DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

The 2020 Presidential Election on Twitter: An Exploration of Candidates' Social Presence, Campaign Momentum, and the Effect of Misinformation

Citation
Lucioni, Sarah. 2021. The 2020 Presidential Election on Twitter: An Exploration of Candidates' Social Presence, Campaign Momentum, and the Effect of Misinformation. Bachelor's thesis, Harvard College.

Permanent link
https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37368561

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

The 2020 Presidential Election on Twitter: An Exploration of Candidates' Social Presence, Campaign Momentum, and the Effect of Misinformation

Sarah K. Lucioni

Department of Computer Science

Department of Statistics

Harvard College

Supervisors

Professor Jim Waldo

Professor Mark Glickman

In partial fulfillment of the requirements for the degree of

Bachelor of Arts in Computer Science and Statistics

March 26, 2021

Abstract

Twitter has evolved from a site of inconsequential information spread to an instant primary source used as the preferred outlet to discuss and witness any semblance of news the emerges each day. Political outreach thrives in this medium due to ease of communication across a diverse and engaged audience. The rapid nature of Twitter creates an ecosystem suited for misinformation spread. Influential misinformation afflicted the platform during the 2016 U.S. presidential election and during the COVID-19 pandemic, causing Twitter to implement policy changes that, while minor, garnered significant media attention and user backlash.

This thesis aims to understand the characteristics of political Twitter as it relates to prominent campaigns, and how misinformation affects campaign momentum. Specifically, we investigate the 2020 presidential election between Donald Trump and Joseph Biden through tweets authored by the candidates and through tweets mentioning the candidates. We evaluate the candidates' social presence on Twitter, their campaign momentum, and the effect of misinformation via sentiment analysis, network analysis, and momentum analysis. In terms of social presence, we find that Trump's tweeting behavior is more active and that the tweeting behavior of both candidates increases in the days surrounding a notable event. We also observe that Biden more frequently publishes emotionally charged messages of trust, anticipation, and joy. Both Biden and Trump exhibit a negative campaign style when mentioning their opponent, but only Trump's negative style resonates with his base as seen via a significant increase in user engagement following Trump's negative tweets. In discussion on Twitter, we discover that general election information is more commonly tied to Biden while non-election information and niche stories are more often tied to Trump.

Regarding campaign momentum, we find that Biden's campaign witnesses more positive momentum relative to Trump's campaign as measured by a sentiment indicator. We then apply sentiment and curvature indicators to explore associations between targeted misinformation events and changes in campaign momentum. Overall, this thesis demonstrates the use of Twitter as a data source to investigate the social presence and momentum shifts of influential members.

Acknowledgements

First and foremost, I would like to thank my two advisors, Professor Jim Waldo and Professor Mark Glickman, for their support and mentorship.

Professor Waldo has been a source of consistent guidance and learning since my first year. He has taught me to think critically and has encouraged me to pursue new challenges every year. Without his support, this thesis would have never come to fruition.

Professor Glickman has helped me understand the real-world applications of statistics. Additionally, thanks go to Professors Glickman, Rader, Blitzstein, and Kohler for making statistics and computer science accessible and fun.

I would also like to thank my friends who make me laugh and have my back even when we are on opposite coasts. Your encouragement and kindness have kept me going all throughout college.

Lastly, I express endless gratitude to my family whose support, encouragement, and love has shown me that I can do hard things. You all enrich my life more than I can fit into words.

Thank you.

Contents

1 Background **1**

1.1 History . . . 1

1.2 Contemporary Media and Definitions . . . 2

1.3 The Function of False Information . . . 6

1.4 Twitter . . . 7

1.4.1 Political Twitter . . . 8

1.4.2 Russian Disinformation Campaign . . . 11

1.4.3 Twitter's Misinformation Response . . . 16

1.5 Moving Forward . . . 18

2 Literature Review **20**

2.1 Analyzing Politics on Twitter . . . 20

2.2 Tweet Sentiment Analysis . . . 22

2.3 Measuring Political Momentum . . . 24

2.4 Uncovering Misinformation Trends . . . 27

2.5 Takeaways . . . 28

3 Data **30**

3.1 Data Overview . . . 30

3.1.1 Twitter API and Data Structure . . . 30

3.1.2 Representative Sample . . . 34

3.1.3 Data Selection . . . 40

3.2 Exploratory Data Analysis . . . 41

3.2.1 Data at Large . . . 41

3.2.2 Candidate Specific Data . . . 47

4 Methods **50**

4.1 Data Preprocessing 50

4.2 Sentiment Analysis 53

4.2.1 SentiStrength and AFINN 53

4.2.2 BERT 54

4.2.3 Model Comparison 56

4.3 Ego Networks 61

4.4 Momentum Strategies 62

5 Results and Discussion **64**

5.1 Social Presence 65

5.1.1 Temporal Characteristics 65

5.1.2 Engagement Style 70

5.1.3 Campaign Style 82

5.1.4 Candidates in Discussion 86

5.2 Campaign Momentum 91

5.3 Effect of Misinformation 97

5.3.1 Wisconsin Ballots 98

5.3.2 Hunter Biden New York Post Article 101

5.3.3 Stop the Steal 104

5.3.4 Discussion 107

6 Conclusion **109**

6.1 Takeaways 109

6.2 Future Directions 110

6.3 Conclusion 111

References **113**

Appendices **120**

A Data **120**

A.1 Tweet Object: Extended Entities Expanded 120

A.2 Tweet Object: Entities Expanded 121

A.3 User Object 122
A.4 List of Twitter Accounts for Filtered Stream 123

B Methods **129**

B.1 Data Handling 129
B.1.1 gDrive API Script 131
B.2 BERT Training 133
B.3 Ego Network Node Size Distribution 133

C Results **135**

C.1 List of Ego-Network Hashtags 135

Chapter 1

Background

1.1 History

> *Like secondhand smoke, misinformation damages the quality of public life...* [12]

The 2000s saw the launch of social media networks that would eventually connect people all around the world, change the global means of communication, and grow into a perfect host and breeding ground for misinformation. While the means of spreading information has evolved throughout human history, false information has continually plagued our society. Historians note the Behistun Inscription circa 515 BC as an early example of propaganda justifying King Darius' rebellion against the previous rulers and his ensuing right to rule [24]. Since then, the prevalence of propaganda and false information has exponentially increased, closely following the trend of communication advancements. The printing press, radio, and internet have all transformed communication and allowed for an influx of misinformation. We have also seen political changes coupled with these social developments.

For instance, the Byzantine Procopius authored *Anecdota* (also known as *Secret History*) in the sixth century AD. Published after his death, the work slandered Emperor Justinian even though Procopius' other works admired the ruler [9]. Notable scenes of misinformation include those where the Emperor is described as

possessed by evil spirits or even suspected to be a demon as well as scenes defaming the Empress in a vulgar light [33]. It is believed that Procopius wrote this false information in order to avoid association with the Emperor's regime in case Justinian's rule was overthrown, causing the fear of execution to become reality.

With the printing press came *pasquinades* and *canards* in the 16th and 18th centuries, respectively. Pasquinades became a genre to distribute false and harmful information surrounding public figures. Canards emerged from pasquinades two centuries later as a form of false information targeted at the gullible. The popularity of this variety of misinformation contributed to the hatred of Marie-Antoinette, precipitating her execution in 1793. The 18th century also introduced *paragraph men*. These short paragraphs, inspired by coffee house gossip, were authored to manipulate public opinion surrounding a public figure, a play, or even a book. The paragraphs would be published in newspapers. These scraps of gossip-tainted news would then proliferate via word of mouth. [9]

1.2 Contemporary Media and Definitions

Today, any entity with Internet access can become a paragraph being. In 280 characters (previously 140 characters), anyone can document or reshare any and every thought on Twitter. As former United States President Donald Trump put it:

Figure 1.1

(The "losses" here are fact checking, healthy debate between all sides, and working with a team of educated journalists). This sentiment is bolstered by the fact that in 2019, Pew Research found that 55 percent of adults in the U.S. often or sometimes get their news from social media [40]. An estimated 73 percent of U.S. adult Facebook users receive their news from the site. Similarly, approximately 71 percent of U.S. adult Twitter users get their news from Twitter. This is extremely dangerous since social media lacks professional ethics. Sites do not require careful examination or critical thinking before allowing users to share posts. Instead, people are driven to portray their "identity" to feel connected to one another online. This means that users will share posts without a second thought as long as they support the same political party, social activism movement, or identify as the same sexuality, religion, race, or ethnic group. The platforms' recommendation algorithms learn that similar material keeps users engaged. Thus, as time stretches on, sites continuously confirm users' biases. This methodology creates echo chambers where bubbles of agreement dampen dissent and preclude critical reflection.

The increased societal fragmentation online in tandem with the rapid pace at which information now spreads fosters an atmosphere for misinformation that has affected politics globally. From Dr. Joan Donavan's Media Manipulation casebook, *misinformation* is "information whose inaccuracy is unintentional, and spread unknowingly" [10]. The following two posts both serve as examples of misinformation. Figure 1.2 is a Facebook post (falsely) claiming that a Joe Biden impersonator wore a mask of Biden's face to give a CNBC interview. This is unintentionally inaccurate (the line is caused by a shadow) [28] and spread to at least 193 other people.

