detection of such content takes time and a wealth of data, which, in a rapidly changing election cycle with high stakes, might not be available. To overcome this limitation, we developed *TrollHunter2020* to hunt for trolls in real-time with several dozen trending Twitter topics and hashtags corresponding to the candidates' debates, the election night, and the election aftermath. *TrollHunter2020* utilizes a correspondence analysis to detect meaningful relationships between the top nouns and verbs used in constructing trolling narratives while they emerge on Twitter. Our results suggest that the *TrollHunter2020* indeed captures the emerging trolling narratives in a very early stage of an unfolding polarizing event. We discuss the utility of *TrollHunter2020* for early detection of information operations or trolling and the implications of its use in supporting a constrictive discourse on the platform around polarizing topics.

CCS CONCEPTS

• **Security and privacy → Social aspects of security and privacy**; **Social network security and privacy**;

KEYWORDS

Real-time Troll Detection, Twitter, 2020 U.S. Elections, Correspondence Analysis, Natural Language Processing (NLP)

ACM Reference Format:
Peter Jachim, Filipo Sharevski, and Emma Pieroni. 2021. TrollHunter2020: Real-time Detection of Trolling Narratives on Twitter During the 2020 U.S. Elections. In *Proceedings of the 2021 ACM International Workshop on Security and Privacy Analytics (IWSPA'21), April 28, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3445970.3451158

IWSPA'21, April 28, 2021, Virtual Event, USA
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-8320-2/21/04...$15.00
https://doi.org/10.1145/3445970.3451158

1 INTRODUCTION

The political disruptions in the U.K. and U.S. in 2016 brought widespread attention to the dissemination of false information online [29, 44]. The false or otherwise unverified information enables "alternative narratives" to proliferate on social media about polarizing topics like elections, man-made disasters, or pandemics [34]. Due to a lack of editorial judgement, fact-checking, or third-party filtering, social media in general, and Twitter in particular, are especially conducive to dissemination of alternative narratives [1, 42]. With the possibility of re-tweeting and linking content with hashtags, Twitter also allows for a special case of alternative narratives as part of information operations campaigns with the goal to provoke emotional response from individual users when discussing polarizing topics [6]. We refer to these alternative narratives as *trolling narratives.* Trolling narratives incorporate fake news, conspiracy theories, and rumors [42], but also personal opinions, comments, memes, and provoking hashtags [4, 36].

In response to the fake news campaigns throughout 2016 [2], Twitter began to publish information operations datasets containing trolling narratives from various state-sponsored troll farms [38]. The detection of political information operations, since then, developed into a serious problem because the state-sponsored troll used a wide array of polarized topics to choose from in creating trolling narratives. Usually, moderators review suspected trolling activity to ban/mute trolling users and flagging/deleting trolling content, but this kind of manual solution has some major drawbacks, including a delay of actions, subjectivity of judgment, and scalability [12, 26]. The need for automated trolling detection on Twitter thus drew the attention of the research community yielding various detection approaches [14, 17, 24, 43].

Most of the existing automated trolling detection approaches utilize the information operations datasets from Twitter [38]. Although these approaches give valuable insight, they do not take into account the evolving nature of the trolling narratives with the development and introduction of polarizing topics (e.g. COVID-19 trolling [31]) nor do they consider the readjustment and pivoting of trolling tactics in response to social media removal and misinformation labeling [28]). State-sponsored trolls, aware that there is an active hunt for dissemination of false information on Twitter, will most likely avoid using the trolling narratives and tactics from 2016. Moreover, given the short time span between posting trolling content and being detected/flagged/removed by Twitter, the trolls will utilize trending topics to capture the moment and muddy discourse in real time (e.g. spreading rumors that Scranton, PA is not the real birthplace of Joe Biden during the campaigning in 2020).

To address these discrepancies, we developed a system for *real-time troll hunting* that captures the contextual development of trolling narratives as they are disseminated on Twitter. Our system, called *TrollHunter2020*, leverages a novel application of a correspondence analysis to hunt trolling narratives in real-time during the 2020 U.S. election cycle, throughout the campaigning period as well as election week and transition period. *TrollHunter2020* provides decision support to an analyst using exploratory techniques to uncover hidden patterns in data. Our contribution of a mechanism for early detection of trolling narratives is a direct response to the imperative for increasing understanding of how alternative narratives evolve within the context of polarizing topics. Because*TrollHunter2020* is intended for use in real-time as an arbiter of developing and intense discourses on Twitter, it raises important ethical concerns about its potential (mis)use and we included an ethical treatise of doing real-time trolling detection research.

2 DETECTION OF TROLLING NARRATIVES

2.1 2016 U.S. Elections

The automated trolling detection runs, to a certain extent, counter to social media platforms' goal to allow for a high degree of desired participation and constructive public discourse, making them reluctant to immediately exclude users exhibiting trolling behaviour to avoid perceptions of excessive control and censorship [12]. However, the need for automated detection of trolling narratives is evident to prevent pollution of online discourse and thwart political information operations. Analyzing the state-sponsor trolling linked to the Russian troll farm Internet Research Agency (IRA), a study found that trolls create a small portion of original trolling content (e.g. posts, hashtags, memes, etc.) and heavily engage in retweeting around a certain point in time of interest (e.g. the Brexit referendum) [24]. A detailed investigation into the trolling activity around the 2016 U.S. elections reveals different state-sponsored trolling with varying tactics: IRA trolls were pro-Trump while Iranian trolls were anti-Trump [43]. Authors in [17] trained a classifier on a Twitter-released IRA trolling dataset and tested it on sample of accounts engaging with prominent journalists and were able to distinguish between a troll and a non-troll with a precision of 78.5% and an AUC of 98.9%, under cross-validation. Another trolling detection algorithm analyzing the writing style of the IRA Twitter trolls looked into the emotional, morality, and sentiment changes showing a 0.94 F1 score [14].

The benefits of using Twitter-released trolling narratives as the training dataset – IRA or other state-sponsor trolls – are obvious because they include accounts, tweets, hashtags, and links confirmed by Twitter as part of a political information operations campaign after an investigation into violating their terms of use [28]. This approach is somewhat limited, however, because it aims to detect *trending* trolling narratives based on their *past* behaviour. State-sponsored troll farms are engineered with the objective of persistence in their information operations, and thus, adapt in response to detection and development of new polarized topics. For example, an automated detection tool trained on the Twitter datasets will likely misclassify most of the trolling narratives proliferated during the COVID-19 pandemic because the tweets, hashtags, and tropes revolve around the false information about "China's responsibility for the pandemic," the "status of COVID-19 treatment and vaccine," and the "origin of the COVID-19 virus" as shown in [31]. To account for this limitation, we identified potentially polarizing events in the context of the 2020 U.S election cycle and tracked trolling narratives on Twitter in real-time to observe how these narratives constantly evolve. For this, we continuously updated our dataset as the 2020 US campaigning and election season unfolded, which allowed our mechanism, called *TrollHunter2020*, to capture the *trending* emerging behaviour as it unfolded on Twitter.