Figure 1.2 Misinformation Targeted at Biden

While the above clearly attacks the Democratic presidential candidate, the spread of misinformation is not partisan. Figure 1.3 shows a tweet of a fake Trump campaign fundraising email asking for donations to help the (now former) President recover from COVID-19 [15]. Again, this tweet exhibits traits of misinformation because the user unintentionally spreads false information.

Figure 1.3 Misinformation Targeted at Trump

On the other hand, *disinformation* is "information that is deliberately false or misleading, often spread for political gain, profit, or to discredit a target individual, group, movement, or political party" [10]. Most notably in recent history, the Russian government funded the St. Petersburg-based Internet Research Agency (IRA) to run a widespread disinformation campaign to polarize the U.S., disrupt politics and elections worldwide, and cast a powerful light on Russia [35]. The IRA-run Twitter accounts were most active in 2016 during the U.S. presidential election, but the accounts started tweeting en masse around 2014. Some of the accounts were activated and used as early as 2012 [14]. Examples from the IRA influence campaign are presented further on in the section.

We abstain from using the term *fake news* since politicians have manipulated the expression to degrade sources that disagree with their political agenda instead of describing a piece of information's veracity. To objectively discuss a story or assertion's veracity, we use the terms *true* or *false* information [50].

1.3 The Function of False Information

The misinformation examples above and the success of the IRA influence campaign show that false information is compelling and hard to quickly decipher, highlighting society's susceptibility to misinformation. In addition to social networks providing a platform to amplify consequential messages across communities, growing societal tensions offer a spark to manufacture captivating false information. Institutional trust is one such tension corroding in part due to widening income inequality. Climate change is another growing tension as effects become increasingly visible and time left to adequately respond runs low. Rise of automation is a third tension instilling fear for one's privacy and job security. An adversary wishing to manipulate an audience can capitalize on the emotional responses that these issues elicit by fabricating novel information that evokes reactions of anger or excitement. These reactions push users to respond by spreading the information in some form. In so doing, the audience lends their credibility to the story, becoming messengers of false information. [51]

False information aims to overwhelm a user, cause confusion, and undermine institutional trust (in a political context). Adversaries producing *disinformation* seek monetary gain, political influence, or simply a desire to distress. As discussed earlier, socio-psychological factors drive spreaders of *misinformation* because of an urge to feel connected and display one's identity. Consumers can accidentally regurgitate disinformation into misinformation due to challenges in recognizing intentionally misleading information. Disinformation constructed with a grain of truth effectively aids this transformation. Hyperbolic and satirical headlines are successful disinformation formats because the hyperbolic or satirical element evokes a strong emotion, while the sliver of truth makes it hard to debunk. Repetition and familiarity are also compelling mechanisms of disinformation. Repetitive stories become familiar, and our brain assigns credibility to that familiarity. Another disinformation strategy capitalizes the meme format. Memes rely on prior shared knowledge about an event. Recounting that knowledge when reading a meme inspires a feeling of connection with the relevant community. The effort required to digest a meme reinforces the subject matter. Additionally, memes are

easily shareable and comfortably consumed. [51]

Social media networks are prime tools to amplify stories of misleading information. Narrowing our scope to Twitter, we can understand why misinformation thrives. The main features of Twitter are the *tweet*, *retweet*, and *reply*. A tweet is simply a short post that consists of some combination of text, links, and images. A retweet is similar to a vote of agreement or personal identification with another user's tweet. A reply is a way to respond to a tweet. The short nature of tweets allows for easy digestion of loads of information, leaving no desire to verify the veracity of each story. This means that repetition and familiarity thrive on Twitter. Additionally, the reply and retweet features allow for users to react to and amplify a message, creating a sense of community. Hyperbolic headlines, satirical stories, and memes are all easily (and commonly) shared on Twitter. Hashtags are another popular mechanism on the network that allow people to derive a sense of connection with like-minded communities. Overall, these attributes ingrain an impulse to keep scrolling and never put one's Twitter feed down. This is part of what makes Twitter profitable; therefore, the company works to feed each user's addiction which in turn fuels the success of misinformation.

1.4 Twitter

Twitter publicly launched on July 15, 2006 [3]. Co-founder and CEO Jack Dorsey explained the origin of the name Twitter, noting:

> *[the team] came across the word "twitter," and it was just perfect. The definition was "a short burst of inconsequential information," and "chirps from birds." And that's exactly what the product was. [38]*

Unbeknownst to Dorsey at the time, the product would take off in the political and current events spheres, transforming the "inconsequential" spread of information into an instant primary source. Politicians and citizens alike can amplify their opinion without a need for factual basis, molding the online climate into a

thriving environment for misinformation.

As of January 2021, the site hosts 340 million users and traffics 500 million new tweets every day [39]. This activity ranks Twitter as only the 13th most popular social media site, falling strongly behind the 2.449 billion Facebook users and 2 billion YouTube users [39]. However, Twitter's monetizable daily active users (mDAUs) have grown drastically throughout 2020 in response to "global conversation around current events" such as the COVID-19 pandemic and the U.S. presidential election [34]. In 2020, Q1, Q2, and Q3 saw a 24 percent, 34 percent, and 29 percent year over year respective mDAU growth, resulting in 187 million mDAUs by the start of Q4 [34].

1.4.1 Political Twitter

User growth in response to current events demonstrates how Twitter is the preferred outlet to discuss and witness any semblance of news throughout a day. Similarly, Twitter has grown to become the preferred platform for consistent political outreach, driven both by the easy means of communication to a diverse and engaged audience as well as by the perpetual use by the 45th U.S. President: Donald Trump. We can measure political engagement by comparing the followings of politicians across different social media platforms. Notably, as of January 2021, Twitter's most followed user is President Barack Obama with 127 million followers [20]. President Obama's Instagram and Facebook followers pale in comparison at 34 million and 56 million, respectively. Similarly, President Joseph Biden has accrued 20 million Twitter followers, 14 million Instagram followers, and eight million Facebook followers. Many other politicians follow the same trend, including Vice President Kamala Harris, Speaker of the House Nancy Pelosi, and Vice President Mike Pence.

President Trump has 89 million followers on Twitter, 24 million on Instagram, and 35 million on Facebook. He joined Twitter in May 2009, initially to market his brand. In 2011, his feed became more personal and the volume of authored

messages quintupled two years in a row [42]. Once elected, President Trump's tweets entirely blurred the lines between Twitter as a political outreach platform and as a national platform to manifest official action. For example, in the month prior to the November election, we saw the former President break the news of his and the First Lady's COVID-19 diagnosis on October 2nd (EST):

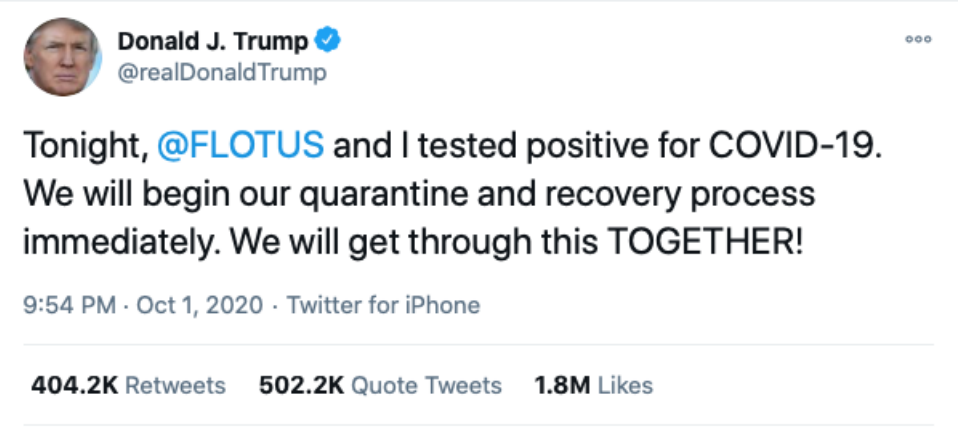


Figure 1.4 Trump Diagnosis (PST)

We also saw the President fire the Director of the Cybersecurity and Infrastructure Security Agency (CISA):

Figure 1.5 Trump Fires Krebs (PST)

Here, we saw the President spread false information to his 89 million followers:

Figure 1.6 Trump Spreads Misinformation (PST)

Figure 1.7 Trump Spreads Misinformation (PST)

The rapid nature of Twitter has given not just President Trump, but any user the ability to snatch a megaphone and amplify serious misinformation. Placing the megaphone in the hands of a world leader attaches credibility (no matter how flimsy) to the false stories they spread, and makes removing the pedestal incredibly difficult in terms of free speech and corporate imposed censorship. We witnessed violence and censorship emerge in physical reality with the January 6th Capitol Riot. Shortly after on January 8th, 2021, President Trump's Twitter account was permanently suspended due to the risk of further incitement of violence [30].