2.2 2020 U.S. Elections

The 2020 U.S. election cycle was characterized by an unprecedented division between the political parties during exceptionally trying times of the COVID-19 pandemic [9]. With many people at home, the public discourse around the 2020 U.S. election cycle naturally took place on social media platforms - with Twitter of special interest serving as a *de facto* press center for the incumbent president Donald Trump. With the previous evidence of information operations on Twitter, and the high stakes of the 2020 U.S. election cycle, researchers began collecting corresponding Twitter data [5], available for identifying trends, tropes, bot accounts, and potential information operations. In their work, Ferrara et al. accumulated a massive dataset of tweets to analyse bot/human activity and trolling narratives spread by banned accounts on Twitter [11]. Their findings indicated highly partisan behavior in retweeting among bots and humans, demonstrating that the political discourse on Twitter was self-reinforcing. Using Twitter's un-hashed Banned User dataset, Ferrara et al. found information operations activity from state-sponsored trolls interacting both with left-leaning and right-leaning users. The trolling, or what they call "distorted" narratives that emerged from their dataset were, expectedly, centered around the QAnon, "-gate", and COVID-related conspiracy theories.

In our study, we are also looking for trolling or distorted narratives, but with a focus on how they emerge in real-time, populating during ongoing polarizing events, instead of their association with known conspiracy theories or known/suspected trolling accounts. Our *TrollHunter2020* is focused on real-time monitoring of early trends on Twitter manifested as hashtags, topics, or themes; it requires a modest amount of data as an input of a correspondence analysis, rather then a typical data classification tool. *TrollHunter2020* outputs the relationship between the top nouns and verbs trolls most likely will use to construct the tolling narratives that later could be identified by the solution proposed by Ferrara et al. or the other mechanisms for automatic trolling detection. In a way, *TrollHunter2020* serves as a real-time early detection system of *present, emerging* trolling narratives that can help improve later detection of associated accounts, state-sponsorship, or other details about a particular information operation. To use *TrollHunter2020* one doesn't need to wait for an account to be banned or flagged by Twitter, but instead, can "tune-in" right during an event of interest (e.g. a debate) and retrieve a high-level overview of "how might trolls distort the narrative around this event."

Trolling has also evolved, and strategies for identifying trolling must evolve as well. Unlike in the past, trolling is no longer just an effort by state-sponsored actors to change perceptions in a target country as evidenced in the Twitter-released sets [18, 38], but it has become a strategy used by people in positions of power to misguide possible voters while using their official platforms [13]. This means that a trolling narrative becomes a vital component in a strategy for public relations employed by politicians. In other words, what initially was deemed as "computational propaganda" has morphed into an actual "political propaganda online" [41]. By propagating alternative narratives from positions of power, those narratives become more legitimate, and in turn, gives cadence to regular Twitter users to become active and participate in the development of an alternative narrative, e.g. amplify the political propaganda online, which resembles an organic discourse compared to the trolling narratives manufactured by the state-sponsored troll farms [35]. The seemingly concerted distortion of a narrative, in this form, could naturally evade detection since it does not come from suspected accounts, does not necessarily use trolling hashtags, nor does it depend on linking content from alternative media. Therefore, we concentrated our lookout for emerging trolling narratives on the textual content of the tweets and it's potency to be developed further with the organic Twitter discourse.

3 TROLLHUNTER2020: DESIGN

3.1 Context Challenges

In building *TrollHunter2020*, we set out to leverage the sophisticated analysis characteristic for the troll hunting models presented above, however, our goal was to achieve an acceptable performance without depending on massive datasets and copious amounts of historical data. The design, therefore, requires unique considerations for applying data mining. First, we needed a real-time data, quickly, as it happens, in order to capture the *trolling opportunity* presented by an event – for example the 2020 vice presidential debate – before the trolls are tagged for misinformation [28], their accounts banned [38], or moved to other platforms like 4chan or Parler [42]. Second, we needed to operate on small datasets that can be quickly accumulated while the event is trending on Twitter instead of waiting for Twitter to retroactively investigate election interference [38]. Third, election events like candidate debates usually generate multiple narratives so we needed to be able to distinguish between each narrative, even with a small amount of data. That required a careful choice of the number of tweets to achieve meaningful troll hunting. In the case of *TrollHunter2020*, the dataset size can be as few as several dozen to few hundred tweets.

Fourth, before an event, it was sometimes unclear what we would be looking for as a *trolling narrative*. For example, during the 2020 vice presidential debate, a fly landed on Vice President Pence's head. There is no way for a troll or a troll detection mechanism to prepare *apriori* for that, but *TrollHunter2020* did need to decipher if that would become a dominant trend in the data. This entails use of unsupervised techniques in order to capture the *present trolling narratives* organically emerging from the event instead of using predefined trolling labels. Fifth, the troll hunting analysis needed to run and present meaningful results quickly. That means that the *TrollHunter2020* must (a) avoid experimentation with hyperparameters; and (b) the analysis must be computationally efficient. This also required a careful synchronisation between the ongoing event, the Twitter API rate for tweeter content accumulation, and the corresponding Twitter activity; Right at the beginning of an event people might not have tweeted enough content yet for us to have gained any meaningful insight, but there potentially could be a burst of tweets or hashtags as the event progressed – say after a heated disagreement – that we could not risk missing.

3.2 Data Preparation

TrollHunter2020 uses the `Tweepy` python library to accumulate tweets corresponding to the specific search terms and hashtags listed in Section 3.5 below for each event of interest. This created datasets where each sample represented one tweet. We used the Twitter search API, so every 3 minutes *TrollHunter2020* could search for tweets, iterating through pre-defined lists of search terms (manually defined terms). *TrollHunter2020* then uses the `spacyr` [3] R library to parse the parts of speech from each individual word in each tweet. The `spacyr` broke the tweets into individual words, so each sample became a single word from a single tweet, and for each word parsed the part of speech, along with its lemma (the lemma is a word stem that combines different forms of the same word into a single word, so "lies," "lied," and "lying" all become "lie"). *TrollHunter2020* performs additional text cleaning using the `tidytext` R library [32] to remove all stop words (e.g. "the" or "an").

Additionally, for each analysis *TrollHunter2020* removes specific words that do not provide useful information. For example, in the context of the 2020 vice presidential debate, the noun "question" and the verb "answer" both did not provide much useful information. These words were not necessarily intuitive, and this step, for now, requires manual intervention during an event to achieve a finer data representation. Next,*TrollHunter2020* performs a full outer join for the dataframe itself using the `dplyr` [40] R library to create samples of each noun and verb lemma pair featured in each individual tweet. We filtered the dataset to only include the top 10 verbs and nouns. Finally, *TrollHunter2020* creates a contingency table where each value represents how many times each verb lemma (the samples) appeared with each noun lemma (the variables). This yielded a final dataset for the correspondence analysis only with 100 values.

3.3 Correspondence Analysis

TrollHunter2020 performs the correspondence analysis using the R `FactoMineR` library [21] to hunt for trolling narratives in real-time on Twitter during the 2020 U.S. election cycle. Correspondence analysis is a technique for visualizing data related to a couple of categorical variables. Correspondence analysis provides "a window onto the data, allowing researchers easier access to their numerical results and facilitating discussion of the data and possibly generating hypotheses which can be formally tested at a later stage" [16]. In other words, a correspondence analysis would "ideally be seen as an extremely useful *supplement* to, rather than a replacement for, the more formal inferential analysis such as *log-linear* or *logistic* models" [10]. In the context of decision support for real-time trolling narrative identification, however, we determined that a primarily exploratory technique is most appropriate. In the context of our mechanism *TrollHunter2020*, we are trying to quickly get a high-level overview of the trolling patterns as they are emerging in the data corresponding to an ongoing event in the 2020 U.S. election cycle. As such, the high-level overview provided in a correspondence analysis is ideal for *TrollHunter2020* to act fast and capture the trolling narratives before they vanish or morph on Twitter.