1.4.2 Russian Disinformation Campaign

In the 2016 U.S. Presidential Election, the Russian IRA disinformation operation successfully disturbed the momentum of political campaigns by pitting a variety of Twitter accounts against one another. Clemson researchers investigated nearly 2 million IRA-linked tweets (in English) between June 19, 2015 to December 31, 2017. These tweets came from 1,311 unique accounts that researchers separated into five distinct categories: Right Troll, Left Troll, News Feed, Hashtag Gamer, and Fearmonger. [35]

Right Troll was the most active category with 617 accounts and 663,740 tweets. These accounts rarely focused on conservative policies such as taxes, regulation,

and abortion; instead, they spread "nativist and right-leaning populist messages" including support of Trump, use of the MAGA hashtag, and attacks on Democrats. The accounts that contained identifying information typically portrayed young, attractive women. Overall, Right Troll accounts accrued nearly one million followers. [35]

The Left Troll category contains 230 accounts with 405,549 tweets. The content focused on cultural identity with a strong spotlight on racial identity. However, these accounts targeted and undermined Hillary Clinton's campaign by painting her as untrustworthy and too centrist. There was a clear objective to speed up Trump's political momentum and slow down Clinton's momentum since Right Trolls supported Trump, and Left Trolls attempted to damage Clinton's image while ostensibly promoting Democratic issues. [35]

The News Feed, Hashtag Gamer, and Fearmonger accounts all had explicit purposes. News Feed accounts mimicked local news accounts across the U.S. The 54 accounts were responsible for 567,846 tweets. Over 100 Hashtag Gamer accounts focused on using left and right political hashtags to incite reactions over divisive social issues. As discussed earlier in Section 1.3, these tweets were engineered to capitalize on emotional responses so that users would interact with and spread the false information presented. Finally, 122 Fearmonger accounts specifically promoted hoaxes, accounting for 10,000 tweets. [35]

The influence operation's effectiveness depended on the coordination among these Twitter accounts. The accounts capitalized on sensitive topics by attacking and defending both sides. Left-appearing accounts *undermined* Clinton's campaign while attacking Republicans and supporting social movements. Right-appearing accounts *supported* Trump's campaign while attacking the Left, promoting racism, and undercutting religious freedom and equality. These topics helped build up a network of like-minded Americans, resulting in large followings on many IRA accounts. The multifaceted accounts incited division by triggering emotional responses and false information spread. [35]

We can observe the potency of the IRA's operation by looking at a handful of accounts. For example, the @TEN_GOP account falsely identified as the Tennessee state GOP account, accumulating almost 150,000 followers in its approximately 1.5 year lifespan. At the same time, Tennessee's legitimate GOP account, @TNGOP, had been active for nine years with only 13,000 followers. The state notified Twitter on multiple occasions that @TEN_GOP was fake. Twitter did not react until August 2017 when they permanently suspended the account 11 months after the initial alert [7]. The @TEN_GOP's 11,341 tweets clearly exhibit far-right sentiments and highlight the IRA's tactic to provoke division. For instance, the following tweets attack Islam:

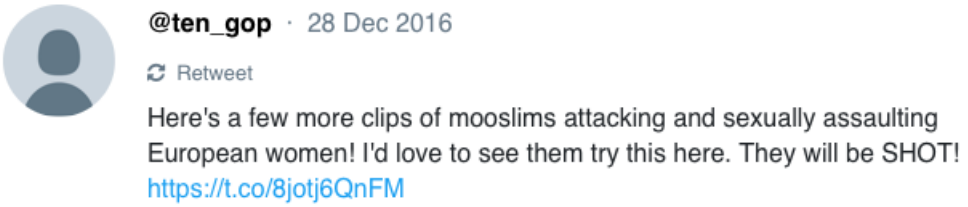


Figure 1.8

Figure 1.9

The following tweets defend specific police shootings of Black men:

Figure 1.10

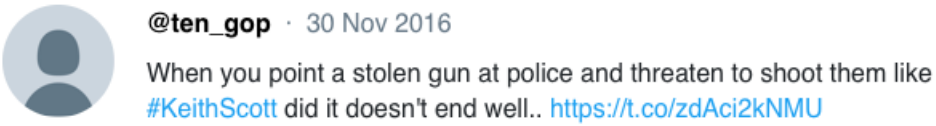


Figure 1.11

The @TEN_GOP fits the category of Right Troll. The 2017 suspension caused further far-right outrage because followers believed that Twitter banned an official state GOP account, not a Russian bot.

The success of the @TEN_GOP account is also attributed to the fact that members close to Trump frequently engaged with the account. Donald Trump Jr., Eric Trump, Kellyanne Conway, and Roger Stone all retweeted and mentioned @TEN_GOP multiple times [5]. This again calls back to our earlier discussion on successful disinformation campaigns. The @TEN_GOP tweets emotionally appealed to far-right influential members who then interacted with the tweets, effectively lending their credibility to the harmful messaging. This cycle bolstered the @TEN_GOP disinformation network.

Investigating tweets by Left Troll accounts also shows how they aided the IRA's operation. We see the following Left Troll account support the racial justice movement:

Figure 1.12

The account also attacks Trump:

Figure 1.13

Figure 1.14

The account further attempts to subvert Clinton's campaign:

Figure 1.15

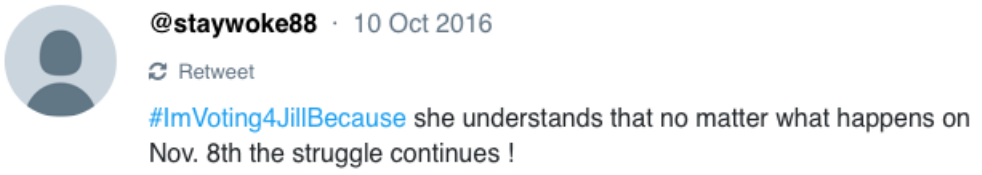


Figure 1.16

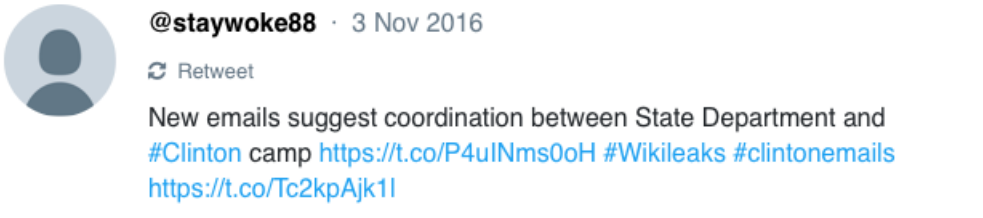


Figure 1.17

The tweet above (Figure 1.17) discusses Clinton's email investigation which became the election's "October surprise." All sides of the IRA accounts heavily mentioned and perpetuated misinformation about the email investigation while they rarely mentioned the leak of an Access Hollywood tape in which Trump brags about sexually assaulting women. Unsurprisingly, the Russian disinformation campaign is suspected to have motivated James Comey's (former FBI director) email investigations in July and ten days prior to the election [22]. Heading into November, Comey's actions and subsequent media coverage distracted from Trump's assault allegations, upending Clinton's momentum and salvaging Trump's campaign. Trump then won the election (despite losing the popular vote).

Overall, the Russian disinformation operation successfully influenced the 2016 presidential election. Instead of tampering with voting machines, the Russian operation meddled with the voters. The social media infiltration evident in the above tweets corroborates Senate Intelligence Committee vice-chair Senator Mark Warner's sentiments that "the extremes are screaming while the majority whispers" [22]. Extreme views elicit raw emotional responses that draw out reactions and push people further apart. This is precisely the goal of a disinformation campaign. Thus, Russian disinformation succeeded by amplifying extremist views that "persuaded enough people to either vote a certain way or not vote at all" [22].

This influence operation leaves us with two questions regarding the 2020 presidential election and misinformation on Twitter. First, what has Twitter done since 2016 to prevent misinformation spread? Second, did misinformation affect the political momentum of presidential campaigns in the 2020 election? We investigate the first question below and address the second question in the remainder of this thesis.

1.4.3 Twitter's Misinformation Response

Since 2016, Twitter has addressed misinformation on their platform. In June, 2017, Twitter wrote a short blog post regarding their approach to bots and misin-

formation. Company representatives accept that the social media network is rife with false information, but they do not take responsibility to correct it, stating "we, as a company, should not be the arbiter of truth" [8]. They also acknowledge that manipulative bots exist and claim they are building new tools to detect and remove them.

In regard to political misleading information, Twitter published additional blog posts with new features. First, in 2019, Twitter claimed that "any attempts to undermine the process of registering to vote or engaging in the electoral process is contrary to our company's core values," and responded by adding a new flag to report misleading information about voting [48]. This feature simply extends the ability to report a tweet. Over a year later, Twitter announced that they were improving account security for major U.S. officials and journalists prior to the 2020 election. The extent of the improved security was the addition of two-factor authentication [47].

Twitter's most substantial new addition to combat misinformation developed from the COVID-19 pandemic. From the onset of the pandemic, a plethora of misleading information has been promoted on Twitter (including from President Trump's account). Thus, in an effort to "make it easier to find facts and make informed decisions about what people see on Twitter," the company created "labels and warning messages that provide additional context and information on some Tweets containing disputed or misleading information related to COVID-19" that may also be utilized to "provide additional explanations or clarifications in situations where the risks of harm associated with a Tweet are less severe but where people may still be confused or misled by the content" [37]. Both labels and warnings can be seen on the President's tweets (Figures 1.5-1.7). This feature effectively retracts Twitter's 2017 sentiment as they begin to grow into their role as a truth mediator. Twitter provides the following charts to understand how and why a tweet may be labeled:

Is the media significantly and deceptively altered or fabricated?	Is the media shared in a deceptive manner?	Is the content likely to impact public safety or cause serious harm?	
✓	✗	✗	Content **may** be labeled
✓	✗	✓	Content is **likely** to be labeled, or **may** be removed.
✓	✓	✗	Content is **likely** to be labeled.
✓	✓	✓	Content is **very likely** to be removed.

Figure 1.18

Misleading Information	Label	Removal
Disputed Claim	Label	Warning
Unverified Claim	No action	No action*
	Moderate	Severe
Propensity for Harm		

Figure 1.19

Finally, confirmed leak reports detail that Twitter is developing Birdwatch, a crowdsourced tool to report misinformation. The company has not made any official remarks as of late 2020, but the tool aims to provide more context for tweets in the form of notes. Currently, it is unclear who will have the ability to annotate tweets. [29]

1.5 Moving Forward

The above section details minor changes that Twitter has implemented since the 2016 elections. The Birdwatch tool will be the biggest step in combating misinformation on the platform; however, it was not launched before the 2020 elections.

Therefore, the rest of this thesis considers the effectiveness of the minor advancements to hinder misinformation spread on Twitter in the 2020 presidential election.

This thesis aims to better understand the characteristics of political Twitter as it relates to prominent campaigns and how misinformation affects their momentum. Specifically, we investigate the 2020 presidential election between Donald Trump and Joseph Biden. We first review literature on how Twitter data has been previously used to analyze politics. This chapter covers how past work has used Twitter to uncover political campaign characteristics, to measure political momentum, to analyze sentiment, and to identify misinformation trends. We then address our data collection process, including how we gather data, what data we use, and how we refine the data.

Following the details of our data, we describe the methodologies applied in our analyses. Then, we present our findings on the candidates' social presence, campaign momentum, and the effect of misinformation. We find differences in the candidates' tweeting behaviors and emotions they convey. We also observe changes in momentum influenced by political events as well as misinformation. Finally, we summarize the takeaways from this thesis, propose future research directions, and conclude our research on the 2020 elections via Twitter.

Chapter 2

Literature Review

In this thesis, we use Twitter data to analyze politics through the lens of online campaign presence and misinformation. Thus, we review existing literature in order to better understand the field and to build tools for use in our research. These topics include: analyzing politics from Twitter data, analyzing tweet sentiment, measuring political campaign momentum, and detecting false information trends. We first examine previous methods for analyzing politics on Twitter to identify characteristics to explore. This research analyzes the sentiment of tweets to both understand overall campaign attitude and to understand the way in which other users address the candidates, which sparks further review of tweet sentiment analysis. We then look into measuring political campaign momentum via Twitter in order to understand how the momentum of presidential campaigns change throughout the election period. Finally, we conclude the literature review by investigating strategies to detect false information trends on Twitter.

2.1 Analyzing Politics on Twitter

In the past few years, analyzing politics via Twitter has become a widening research field. The 2016 U.S. and Austrian elections have both been analyzed through a combination of sentiment analysis and networking methods. Kušen and Strembeck in *Politics, sentiments, and misinformation: An analysis of the Twitter discussion on the 2016 Austrian Presidential Elections*, outline a clear

research plan to understand the characteristics of the two competing campaigns through Twitter data [18]. They first look at the tweeting behavior of each candidate, including temporal characteristics (such as the total number of daily tweets per candidate), the engagement style, and the campaign style. Examining the engagement style involves looking for commonly used words or phrases, common sentiment, and user reactions through favorites, retweets, and replies. The campaign style is categorized as either positive or negative: a positive campaign tweets information about future plans and past successes while a negative campaign spreads misinformation and attacks the opponent. Analyzing tweet sentiment in messages addressing one's opponent helps determine the campaign style.

The second part of Kušen and Strembeck's research aims to understand how other users mention the candidates. They construct ego networks (graphs centered at a pre-determined node) from hashtags to visualize the sentiment and associated topics surrounding the candidates. Each hashtag is a node and all hashtags present in one tweet are connected by edges. Additionally, the networks are centered at a candidate hashtag. The hashtag nodes are sized relative to their frequency and are categorically colored to be either supporting, against, important, or general. The constructed ego networks in the paper show a noticeable difference in the way users refer to the two candidates in the Austrian election. The winning candidate has fewer nodes, and his associated hashtags are often more positive. The losing candidate has visibly more hashtags in his network that directly oppose him. These visualizations reveal insights about how Twitter users view the candidates and what topics users associate with each campaign.

Overall, Kušen and Strembeck find that the winning candidate typically authored neutral scoring tweets while his opponent published more emotional messages. They also find that negative information spreads faster and longer compared to positive or neutral information. This finding makes sense given that emotional and false information have more engaging qualities. The researchers note that "the emotional dimension of a social media discussion is of particular importance as an emotional debate over a controversial topic often develops more dynamically

and unpredictably than an objective discussion," which highlights interest in false information's ability to affect a campaign [18]. Finally, Kušen and Strembeck find that the winner's tweets had more favorites and retweets while the opponent had more replies. This reinforces our intuition that a favorite or retweet is a vote of confidence and agreement while a reply is a way to discuss or express discontent with a tweet. We adapt these methods to explore the Biden and Trump campaigns' characteristics in the 2020 election.

2.2 Tweet Sentiment Analysis

Sentiment analysis uses natural language processing to identify subjective information about a body of text such as the overall tone or the emotions present. The most basic task classifies text as positive, negative, or neutral. More advanced tasks seek to classify the emotional states that text portrays, such as happy, sad, surprised, fearful, etc. In Chapter 1, we discussed how emotional tweets drive connection between individuals and elicit stronger responses from both allied and opposing sides. Therefore, understanding the emotional context of tweets is important to analyze Twitter data.

Sentiment classification of tweets is challenging because the nature of text on Twitter differs from traditional text sources which are typically longer and grammatically formal. Tweets are restricted to 280 characters (circa 2017), yet, the majority of users use less than 50 characters per tweet [36]. This means that the body of text to analyze is very short! (E.g., the previous sentence is 58 characters). This brief length means that there are hardly any words from which to derive sentiment. Thus, each word has a lot of power to sway the tone of a tweet. Additionally, tweets are commonly laden with irony, adding another layer of difficulty to the classification task. These challenges make tweet sentiment analysis an important area of research to review.

Methods for sentiment analysis of Twitter range from a sentiment score lookup

dictionary to deep learning-based models. The shared initial step for many of these methods is to preprocess the tweet text. Preprocessing techniques help reduce noise and reorganize content so that the classification systems can better understand the message. In a handful of early papers on Twitter sentiment analysis, Agrawal et al. and Angiani et al. showed increased accuracy in sentiment classification when they stemmed words during preprocessing [1, 2]. Stemming is the process of reducing a word to a corresponding base word by removing a suffix. For example, "great," "greatly," and "greatest" would all fall into the word stem bucket of "great." This technique reduces entropy and increases the prevalence of the word stem. Angiani et al. compared stemming to stopword removal, emoticon-to-word transformation, negation handling, and acronym expansion. They find that stemming improved classification accuracy on two separate test sets by approximately 1-4% compared to the other techniques. Transforming emoticons into text strings is another common technique seen to improve accuracy in the BERT (Bidirectional Encoder Representations and Transformers) model structures of Nguyen et al. and Azzouza et al. [26, 4]. The former along with Zhao et al. also find that mapping URLs and user mentions to predetermined tokens effectively reduces noise [16].

Once the text is preprocessed, there are a plethora of methods to classify sentiment. The most basic approaches use Support Vector Machine, Multinomial Naive Bayes, and Maximum Entropy classifiers [32]. However, these methods benefit from an extensive vocabulary resulting in high training cost. Thelwall et al.'s SentiStrength tool is a widely used lookup dictionary-based model to extract sentiment from informal English text [44]. The tool classifies the strength of both the positive and negative sentiments by returning two scores from a 1 to 5 scale (1 is neutral, while 5 indicates strong positive or negative emotion). The core algorithm scores words by looking them up in a predetermined sentiment word strength list. The tool also corrects words with too many repeated letters (e.g. hellooooooo $\rightarrow$ hello) and simultaneously boosts the sentiment score to account for the effect of repeated letters. Other words contained in a booster list can also increase or decrease the score (e.g., very, extremely). The tool is further trained

by applying an algorithm that optimizes individual word strengths by testing if a one point increase or decrease in sentiment score would increase the overall classification accuracy. SentiStrength predicted positive emotion with 60.6% accuracy which beat all of the earlier listed simple models. It accurately classified negative emotion with 72.8% accuracy which beat most simple models except SVMs and simple logistic regression.