A correspondence analysis is a method for comparing two categorical variables. It works by first putting the two variables in a contingency table, then taking the χ^2 statistic, which can be put into a matrix C of elements c_{ij} which can be calculated like this:

$$c_{ij} = \frac{n_{ij} - E_{ij}}{\sqrt{E_{ij}}}$$

Where E_{ij} represents the expected value if all f the values in the contingency table are the same.[10]. To get the first two dimensions from this matrix, we use a matrix decomposition.

The matrix decomposition solves:

$$C = U \Delta V'$$

Where U represents the eigenvectors of CC', V contains the eigenvectors of $C'C$, and Δ represents the diagonals [10]. The values that we are primarily interested in are the first two columns of V, which, for our purposes, we plotted to give us an overview of the data.

3.4 Data Collection

We decided not to work on a Twitter dataset with already identified trolling users and trolling tweets (e.g. the IRA troll dataset in [14] or the Kaggle dataset [7]) nor to rely on user's reports of twitter trolls/post like in [12]. Instead, we utilized the Twitter API to collect tweets related to key events during the 2020 U.S. election cycle that were sure to spark divisive content on Twitter. *TrollHunter2020* allows edits to our constructed set of target words based on the trending keywords/hashtags on Twitter as well as tweets that mentioned either of the two presidential candidates: "@realDonaldTrump" or "@JoeBiden". The creation of the dataset for any given event is contextually dependent, therefore, *TrollHunter2020* develops it in real-time so the dataset can be as holistic and representative as possible. For example, during the vice presidential debate on October 7, 2020, when a fly landed on Vice President Pence's head, "fly" was added to the dataset (manually, by our subjective judgment), since it became quickly clear that this keyword would be used to inject trolled content into the discourse surrounding the debate, regardless of the irrelevance to the political event itself. We manually selected approximately twenty different terms/hashtags for each political event in order to limit our analysis to the *most* topical issues on Twitter during and after significant events took during the 2020 U.S. elections cycle.

4 TROLLHUNTER2020: REAL-TIME DETECTION OF TROLLING NARRATIVES

We collected and analyzed a series of tweets from a few different events that offered a fertile ground for trolling during the 2020 U.S. election cycle. We selected the debates as well as entire period from the election to the confirmation of the results, because these events generated the most discourse on Twitter, and by extension, the most alternative narratives around the candidates.

4.1 Vice Presidential Debate

For the initial iteration of data collection, *TrollHunter2020* required that search terms be only one word, so we had to shorten phrases down to single words to use them to collect tweets. We collected tweets using the trending hashtags, themes, and the official and personal Twitter handles of Senator Harris and Vice President Pence:

- **#BidenHarris2020** a Biden/Harris campaign association.
- **#DebateNight, #VPdebate** a hashtag used to associate a tweet with the vice presidential debate.
- **#PenceKnew** refers to Pence knowing about the seriousness of COVID-19 despite Trump's public downplaying of the events related to the pandemic.
- **abortion** was discussed widely on Twitter as a topic because of Amy Coney Barrett's recent Supreme Court nomination by President Trump.
- **blm, brutality**, short for "Black Lives Matter," was discussed in the context of police brutality during the debate.
- **catholic** refers to Former Vice President Biden's and Amy Coney Barrett's Catholic faith
- **china** in regards to trade with China, as well as China as the source of COVID-19, and the candidates' relationships with China.
- **deal** for Green New Deal, a proposed legislation by Rep. Alexandria Ocasio-Cortez ("AOC") and Sen Ed Markey
- **roe** refers to Roe vs. Wade, feared would come under attack with a conservative Supreme Court
- **speaking** refers to the instances where Vice President Pence interrupted Senator Harris, and Senator Harris said the words "I'm speaking."
- **swine** refers to the "swine flu" which Vice President Pence used to compare to COVID-19.

Figure 1 shows the resulting *TrollHunter2020* correspondence analysis for real-time hunting of trolling narratives on Twitter during the 2020 vice presidential debate. *TrollHunter2020* shows that President Trump (the terms "@RealDonaldTrump" and simply "Trump") is closely associated with the verb "lie." The association is very strong (both nouns and the verb are far from the origin and the angle between the lines connecting them to the origin is very small, resulting in a higher value for the cosine). This suggests that even though he was not present at the debate, a dominant narrative during the debate is that President Trump lies. This might be an obvious notion, but the utility of *TrollHunter2020* in this case is precisely identifying these relationships and emerging direction of trolling. Trolls are aware that Twitter or anyone tracking their activity will be focused on the salient topics identified above (e.g. #PenceKnew or #AOC) and might not look for a hashtag like #TrumpLiedPeopleDied used towards right-leaning users.

When we look at the overall dimensions, Dimension 1 accounts for 33.82% of the inertia of the data and shows how political or procedural a noun/verb pair is. Mentions of "@VP" and "Pence" are both closer to $X = 0$, while "Harris," "Kamala" and "@KamalaHarris," are slightly more to the right, and "Trump" and "@realDonaldTrump" are both even further to the right. Dimension 2, which accounts for 28.08% of the inertia in the data, shows Republicans versus Democrats. "Trump" is all the way on the bottom, and "Harris" and "Kamala" all the way on the top. This indicates that Harris seemed to be more against Trump than against Pence. The strong association between "look" and "Kamala" or "Haris" gives an early warning of the trolling narrative about that she looks upset and nervous leading to a resurgence of hashtags like #KamalaHarrisDestroyed (initially used during the Democratic candidates' debates in conjunction with the hashtag #DemDebateSoWhite [23]). Another alternative narrative relates to the "condescending look" on Senator Harris's face when protesting to Vice President Pence "I am speaking," presenting her as "arrogant, rude, smug" and ultimately a

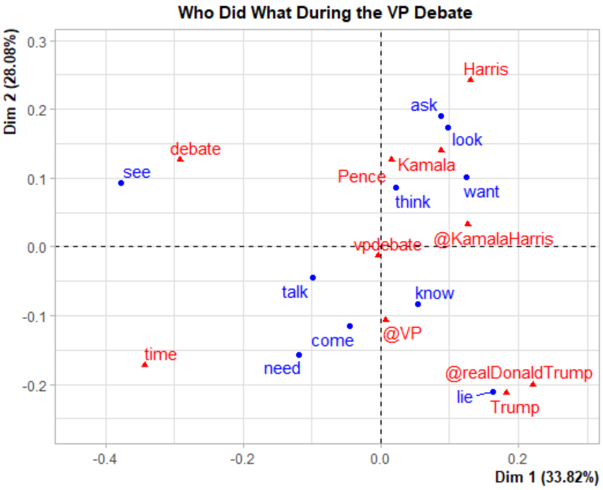


Figure 1: Correspondence Analysis of Top Verbs and Nouns used on Twitter to Describe the Vice Presidential Debate

seed for a sexist and misogynistic attacks with hashtags like #HeelsUpHarris [37]). Finally, it is interesting seeing who is and isn't present. While President Trump was very present in the data, and accounted for two of the identified nouns, Former Vice President Biden did not appear at all in the top nouns, even though one of the terms we used to identify debate-related tweets is the hashtag "#BidenHarris2020." This means that Pence may have needed to answer for things that Trump did, but Harris may not have needed to do the same for Biden. This distinction could be vital for early discrimination of bot accounts that act on behalf of a nation-state actor, placing Senator Harris as the main target of the emerging trolling narratives after the vice presidential debate.