Newer sentiment techniques use word embeddings and deep learning. Word embeddings map words to vectors such that words with similar meaning have close vector representations. On top of this technique, deep neural network architectures have greatly improved sentiment analysis and other NLP tasks. Convolutional Neural Nets (CNNs) have been researched for text processing, showing an improvement over lexicon-based methods as in Pota et al.'s work in classifying political tweets [31]. However, CNNs still remain best fit for image processing while Recurrent Neural Networks (RNNs) are better for modeling sequential data. Long Short-Term Memory (LSTM) RNNs have shown the best performance in sentiment classification since they can overcome the vanishing gradient problem [32]. Currently, the state-of-the-art language model is Google's 2019 BERT (Bidirectional Encoder Representations and Transformers) model. BERT models are designed to understand short phrases, making it a well-suited model for tweets. Incredible performance comes from BERT's ability to understand the context of a phrase [25]. Each word is processed in relation to the rest of the sentence helping grasp a fuller understanding of the text. In their paper *BERTweet: A pre-trained language model for English Tweets*, Nguyen et al. show the effectiveness of a BERT architecture in classifying the sentiment of tweets. They find a 5% and 4% classification accuracy improvement for two SemEval (Semantic Evaluation) test sets.

2.3 Measuring Political Momentum

Momentum strategy is a popular market investment strategy that aims to buy shares when they are gaining momentum and sell shares once they have reached

a predicted peak. In the paper *Using Social Media and News Sentiment Data to Construct a Momentum Strategy*, Shi constructs four momentum strategies using social media sentiment data [41]. The resulting strategies show improved performance, suggesting that social media data can be used to deduce momentum. Watts et al. adapt this economic strategy to instead model the momentum of political campaigns via campaign Twitter data [52]. They present three indicators as measures of political campaign momentum. We cover four of the proposed indicators; two from Shi's work and two from Watts et al.

The first two indicators are the Sentiment Indicator (SI) and the Curvature Indicator (CI) proposed by Watts et al. The SI compares the candidates' tweets via sentiment count. Each tweet receives a positive, negative, or neutral classification and a corresponding integer sentiment score. The total sentiment count is the sum of all sentiment scores within a time period. Watts et al. use the AFINN library to calculate tweet sentiment. This is a lookup dictionary model similar to SentiStrength. However, instead of bounding the entire tweet sentiment score between [-5, 5], the AFINN library bounds each word's score between [-5, 5], so a tweet's sentiment score can be outside this range. A higher sentiment count indicates positive momentum. Watts et al. validate the use of the SI by showing a high correlation between the SI and the outcome of the primary elections for the 2016 U.S. presidential race. Trump had the highest SI value throughout the primary period and he won the nomination. Sanders had a higher SI value for the first eight debates, but Clinton finished the ninth debate with a higher SI value indicating a momentum shift. She then won the Democratic nomination. Thus, the results support the use of the SI to measure campaign momentum.

The Curvature Indicator (CI) serves to compare the momentum of different candidates by measuring the direction of momentum daily or for a designated time period. For a single, daily value, the CI is either 1 or -1 depending on the curvature of the overall trend. For a time period, the CI is the linear trend of daily CI values over the interval. To calculate the CI, Watts et al. first construct a time

series $P(t)$ from the data. $P(t)$ is the time series of some tweet attribute such as:

$$\begin{aligned} P_s(t) &= \text{ total sentiment count for day } t \\ P_c(t) &= \text{ number of tweets mentioning candidate } c \text{ on day } t \\ P_d(t) &= \text{ number of tweets by candidate } c \text{ on day } t \end{aligned}$$

One crucial assumption is that $P(t)$ is twice differentiable. Thus, the second derivative can be approximated as a second difference as follows with $h = 1$ since one considers a single data point each day:

$$P''(t) = \frac{P(t-h) + P(t+h) - 2P(t)}{h^2}$$

Watts et al. then fit a linear trend to the approximated second derivatives, resulting in the CI momentum trend. A positive slope shows positive momentum, and a negative slope shows negative momentum.

Shi describes two common momentum indicators: Rate of Change (ROC) and Relative Strength Indicator (RSI) [41]. ROC measures the velocity of change over a time interval, τ. Formally, ROC is defined as:

$$\text{ROC} = \frac{P(t+\tau) - P(t)}{P(t)} \cdot 100$$

where $P(t)$ is the time series measure at time t and $P(t+\tau)$ is the measure at τ past t. The $sign(\text{ROC})$ measures the direction of the momentum movement while the magnitude of the ROC measures the strength of the change. The second indicator Shi describes is the Relative Strength Indicator (RSI), originally introduced by Wilder in 1978. Similarly to ROC, RSI measures both the speed and magnitude of $P(t)$ changes. RSI is favorable because it attempts to reduce the erratic movements that are common among momentum indicators by preventing or removing these

distortions. In the economic context of stock price fluctuation, RSI is defined as:

$$\text{Smoothed Average Gain} = \frac{\text{sum of gains in } t \text{ days}}{t}$$
$$\text{Smoothed Average Loss} = \frac{\text{sum of losses in } t \text{ days}}{t}$$
$$\text{Relative Strength (RS)} = \frac{\text{smoothed average gain}}{\text{smoothed average loss}}$$
$$\text{RSI} = 100 - \frac{100}{RS}$$

Other indicators proposed by both Shi and Watts et al. suggest the inclusion of exponential smoothing to decrease the effect of events that happened farther in the past. These momentum indicators provide a tool to measure the momentum of each political campaign and observe whether misinformation affected said momentum.

2.4 Uncovering Misinformation Trends

False information camouflages well online among political half-truths and stories that are hard to debunk. This makes algorithmically flagging online false information challenging. Vosoughi et al. corroborate the wide existence and success of online false information. In their paper, *The Spread of True and False News Online*, they find that novel information is more likely to spread, noting that "falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information" [50]. False information is more novel, and thus is set up for amplification success. For example, Vosoughi et al. found that "falsehoods were 70% more likely to be retweeted than the truth," effectively extending the online success of false information. They also recorded more pronounced effects for false political news.

Research by the Knight Foundation on the effect of disinformation on the 2016 U.S. presidential election outlines a handful of techniques that are used to help uncover false information. First, they note that accounts that spread false information are "extremely densely connected" [11]. Thus, identifying dense networks

that repeatedly link to false stories is one effective method to target the spread of influence campaigns. The report also discusses how coordinated accounts extend the life cycle of targeted news stories or hashtags that opportunistically fit the agenda of the influence operation. A case study on these coordinated bodies showed that "aggressive action against fake news outlets can be effective at containing the spread of fake news" [11].

We can apply these observations to help uncover false information trends. Namely, we can look for densely connected bodies of accounts that tend to share known false stories. We can also look for links or hashtags that are sustained longer than normal. In future attempts to hinder the spread of false information, we should target fake news outlets to stop stories before they can be amplified.

2.5 Takeaways

We acquire several useful insights from the reviewed literature. First, we obtain a clear research outline for analyzing the 2020 presidential election via Twitter. We adapt the methods presented to explore the Biden and Trump campaigns' characteristics, including their respective temporal characteristics, engagement style, and overall campaign style. Then, we develop several ego networks to visualize how other users discuss the candidates.

From reviewing past research in the field of tweet sentiment analysis, we learn that preprocessing tweets is necessary to improve sentiment classifications. This preprocessing step includes general cleanup such as turning emojis into words and mapping URLs to a single URL token. We also learn about different sentiment analysis methods ranging from lexicon-based methods to deep learning networks. SentiStrength is a lexicon-based method that has decent performance for classifying the sentiment of short texts. The BERT model is a state-of-the-art deep learning network that was developed specifically to better understand short texts based on context. These are the two main sentiment methods we apply.

Next, came measures of political momentum. Four reviewed momentum indicators can be adapted to measure political campaign momentum via Twitter: the Sentiment Indicator, the Curvature Indicator, the Rate of Change measure, and the Relative Strength Indicator. The former two have already been modified to measure political momentum on Twitter. The latter two are typically used to measure the momentum of stocks, but we can construct time series Twitter data to instead measure political campaign momentum.

Finally, we understand that false information is hard to uncover online, but that there are certain trends that may indicate misinformation spread. One takeaway is that links or hashtags that have a longer than normal lifespan may suggest suspicious activity.

This literature influences the research we present in the following chapters.

Chapter 3

Data

This chapter outlines the data we use in our analysis of the 2020 presidential campaigns. We first detail the data collection process. This process includes understanding the Twitter API and the data objects that are returned. We then explain what information we retain from the collected objects. Next, we discuss the plausibility of our sample as a representative sample. We then move on to describe how we collect a political sample. The following exploratory data analysis section provides information on the substance of our data. We investigate the size of our daily samples, note missing data, and present related summary statistics. We conclude the chapter by detailing our data selection for analysis.