4.2 Final Presidential Debate

With improvements to *TrollHunter2020*, we were able to search phrases as well as single words/hashtags during the final presidential debate which took place on October 22. Juxtaposing the dataset from the vice presidential debate, the final presidential debate featured more trending phrases/keywords than hashtags which is reflected in the second *TrollHunter2020* dataset. In addition to tracking the Twitter handles of each candidate, we collected tweets involving the following trending hashtags and themes:

- **#Debates2020** for collecting tweets referencing the 2020 debates
- **#auditTrump** referencing the President's alleged tax returns which leaked the weekend previously
- **@kwelkernbc** referencing the debate moderator, Kristen Welker's, Twitter handle
- **prepaid tax** a phrase President Trump used during the debate to attest to how much tax he truly paid to the IRS
- **Operation Warp Speed** referencing the Presidential endeavor to quickly develop a vaccine for COVID-19
- **H1N1** criticizing the Obama-Biden administration's handling of the H1N1 virus in comparison to the Trump administration's handling of the COVID-19 pandemic
- **bidencare** a term Biden coined during the debate to refer to his healthcare plan building off of "Obamacare"
- **beautiful healthcare** was the phrase used by President Trump to describe his healthcare proposal
- **lowest IQ** referencing a phrase President Trump used to describe illigal immigrants who returned for U.S. court proceedings
- **who built the cages** a poignant question asked by President Trump to former Vice President Biden who had criticized the Trump administration immigration detention policy
- **least racist** "person in this room" was President Trump's phrase used to defend himself against accusations that his response to the Black Lives Matter movement inflamed racial tensions and emboldened racism
- **poor boys** mistakenly referenced by Biden when discussing the white nationalist group "Proud Boys"
- **fracking** to include tweets discussing one of the most controversial discussions of the night, over the position of both candidates on fracking and general environmental policy
- **I know more about wind than you do** a phrase flaunted by President Trump to critique Biden's knowledge of renewable resources like wind

Regarding the large amount of trending quotes from the presidential debate, it was interesting to see that on Twitter, users seemed to be more focused on the precise wording of the presidential candidates versus during the vice presidential debate when users tended to focus more on the topics and issues discussed. In our initial correspondence analysis, shown in Figure 2, in the first dimension (the x-axis), which accounts for 86.09% of the inertia in the data, the primary trend was a verb/noun pair which was "build," and "cage." Following this early warning from *TrollHunter2020*, "build," and "cage" were contextualized in an emerging and prevalent trolling narrative after the presidential debate about "who build the cages" targeting President Obama about his decision of building immigration cages, additionally amplified with hashtags like #ObamaCages and #DeporterInChief.

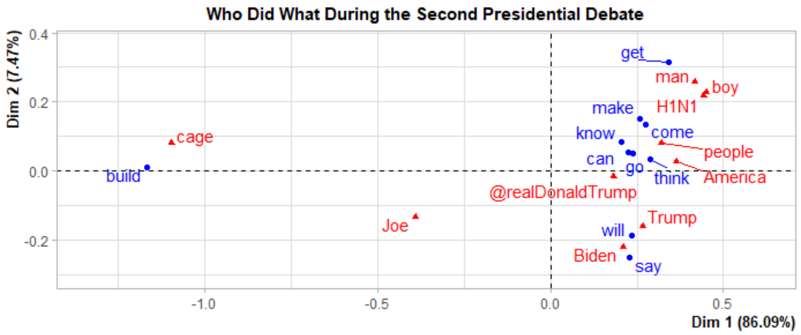


Figure 2: Correspondence Analysis of Top Verbs and Nouns used on Twitter to Describe the Final Presidential Debate

Using *TrollHunter2020* is an iterative manner, the strong patterns initially identified could be removed to allow for the analysis of additional patterns in the data. To continue analyzing the tweets, We re-ran *TrollHunter2020* without the words "cage" and "build" with the results shown in Figure 3(a). The first dimension in the correspondence analysis, which accounts for 43.71% of the inertia in the data, shows for small values the names of the candidates. "H1N1" is approximately at Dimension 1 = 0. The term "come" and "man" could be a reference to Biden's quote "come on man" in reference to an interruption by Trump in the first presidential debate.

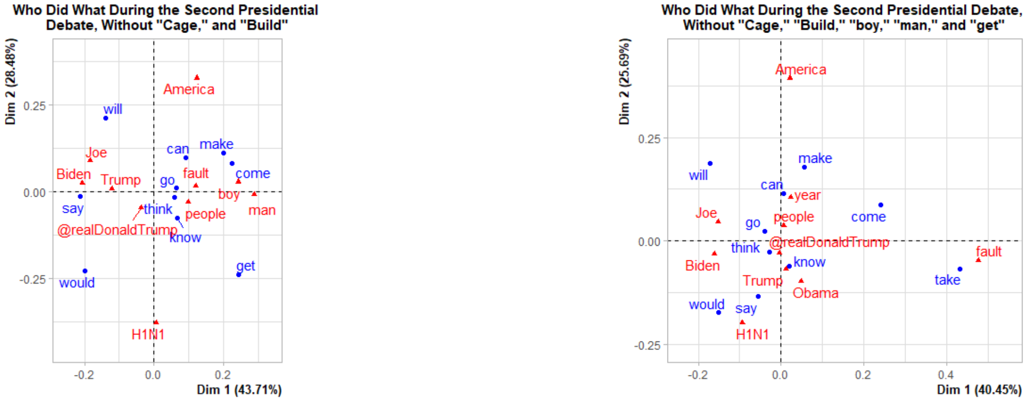
The pattern that correlates with higher valued for the first dimension is a little ambiguous, so we elected to remove a few more terms, "boy," "man" and "get" for a second iteration. As shown in Figure 3(b), the reduced dataset clarified the results further. The words "Joe" and "Biden" are all the way to the left, "H1N1" moves slightly to the left. The removed words make space for "Obama" whose name is a lot closer to "Trump" and "@realDonaldTrump" than it is to "Joe" or "Biden." The verb "take" and the noun "fault" significantly further to the right, indicating that a lot of people were likely trying to attribute blame to Joe Biden about the mass deportations under the Obama administration and his involvement as a Vice President [8].

This *TrollHunter2020* early warning sheds additional light on the emerging trolling narrative targeting Joe Biden's previous White House legacy of more than 3 million immigrants removed from the US [25] with hashtags like #ObamaCabal and #BidenCrimeFamily. Alternative narratives surrounding the Biden family gained significant traction after President Trump himself tweeted out the hashtag #BidenCrimeFamiily in a seemingly innocuous attempt to evade Twitter's censorship of the correct hashtag. This points to the usefulness of *TrollHunter2020* to provide concrete warnings of the danger of propagating trolling narratives from positions of power, since it adds additional legitimacy and encourages Twitter followers and supporters to continue to disseminate those views, regardless of the accuracy of the narrative itself.

4.3 Election Week

Due to the influx of mail-in ballots, unlike previous election cycles, it was expected that a winner of the 2020 U.S. Presidential Election would not be called on Election Night, scheduled for Tuesday, November 3, 2020 [15]. Therefore, for the purposes of this event, we collected data about "Election Night" throughout the week, leading up to the call by Associated Press (and others) on Saturday, November 7, 2020 that declared former Vice President Biden the projected winner of the 2020 Presidential Election and therefore the President-Elect [22]. Although the election does not officially end when media outlets project the winner of the race, for the purposes of analyzing public discourse, the projection by news outlets serves as a critical shifting point in conversation regarding the outcome, and therefore, allowed us a break point on when to stop data mining. For "Election Week," we collected data using the following hashtags and search terms:

- **#AmericaDecides2020** used to refer to everything election-related on November 3, 2020
- **#ElectionNight** used to reference election-related content on the day of the elections
- **#TrumpMeltdown** originated in reference to a speech by President Trump on the night of Thursday, November 5 that some perceived as a disheveled, tired, disjointed appearance and message from the president
- **#StopTheSteal** coined by Trump supporters fearful that the election was stolen by Joe Biden and the Democratic party
- **#StopTheCount** went viral in the days after Election Day, when mail ballots were still being counted throughout the country, urging election workers to stop counting ballots
- **#CountEveryVote** also went viral in the days after Election Day, directly contradicting the previous hashtag, urging election workers to continue to count ballots
- **#IHaveWonPennsylvania** was a claim falsely made by President Trump before the state of Pennsylvania had concluded its vote-tallying



(a) "cage" and "build" removed

(b) "boy," "man' and "get" removed

Figure 3: Iterative Correspondence Analysis of Top Verbs and Nouns used on Twitter to Describe the Final Presidential Debate

- **vote by mail** became a main subject of debate during the election by President Trump who made allegations that it was insecure and contributed to election fraud
- **mail ballot fraud** to track tweets that pushed unsubstantiated narratives of vote-by-mail fraud
- **poll watch**-ing was something that was pushed by President Trump in order to provide additional scrutiny in polling places but it became controversial for fear of poll watchers intimidating voters
- **victory** was included in our dataset to track any premature declarations of victory by either candidate or either candidate's supporters before the race could be accurately projected
- **count** was debated on Twitter throughout Election Week, so by using the word "count" in our dataset, *TrollHunter2020* could identify content related to "stop the count" as well as "count all votes"
- **rigged election** to track tweets spreading divisive information that the election was "rigged" or illegitimate
- **steal election** was utilized to identify tweets accusing any party of perpetrating the "rigging" of the election
- **magically** was the term used by President Trump to describe the vote tallying on Election Night that ultimately swung the vote in key swing states toward Joe Biden

In the Election Night correspondence analysis, the first dimension, which accounts for 64.87% of the inertia in the data, shows slightly different trends for nouns and verbs. Lower values for nouns in the first dimension show specific politicians, like "Trump," or "Biden," with specific election procedural nouns appearing as higher values, like "voter," and "ballot." Interestingly, "Biden" is more likely to be associated with "ballot" which feeds to the distorted narrative of ballot harvesting in Texas orchestrated by a high-level staff member on behalf of Joe Biden [30]. For verbs, the lower values show terms like "concede," and "win" that indicate election results, while more active verbs like "can," "make," and "vote" appearing as higher values in the first dimension. In addition, the more obvious the election outcome became, the verb "concede" was getting closer to the nouns "Trump," "President," and "@realDonalTrump." This is a confirmation of the culminating tension on the election results with Trump's refusal to concede (e.g. #NeverConcede, #StopTheSteal, #BidenCheated2020 #TrumpWon, etc) and the apparent win of Joe Biden (e.g. #BidenWon, #TrumpConcede, etc).

The second dimension which accounts for 17.94% of the inertia in the data shows words more heavily associated with Republicans as higher values. These include nouns like "@RealDonaldTrump," as well as verbs like "censor." Lower values in the second dimension indicate terms more heavily associated with Democrats, such as "Biden," "vote," and "TrumpMeltdown," which refers to the hashtag "#TrumpMeltdown." The verb "censor" is most closely related to the noun "@realDonaldTrump," which could be a reference to Twitter's decision to label many of Trump's tweets as being potentially misleading [38]. The verb "will" is about equally close to both of the candidates, suggesting that in the dataset people were speculating about what both candidates were doing.

4.4 Election Aftermath

As previously noted, in order to differentiate between election week and beyond, we divided our data based on the call by Associated Press (and others) on November 7 that declared the projected winner to be Joe Biden [22]. But it was quite clear, that unlike other elections, the 2020 U.S. Elections would be host to a multitude of legal challenges, conspiracies, and controversies regarding the legitimacy of the results. Therefore, we collected additional data associated with some of the most-debated issues on Twitter in the aftermath of election week. As in previous *TrollHunter2020* instances, we tracked the Twitter handles of the two presidential candidates (@realDonaldTrump and @JoeBiden) as well as a series of terms and hashtags that were trending throughout the next two weeks after November 7, 2020.

- **#25thAmendment** circulated on Twitter as a response to President Trump's lack of public appearances after the election leading some users to believe he should turn over presidential responsibilities to Vice President Mike Pence

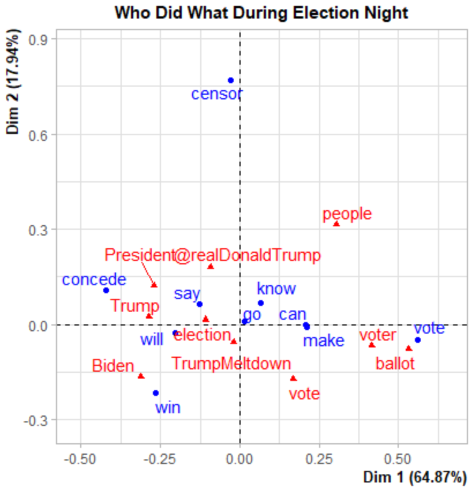


Figure 4: Correspondence Analysis of Top Verbs and Nouns used on Twitter to Describe Election Night

- **#ByeByeTrump** referenced a taunt to Donald Trump after it appeared he had lost the election and would be leaving the White House
- **#fireTrump** was used to reference President Trump's appearances on the television show, *The Apprentice*, which he often said "you're fired!" to contestants
- **#BYEDON** is a play on words combining "Biden" + "Bye Donald" and was used as a Biden campaign slogan that gained significant traction on Twitter
- **#TimeToTransition** urged the President to stop holding up formal transition processes and funds through the General Services Administration (GSA)
- **#PresidentElectJoe** referenced the projected winner and thus, President-Elect, Joe Biden
- **#ItsTimeToConcede** was a plead with President Trump to concede the race instead of continuing to pursue post-election litigation
- **concede** was included to gather tweets referencing the debate over whether or not the incumbent President should concede the race in a speech to Joe Biden
- **BIDEN WON** was a viral phrase in the initial aftermath of news networks calling the race on November 7
- **by a lot** in reference to President Trump falsely claiming that he had won the popular vote by large margins
- **just for men** went viral on Twitter after Giuliani gave a press conference in which it appeared his hair dye had started to bleed down the sides of his cheeks

The correspondence analysis of the election aftermath suggests that the words "Biden" and "win" are close to one another in the lower left hand corner and their separation from the rest of the terms indicates the strength of that pattern. This shows that one of the of the dominant narratives is increased probability that Biden is winning the election. This is corroborated by the relative closeness of the "@realDonaldTrump and "#ByeDon" indicating emergence of trolling against the then-sitting president. The splitting in polar-opposite narratives comes apparent with the closeness of "#PresidentBiden" to the word "people" indicating a relationship between Biden and people. This could be a reference to his efforts to emphasize that he's a President for all people; yet it could also reference the people voting and choosing Biden, or Biden subverting the will of the people by trying to steal the election.