3.1 Data Overview

3.1.1 Twitter API and Data Structure

The data we analyze in this research consists of tweet objects streamed from Twitter. A tweet object can be an original tweet, a retweet, or a reply. On an average day, approximately 500 million tweets are authored [43]. We collect a daily sample of this corpus using Twitter's API and the Tweepy Python library. Twitter offers multiple tiers of API access that put varying limits on the amount of tweets one can stream. The four main APIs currently available for public use are: v2, Standard v1.1, Premium v1.1, and Enterprise. Several early access endpoints were released for v2 in August, 2020, but Twitter has not yet depreciated v1.1 [6]. Each

API has documentation on Twitter's developer site. However, Twitter obscures the rate limits for every tier, making it hard to figure out how much data one can actually download.

In v2, the basic access level is free. This tier appears to limit a developer to 500,000 tweets a month (approximately 17,000 tweets daily which is roughly a 0.0034% sample). The Academic Research tier within the basic access level increases this limit to 10 million tweets a month (approximately 333,000 tweets daily, roughly a 0.067% sample). Similarly to v2, Standard v1.1 has a 500,000 tweets per month cap. Premium v1.1 increases the monthly limit and number of tweets per request, but requires a paid subscription. Both the pricing and higher rate limit details are only available upon request. The Enterprise API provides the highest level of access. For example, Enterprise developers have access to stream a 10% random sample of all publicly available tweets. A 10% sample is roughly 50 million tweets a day. Additionally, Enterprise developers have access to search a full archive of public tweets. Enterprise subscriptions and pricing information are only available upon application.

We access the Standard v1.1 API via the Tweepy Python library. We create a political filtered stream which allows us to select relevant tweets from the real-time, public stream. Our filter is based on political Twitter handles which we cover in more detail in Section 3.1.3. We run seven concurrent streams. The decision to run seven concurrent streams is based on trial-and-error to balance Twitter's API limits and the entirety of our political filter. Each of the seven streams makes a request for data, then sleeps for a random period of time between one minute and one hour. This sleep is necessary to maintain and process a manageable amount of data. Without the sleep, the data files are intractable.

The request returns data as tweet objects. These objects are messy and contain a plethora of unnecessary fields. For example, the following tweet is encoded by the subsequent object. (The compressed fields can be found in Appendix A.1, A.2, and A.3. Notice that the user field contains a user object that includes 38

additional fields shown in Appendix A.3.)

Figure 3.1 Example Tweet

```
▼ object {30}
    created_at : Wed May 06 03:44:48 +0000 2020
    id : 1257878959573868500
    id_str : 1257878959573868544
    text : @realDonaldTrump @honeywell https://t.co/O6tAvOwtmS
  ▶ display_text_range [2]
    source : <a href=\"http://twitter.com/download/android\" rel=\"nofollow\">Twitter for Android</a>
    truncated : false
    in_reply_to_status_id : 1257873257732214800
    in_reply_to_status_id_str : 1257873257732214784
    in_reply_to_user_id : 25073877
    in_reply_to_user_id_str : 25073877
    in_reply_to_screen_name : realDonaldTrump
  ▶ user {38}
    geo : null
    coordinates : null
    place : null
    contributors : null
    is_quote_status : false
    quote_count : 0
    reply_count : 0
    retweet_count : 0
    favorite_count : 0
  ▼ entities {5}
    ▶ hashtags [0]
    ▶ urls [0]
    ▶ user_mentions [2]
    ▶ symbols [0]
    ▶ media [1]
  ▼ extended_entities {1}
    ▶ media [1]
    favorited : false
    retweeted : false
    possibly_sensitive : false
    filter_level : low
    lang : und
    timestamp_ms : 1588736688880
```


Figure 3.2 Tweet Object of Example Tweet

Since the tweet in Figure 3.1 is a reply, we see the "in_reply" fields populated in the corresponding tweet object shown in Figure 3.2. If this tweet was a retweet, the object would include an additional field containing the full tweet object of the retweeted tweet. An original tweet contains neither the "in_reply" fields nor the extra retweet object. Accumulating hundreds of thousands of these tweet objects results in quite large data files. To combat this problem, we create a thinned object that contains only the fields we want to analyze. Specifically, we retain the following fields: `text`, `tweet_id`, `hashtags`, `mentions`, `in_reply`, `retweet`, `retweet_count`, `favorite_count`, `favorited`, `quote`, `quote_status`, `user`. The

retweet field stores a thinned retweet object, and the user field stores a thinned user object.

We thin the tweet objects as they stream in. At the end of a day, we combine the data from the seven streams into one daily dataset. This combination requires deduplication and the removal of empty and error-filled objects. The basic access tier does not always grant a reliable stream, meaning that objects can be empty, incomplete, or filled with other errors. The Enterprise tier offers the highest level of reliability. Stream disconnection is another potential problem. This means that we have to monitor our streams so that we can reconnect as quickly as possible in case of a disruption.

With these issues managed, we produce a daily dataset of approximately 300,000 tweets. Since we wish to use the data for inference and analysis, the natural question becomes: is our roughly 0.06% sample sufficiently representative of Twitter?

3.1.2 Representative Sample

In this subsection, we examine whether or not our daily dataset is representative of the full, 100% daily Twitter stream. Theoretically, since we are taking a random sample of Twitter, our datasets should be statistically similar. The following analysis aims to confirm this hypothesis.

In 2019, Kalev Leetaru examined the similarity between the 100% and 1% Twitter streams [19]. Leetaru does not explicitly outline his methods, but he gives a general overview of the process that led him to conclude that the 1% sample is a fairly accurate representation of the full 100% stream. He considers each day as one data point in the range of January 2012 to October 2018. Then, he compares the daily frequency of three keywords. Leetaru selects one common topic that should appear often ("climate change"), one small-sized topic ("global warming"), and one micro-sized topic ("IPCC"). He then tests to see if a positive correlation exists between the number of keyword mentions in the 1% sample and

the 100% stream. If the sample is representative, a strong, positive correlation should exist. Leetaru finds that all three topics have strong, positive correlations (see Table 3.1). Unsurprisingly, an extremely infrequent, fourth topic shows not as strong a correlation as the common, small, and micro topics (see row 4 of Table 3.1).

Keyword	Daily Average Frequency in 1% Sample	Correlation Between Sample and Full Stream
climate change	33,500	0.96
global warming	10,500	0.94
IPCC	850	0.94
carbon sequestration	0-10	0.79

Table 3.1: Leetaru Correlations

In a perfect world, we would mimic Leetaru's analysis to determine if our sample is representative. However, we do not have nearly as much data, but we do have three "varieties" of datasets. Each variety of dataset was collected using the same filtering scheme. They differ due to a variable random sleep implemented in order to throttle the daily sample size. These three varieties are as follows:

Variety Name	Random Sleep	Sample Size	Collection Dates
Large	None	∼1% (5 million)	06/24/2020
Small	1 min - 1 hour	∼0.06% (300,000)	08/26/2020 - present
Micro	1 min - 2 hours	∼0.03% (150,000)	06/26/2020 - 08/25/2020

Table 3.2: Dataset Varieties

Following Leetaru's work, let's assume that our one day of 1% sampling is representative of the full stream. In other words, the large sample is representative of a true day on Twitter. We now want to determine if the small and micro samples are representative of the 1% sample. If true, by transitivity, we can conclude that the small and micro samples are representative of the full stream.

We will perform two Chi-square tests of equal proportions to see how well the data from the micro sample fits the distribution from the large, representative sample.

The quantitative measures we look at are tweet word count and proportion of stopwords per tweet. We refrain from using candidate-related measures (such as the average number of Biden or Trump mentions) since we should expect this statistic to fluctuate based on the news cycle and campaign momentum. Before performing the tests, we look at the distribution of word count and stopword proportion across all three dataset size varieties.