Interesting, the correspondence analysis shows that all of the verbs except for "win" and "lose" are concentrated in the bottom right quadrant of the graph. This means that those two words are treated differently from the other verbs. The other verbs: "concede," "would," "can," "know," "make," "think," "say" and "will" are all mostly speculative terms suggesting some degree of speculation or confusion. Note that this collection of confused terms focuses around the word "Trump," "#PresidentBiden," "people" and "President." These indicate a high degree of speculation and hence potential for trolling. The first dimension, which accounts for 64.91% of the inertia in the data shows a gradient from the low values which indicate victory, with verbs like "win" and "Biden" (ostensibly the winner of the election), to more procedural nouns like "vote" and "election," shifts to terms involving losing like "concede" and "lose" before shifting to taunting terms like "#ByeDon." The second dimension, which accounts for 19.69% of the inertia in the data seems to indicate a spectrum from winning to losing, with lower values in the second dimension indicating winning, with "President," "will" and "people," which all indicate a victory regarding who will be president after a democratic election (the word "win" being an outlier, possibly because the word "win" was used in a large number of tweets referring to Trump, as in misinformation that Trump "won" the election), before transitioning to speculation about concessions with words like "concede," and finally ending with taunts like "#ByeDon" and the word "lose."

4.5 Capitol Insurrection

January 6th, 2021 was bound to be a tense day, as it was the day in which the new 117th Congress would certify each state's electoral votes for the presidency of Joseph R. Biden. It would also mark the symbolic end of Trump campaign lawsuits challenging the outcome

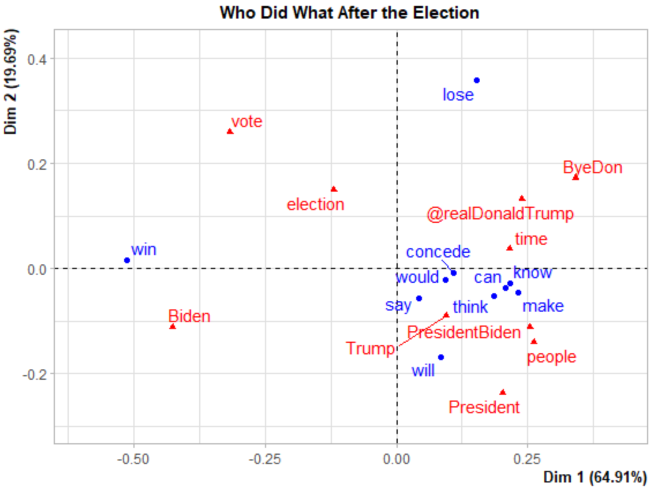


Figure 5: Correspondence Analysis of Top Verbs and Nouns used on Twitter to Describe Events after the Presidential Election

of the election. But tension turned to violent insurrection during the formal certification process, when an armed pro-Trump mob infiltrated the Capitol Building, causing the certification process to temporarily stop as staff and legislators took shelter. We immediately set out to capture tweets about the event as it unfolded, and continued tracking discourse surrounding the insurrection and into the following week, after Twitter permanently banned “@realDonaldTrump” from its platform, and as calls for impeaching President Trump grew, alleging that he was responsible for inciting the violence on the Capitol [39].

- **Trump** was the main target of criticism and the main inspiration to the mob which overtook the Capitol in a last-ditch effort to preserve his presidency
- **Capitol, insurrection** in reference to the riots and the insurrection at the federal building
- **sedition** caught attention after President-Elect Biden addressed the nation, calling the actions of those at the Capitol “borders on sedition" [33]
- **coup** or “attempted-coup" was another term used frequently on Twitter grappling with the armed mob’s actions and inferring their objectives
- **impeachment** as a consequence for President Trump for his role in enraging his base and inciting the mob
- **MAGA** in reference to the accusations toward President Trump’s base
- **ANTIFA**, standing for anti-facist, in reference to early misinformation that claimed the mob was undercover-antifa organizers, not Trump loyalists

In the correspondence analysis, “Trump” is closest to the verbs “impeach” and “resign.” This makes sense given that Trump was impeached during the time for which we collected tweets. Additionally, there were many people did “call” for “Trump” to “resign”, and all three of those words are relatively close to one another, indicating that a lot of people expressed interest in Trump’s departure from power. An interesting pattern is how far “Trump” is from “sedition” and “treason,” indicating that Trump was not associated with those terms very much in this dataset. This means that people may not have been particularly interested in accusing Trump of sedition or treason, but the prevalence of those terms might indicate that a lot of people felt that someone “should” be charged. The noun that “Antifa” is closest to is the noun “Capitol.” This proximity could be due to theories that Antifa was behind the storming of the US Capitol, or people comparing protests in which Antifa participated relative to the storming of the Capitol.

In total, the first two dimensions account for 82.8% of the total inertia in the data. The first dimension, which accounts for 50.23% of the inertia in the data shows a gradual shift from individual to groups, with some of the smallest values representing verbs like “impeach” and “resign,” both of which are individual acts, and nouns including “Trump” and “impeachment” both of which refer to individuals. As the first dimension grows greater, the terms refer to groups and larger numbers of people, like “Antifa” and “MAGA” (which has had a shift in meaning from Trump’s 2016 campaign slogan to a derisive term to describe his supporters, as well as a way for his supporters to continue to show their support). The second dimension, which accounts for 32.57% of the inertia in the data, shows a shift from higher values, that include procedural/results-oriented nouns, like “sedition,” and “treason” as well as speculative verbs, like “should” and “can,” while lower values refer to more specific entities, like “Trump,” “Antifa,” and “Capitol,” as well as verbs that describe those entities, “say,” “call.”

5 DISCUSSION AND CONCLUSION

5.1 Ethical Considerations of Using TrollHunter2020

We openly acknowledge that *TrollHunter2020* could be abused in multiple contexts for nefarious purposes with malicious repurposing of what constitutes an “alternative, distorted, or trolling narrative.” For example, in response to popular uprisings in late May 2020, sparked by the killing of George Floyd by police officers

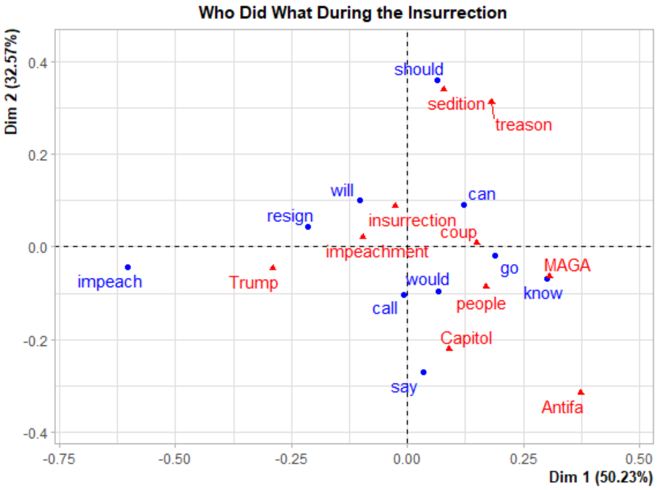


Figure 6: Correspondence Analysis of Top Verbs and Nouns used the week following the attempted insurrection

in Minneapolis, many activists took to Twitter to spread awareness of the incident, using hashtags like #BlackLivesMatter, #BLM, #iCantBreathe, and #RestInPower [27]. If *TrollHunter2020* is used without considering the developing socio-political landscape in this case, it might capture combinations of words and nouns that look like an alternative narrative, but might very well be a legitimate trending Twitter topic about a mass struggle. The narrow use of *TrollHunter2020* might prevent this narrative to achieve the intended goal of social action -in fact, the Black Lives Matter (BLM) movement was born out of the the hashtag on Twitter- which is the opposite effect of what *TrollHunter2020* was developed to achieve. Additionally, with a tool like *TrollHunter2020*, opposition groups have an opportunity to craft emerging counter-narratives in real-time in public spaces like Twitter by combining opposite matching of the noun-verb or noun-noun relationships. For example, #AllLivesMatter emerged out of retaliation to the initial BLM narrative and movement. Therefore we strongly recommend the use of *TrollHunter2020* only as an early, high overview troll hunter with special consideration of the contemporary socio-political landscape in deciding narratives that actually pollute public discourse.

5.2 Implications of Using TrollHunter2020

Zannettou et al. showed that alternative narratives flow through multiple social media platforms like Reddit, Twitter, and 4chan [42]. The introduction of real-time trolling narrative detection, in general, could potentially have an effect of moving trolls to less regulated platforms. A recent example of such a migration from Twitter to Parler, Rumble and Newsmax was witnessed after Twitter actively labeled and removed false information on the platform during the 2020 U.S. elections [19]. An opposite effect is also possible, where trolls or fringe Web communities disseminating trolling narratives could be attracted on Twitter by exploiting the limitations of *TrollHunter2020* to precisely identify the potent or most probable candidate of words for the emerging trolling narratives. Aware of the limitation of the correspondence analysis of *TrollHunter2020*, trolls could come up with alternative tropes or words, or hashtags, for example, “demagogue” instead of “lie” (for the vice presidential debate), “blue-pencil” instead of “censor” (for the final presidential debate), and “triumph” instead of “win” (for the Election Night), that could evade detection.

5.3 Inherent Limitations of TrollHunter2020

TrollHunter2020, like every automated detection mechanism, comes with a set of limitations. Even though we attempted to capture the most popular topics on Twitter surrounding the events we reviewed, they might not represent the complete picture around a given election event, especially around events that are discussed not just on Twitter but on many other social media platforms and forums. The selection of hashtags, themes, and accounts to be involved was rather limited, and one might yield different lists of top words to be used for the correspondence analysis if they employ a different selection criteria. *TrollHunter2020* provides a real-time high overview of emerging trolling narratives and is possible to misidentify or completely miss a combination of words, hashtags, and accounts. *TrollHunter2020* does not identify trolling accounts nor discriminates between types of accounts, e.g. bot versus human accounts. *TrollHunter2020* highlights some of the biggest narratives circulating, and should only be used to help clarify and analyze the overall trends in the data in a decision support role rather than an automated hunter of trolls on Twitter.

5.4 Future Adaptations of TrollHunter2020

Juanals and Minel used a very structured analysis to analyze the dissemination of information across different Twitter accounts as an approach to supplement their correspondence analysis [20]. It might be interesting to incorporate some of these more structured elements into a future iteration of *TrollHunter2020* to see the outcome of the overall troll hunt. For example, *TrollHunter2020* could be enhanced to to track the spread of messaging from the President to their cabinet members' Twitter accounts or from any person of interest during and election and their closes collaborators on Twitter. Another possible adaptation of *TrollHunter2020* is to yield a comparative correspondence analysis between two datasets created around a highly publicized event, for example, a Twitter dataset and a Parler dataset. The benefit of such a comparison is to broaden the scope of detection of early trolling narratives, given that social media becomes more segmented based on ideological preferences and as a result of active trolling detection from Twitter. In this context, it would be interesting to employ *TrollHunter2020* to monitor the ongoing Twitter discourse around the approval, distribution, administration, and monitoring of COVID-19 vaccines.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives* 31, 2 (2017), 211–36.

[2] Yochai Benkler, Robert Faris, and Hal Roberts. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics.* Oxford University Press, Oxford, UK.

[3] Kenneth Benoit and Akitaka Matsuo. 2020. *spacyr: Wrapper to the 'spaCy' 'NLP' Library.* https://CRAN.R-project.org/package=spacyr R package version 1.2.1.

[4] David A. Broniatowski, Amelia M. Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C. Quinn, and Mark Dredze. 2018. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. *American Journal of Public Health* 108, 10 (2018), 1378–1384. https://doi.org/10.2105/AJPH.2018.304567

[5] Emily Chen, Ashok Deb, and Emilio Ferrara. 2020. #Election2020: The First Public Twitter Dataset on the 2020 US Presidential Election. arXiv:cs.SI/2010.00600

[6] Bryn Alexander Coles and Melanie West. 2016. Trolling the trolls: Online forum users constructions of the nature and properties of trolling. *Computers in Human Behavior* 60 (2016), 233 – 244. https://doi.org/10.1016/j.chb.2016.02.070

[7] DataTurks. 2020. Tweets Dataset for Detection of Cyber-Trolls. https://www.kaggle.com/dataturks/dataset-for-detection-of-cybertrolls

[8] Ryan Devereaux. 2020. Will Biden Dismantle Trump's Immigration Police State? https://theintercept.com/2020/11/17/biden-immigration-obama-trump/

[9] Michael Dimock and Richard Wike. 2020. America is exceptional in the nature of its political divide. https://www.pewresearch.org/fact-tank/2020/11/13/america-is-exceptional-in-the-nature-of-its-political-divide/

[10] Brian S. Everitt and Graham Dunn. 2010. 4 Correspondence Analysis. In *Applied Multivariate Data Analysis* (2nd edition ed.). Wiley, Chichester, 74 – 92.

[11] Emilio Ferrara, Herbert Chang, Emily Chen, Goran Muric, and Jaimin Patel. 2020. Characterizing social media manipulation in the 2020 U.S. presidential election. *First Monday* 25, 11 (2020/11/30 2020). https://doi.org/10.5210/fm.v25i11.11431

[12] Paolo Fornacciari, Monica Mordonini, Agostino Poggi, Laura Sani, and Michele Tomaiuolo. 2018. A holistic system for troll detection on Twitter. *Computers in Human Behavior* 89 (2018), 258 – 268. https://doi.org/10.1016/j.chb.2018.08.008

[13] Christian Fuchs. 2018. *Digital demagogue: Authoritarian capitalism in the age of Trump and Twitter.* Pluto Press.