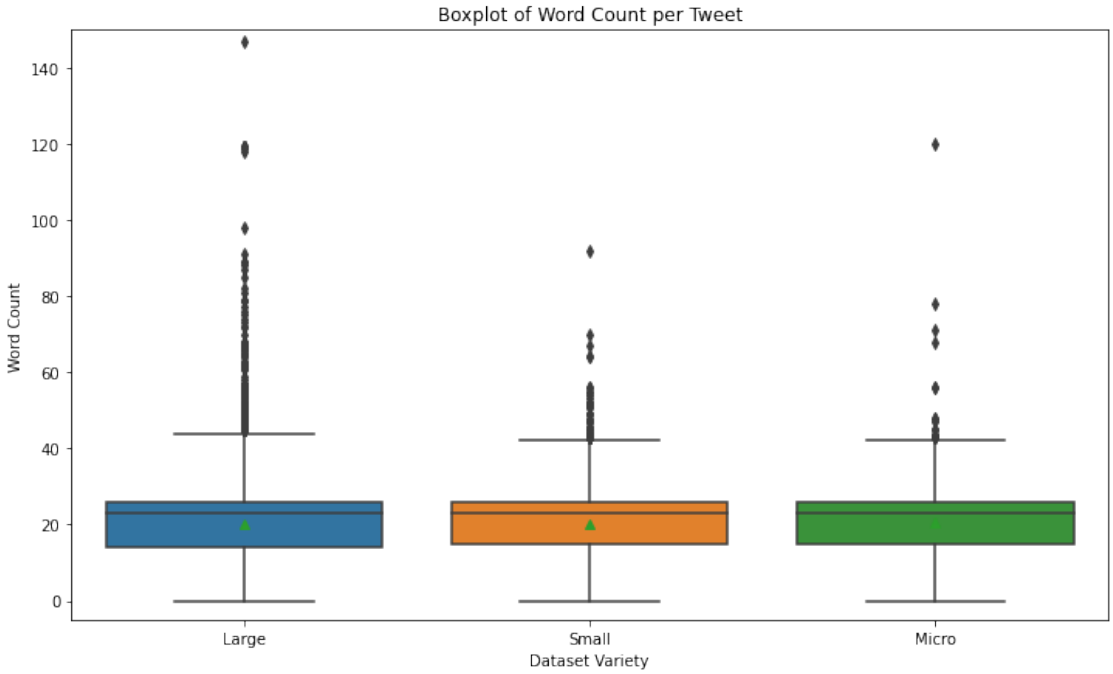


Figure 3.3 Boxplot of Word Count per Tweet

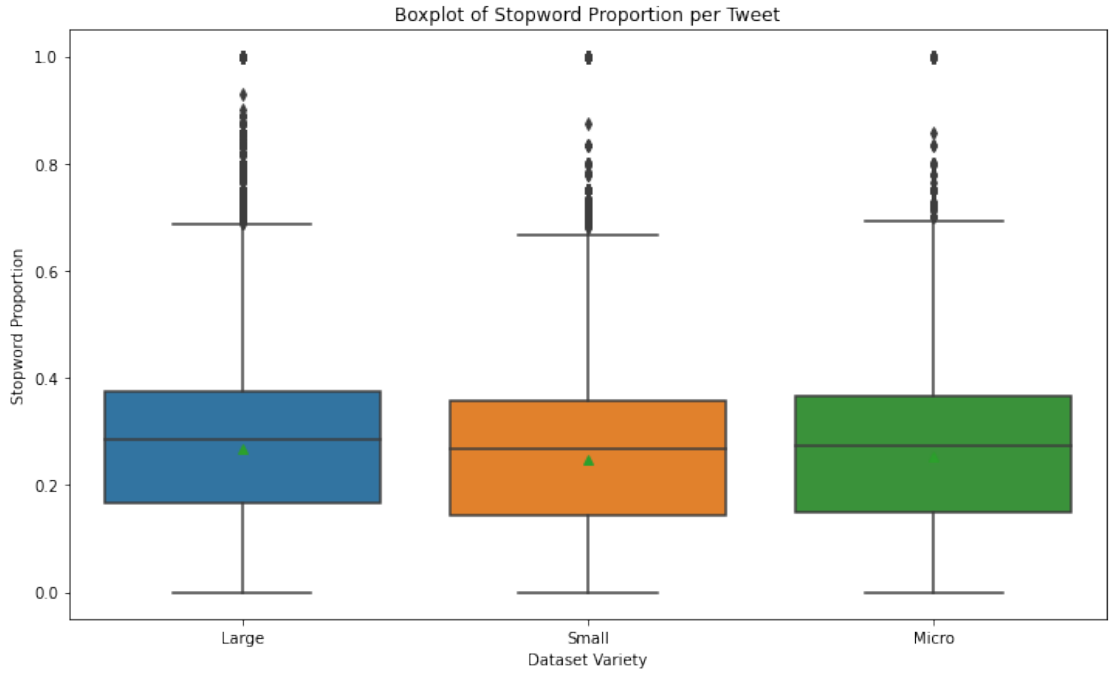


Figure 3.4 Boxplot of Stopword Proportion per Tweet

The distributions look decently similar. For both measures, the four quartiles match quite closely across all three datasets. The means (denoted by the line in the box) and medians (denoted by the green triangle in the box) also align. Now, we perform the Chi-square tests of equal proportions.

For both tests, we split the data into quintiles based on the large sample. Then, we compare the number of tweets that fall into each quintile via the test of equal proportions. The following tables present the contingency tables for the two quantitative measures. The quintile values are listed below the quintile label in Tables 3.3 and 3.4. The contingency table based on word count is as follows:

	Quintile 1 [0,12]	**Quintile 2** (12, 20]	**Quintile 3** (20, 24]	**Quintile 4** (24, 27]	**Quintile 5** (27, 304]	**Total**
Micro	29,426	30,408	33,430	35,144	25,592	154,000
Large	1,017,253	915,083	993,832	1,043,075	815,200	4,784,443
Total	1,046,679	945,491	1,027,262	1,078,219	840,792	4,938,443

Table 3.3: Chi-Square Contingency Table Based on Word Count

The Chi-square value is $\chi^2_{df:4} = 505.75$ and the p-value is $p < 0.0001$. The effect

size via Cohen's W (which is the effect size measure for tests of equal proportions) is $W = 0.01012$. Thus, even though the Chi-square test concludes that the micro and large samples are statistically different based on word count, the very small effect size shows that this significant difference is practically trivial. Let's compare the relative distributions to double-check the validity of this claim that the distributions are roughly the same.

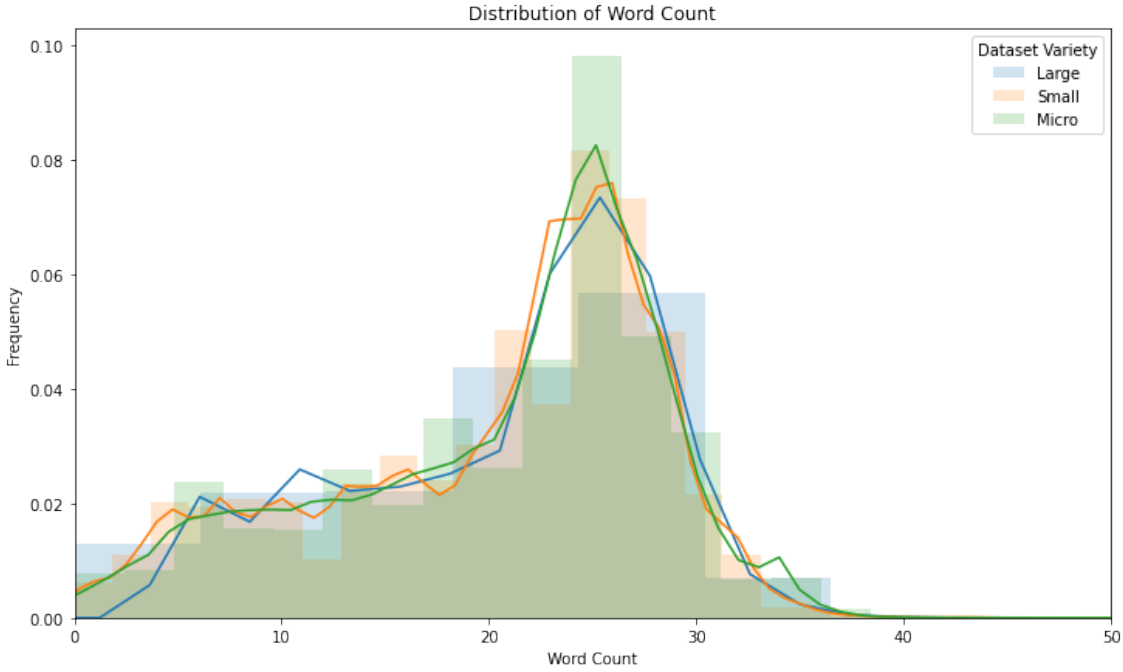


Figure 3.5 Distribution of Word Count

Figure 3.5 shows that as expected, the relative distributions are roughly the same for word count.

Next, let's perform the same analysis for the proportion of stopwords. The contingency table based on stopword proportion is as follows:

	Quintile 1 [0, 0.12]	**Quintile 2** (0.12, 0.25]	**Quintile 3** (0.25, 0.32]	**Quintile 4** (0.32, 0.4]	**Quintile 5** (0.4, 1]	**Total**
Micro	33,867	36,391	27,727	31,216	24,799	154,000
Large	962,276	1,071,672	844,083	1,026,110	880,302	4,784,443
Total	996,143	1,108,063	871,810	1,057,326	905,101	4,938,443

Table 3.4: Chi-Square Contingency Table Based on Stopword Proportion

The Chi-square value is $\chi^2_{df:4} = 898.69$ and the p-value is $p < 0.0001$. The effect size via Cohen's W is $W = 0.01349$. Similarly to the word count analysis, even though the Chi-square test concludes that the micro and large samples are statistically different based on stopword proportion, the very small effect size demonstrates that this significant difference is practically trivial. Once again, let's compare the relative distributions to double-check the validity of this claim that the distributions are roughly the same.