[14] Bilal Ghanem, Davide Buscaldi, and Paolo Rosso. 2019. TexTrolls: Identifying Russian Trolls on Twitter from a Textual Perspective. arXiv:cs.CL/1910.01340

[15] Shane Goldmacher. 2020. A Winner on Election Day in November? Don't Count on It. https://www.nytimes.com/2020/06/24/us/politics/november-2020-election-day-results.html

[16] Michael Greenacre. 2016. Correspondence analysis in medical research. *Statistical methods in medical research* 1, 1 (2016), 97–117.

[17] Jane Im, Eshwar Chandrasekharan, Jackson Sargent, Paige Lighthammer, Taylor Denby, Ankit Bhargava, Libby Hemphill, David Jurgens, and Eric Gilbert. 2020. Still out There: Modeling and Identifying Russian Troll Accounts on Twitter. In *12th ACM Conference on Web Science (WebSci '20).* Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3394231.3397889

[18] Jane Im, Eshwar Chandrasekharan, Jackson Sargent, Paige Lighthammer, Taylor Denby, Ankit Bhargava, Libby Hemphill, David Jurgens, and Eric Gilbert. 2020. Still out There: Modeling and Identifying Russian Troll Accounts on Twitter. In *12th ACM Conference on Web Science (WebSci '20).* Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3394231.3397889

[19] Mike Isaac and Kellen Browning. 2020. Fact-Checked on Facebook and Twitter, Conservatives Switch Their Apps. https://www.nytimes.com/2020/11/11/technology/parler-rumble-newsmax.html

[20] Brigitte Juanals and Jean-Luc Minel. 2017. Information Flow on Digital Social Networks during a Cultural Event: Methodology and Analysis of the "European Night of Museums 2016" on Twitter. In *Proceedings of the 8th International Conference on Social Media & Society (SMSociety17).* Association for Computing Machinery, New York, NY, USA, Article 13, 10 pages. https://doi.org/10.1145/3097286.3097299

[21] Sébastien Lê, Julie Josse, and François Husson. 2008. FactoMineR: A Package for Multivariate Analysis. *Journal of Statistical Software* 25, 1 (2008), 1–18. https://doi.org/10.18637/jss.v025.i01

[22] Jonathan LeMire, Zeke Miller, and Will Weissert. 2020. Biden defeats Trump for White House, says 'time to heal'. https://apnews.com/article/joe-biden-wins-white-house-ap-fd58df73aa677acb74fce2a69adb71f9

[23] Maureen Linke and Elisa Collins. 2019. Bot-Like Activity Pushed Divisive Content on Race During Debates. https://www.wsj.com/articles/bots-pushed-divisive-content-misinformation-on-race-during-debates-11564678048?mod=rsswn

[24] Clare Llewellyn, Laura Cram, Adrian Favero, and Robin L. Hill. 2018. Russian Troll Hunting in a Brexit Twitter Archive. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries (JCDL '18).* Association for Computing Machinery, New York, NY, USA, 361–362. https://doi.org/10.1145/3197026.3203876

[25] Department of Homeland Security. 2017. Aliens Removed or Returned: Fiscal Years 1892 to 2017. https://www.dhs.gov/immigration-statistics/yearbook/2017/table39

[26] F. Javier Ortega, Jose A. Troyano, Fermin L. Cruz, Carlos G. Vallejo, and Fernando Enriquez. 2012. Propagation of trust and distrust for the detection of trolls in a social network. *Computer Networks* 56, 12 (2012), 2884 – 2895. https://doi.org/10.1016/j.comnet.2012.05.002

[27] Russell Rickford. 2016. Black Lives Matter: Toward a Modern Practice of Mass Struggle. *New Labor Forum* 25, 1 (2016), 34–42. https://doi.org/10.1177/1095796015620171

[28] Yoel Roth and Nick Pickles. 2020. Updating our approach to misleading information. https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html

[29] David Sanger. 2020. Pompeo Ties Coronavirus to China Lab, Despite Spy Agencies' Uncertainty. https://www.nytimes.com/2020/05/03/us/politics/coronavirus-pompeo-wuhan-china-lab.html

[30] Kirk Semple. 2020. No, a high-level member of the Biden campaign was not arrested in Texas. https://www.nytimes.com/2020/11/17/technology/no-a-high-level-member-of-the-biden-campaign-was-not-arrested-in-texas.html

[31] Filipo Sharevski, Peter Jachim, and Paige Treebridge. [n. d.]. TrollHunter [Evader]: Automated Detection [Evasion] of Twitter Trolls During the Coronavirus Pandemic. In *Proceedings of the New Security Paradigms Workshop (NSPW '20).* Association for Computing Machinery, New York, NY, USA, 1–12.

[32] Julia Silge and David Robinson. 2016. tidytext: Text Mining and Analysis Using Tidy Data Principles in R. *JOSS* 1, 3 (2016). https://doi.org/10.21105/joss.00037

[33] Andrew Solender. 2021. Biden Says Capitol Breach is Insurrection That Borders on Sedition. https://www.forbes.com/sites/andrewsolender/2021/01/06/biden-says-capitol-breach-is-insurrection-that-borders-on-sedition/?sh=29c668975dbb

[34] Kate Starbird. 2017. Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter.

[35] Kate Starbird, Ahmer Arif, and Tom Wilson. 2019. Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 127 (Nov. 2019), 26 pages. https://doi.org/10.1145/3359229

[36] Leo G Stewart, Ahmer Arif, and Kate Starbird. 2018. Examining trolls and polarization with a retweet network. In *Proc. ACM WSDM, workshop on misinformation and misbehavior mining on the web. 2018.*

[37] Karen Tumulty, Kate Woodsome, and Sergio Pecanha. 2020. How sexist, racist attacks on Kamala Harris have spread online — a case study. https://www.washingtonpost.com/opinions/2020/10/07/kamala-harris-sexist-racist-attacks-spread-online/?arc404=true

[38] Twitter. 2020. Information Operations. https://transparency.twitter.com/en/reports/information-operations.html

[39] Noah Weiland. 2021. Impeachment Briefing: Trump is Impeached, Again. https://nyti.ms/39KmWNR

[40] Hadley Wickham, Romain François, Lionel Henry, and Kirill Müller. 2020. *dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr R package version 1.0.2.

[41] SC Woolley and D Guilbeault. 2017. *Computational propaganda in the United States of America: Manufacturing consensus online.* Technical Report. 1–29 pages.

[42] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Nicolas Kourtelris, Ilias Leontiadis, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2017. The Web Centipede: Understanding How Web Communities Influence Each Other through the Lens of Mainstream and Alternative News Sources. In *Proceedings of the 2017 Internet Measurement Conference (IMC '17).* Association for Computing Machinery, New York, NY, USA, 405–417. https://doi.org/10.1145/3131365.3131390

[43] Savvas Zannettou, Tristan Caulfield, William Setzer, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2019. Who Let The Trolls Out? Towards Understanding State-Sponsored Trolls. In *Proceedings of the 10th ACM Conference on Web Science (WebSci '19).* Association for Computing Machinery, New York, NY, USA, 353–362. https://doi.org/10.1145/3292522.3326016

[44] Savvas Zannettou, Michael Sirivianos, Jeremy Blackburn, and Nicolas Kourtellis. 2019. The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans. *J. Data and Information Quality* 11, 3, Article 10 (May 2019), 37 pages. https://doi.org/10.1145/3309699