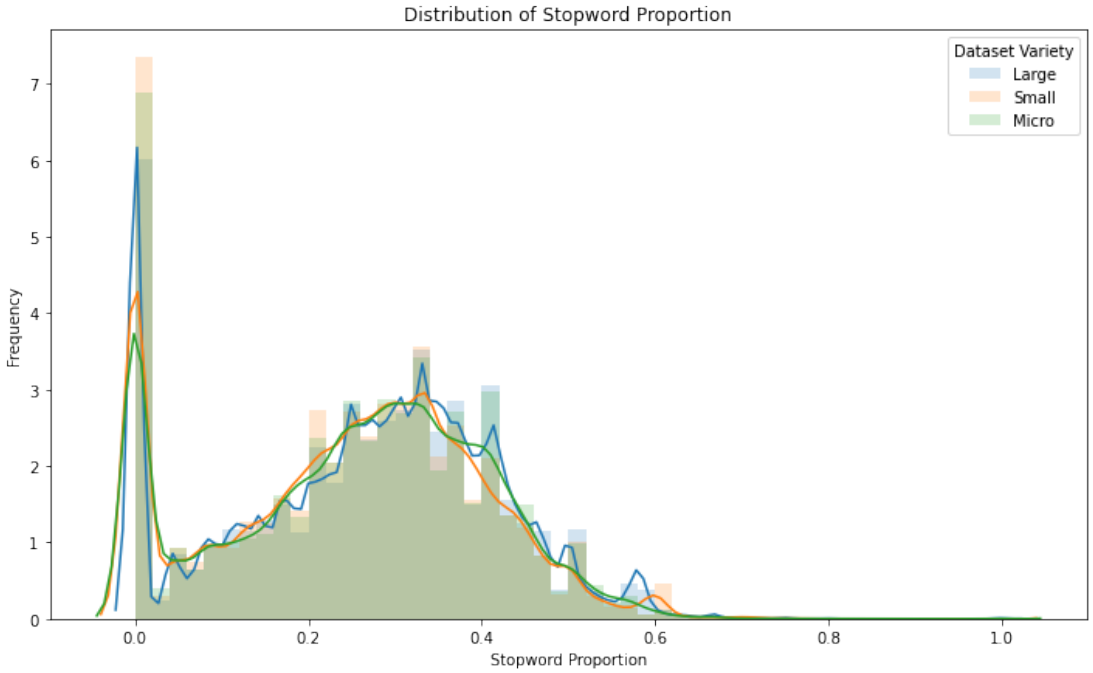


Figure 3.6 Distribution of Stopword Proportion

Figure 3.6 shows that as expected, the relative distributions are roughly the same for stopword proportion. Therefore, even though our tests show significant difference between the micro and large samples for both of our quantitative measures, the very small effect sizes show that the difference is practically trivial. We conclude that the micro sample is roughly representative of the large sample. It then follows that the small sample is also roughly representative of the large sample. Overall, the micro and small samples are both sufficiently representative of a day on Twitter.

3.1.3 Data Selection

Concluding that our samples are roughly representative of Twitter actually means that our samples are representative of a specific subset: political Twitter. We effectively subset our stream by constructing a filter to select tweets related to politically affiliated Twitter accounts. We specifically select this political sample because we want to analyze the political sphere. The handles in our filter are taken from pre-made Twitter lists (lists that group related accounts together such as us-house which groups together all house members) and manually added accounts. The lists and categories of manually added accounts are:

TwitterGov
- us-house
- us-senate
- us-cabinet
- us-cities
- us-election-officials
- us-presidential-candidates
- us-governors

News
- C-Span
  - political-reporters
- Fox News
  - fox-accounts
  - shows-hosts
  - all-fox-news
- CNN
  - bestpoliticalteam
  - cnn-news
- MSNBC
  - correspondent-list
  - msnbc-shows-and-hosts1
- New York Times
  - new-york-times-politics
  - elections-2020
- Wall Street Journal
  - wsj-politics
- ABC
  - abc-news-staff
  - politics
  - mediaorgs
- NBC News
  - politics-reporter-embed
  - archive-election-embeds
  - nbc-news-msnbc-brands

News cont.
- CBS News
  - cbs-news-2020-politics
  - cbs-news-politics-team
  - cbs-news-team
- Slate
  - right-leaning-tweets
  - left-leaning-tweets

Pundits
- PunditFact
  - political-pundits
- Max Temkin
  - u-s-politics-2020
- Matt Lewis
  - conservatives
  - liberals
  - media
  - political-journalists
- Ari Herzog
  - alt-and-rogue1
  - white-house-reporters1

Manually Added Categories:
- Candidate family members
- Prominent political and media figures
- Candidates in contentious house/senate races
- Prominent activist/interest groups
- Media organizations to ensure representation of the entire political spectrum

Figure 3.7 Categorized Twitter Lists for Filtered Stream

In total, the final filter includes 2,684 accounts (see Appendix A.4 for full list of accounts).

Our analysis further focuses on the Democratic and Republican presidential candidates: Joe Biden and Donald Trump. Thus, we additionally reduce our data during analysis. Specifically, within our sample, we look at the tweets authored by each candidate (@JoeBiden, @realDonaldTrump), the interaction with candidate tweets (i.e., retweets of and replies to candidate tweets), and any tweet that contains "biden" or "trump" (we lowercase the tweet before checking for these keywords in order to capture all capitalizations). The next section provides an exploratory data analysis adding details to this choice of data selection.

3.2 Exploratory Data Analysis

3.2.1 Data at Large

This EDA and the research in the following chapters analyze data between the dates of August 1, 2020 through November 7, 2020. This range allows us to examine important events starting from the Democratic and Republican National Conventions (DNC, RNC) and ending with the day that the Associated Press called the election. These dates mean that we use both micro and small sized datasets since we only have micro datasets before August 26th. Figure 3.8 shows the variety in the sample sizes.

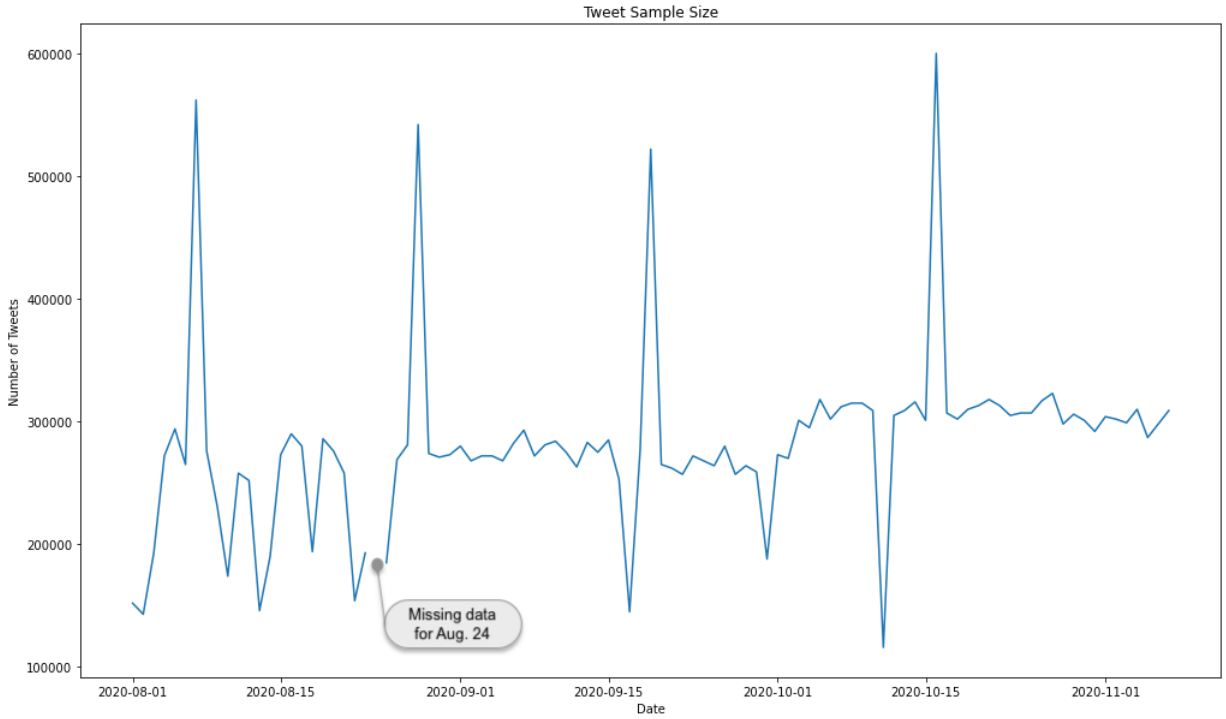


Figure 3.8 Tweet Sample Size

First, we want to note the day of missing data on August 24th. This missingness occurred due to an API disconnection. Next, notice how the sample size is much more erratic from August 1st to August 25th. This makes sense because this range corresponds with the micro dataset which employs a random sleep between one minute to two hours. Once we begin using the small dataset (August 26th to November 7th), we see a much more uniform pattern, yet, a few days still dip or spike. These dips and spikes can be attributed to the randomness of the one minute to one hour sleep of the seven concurrent streams. While the above figure reflects the sample sizes, some of the tweet objects likely contain errors and cannot be processed for analysis. Therefore, we show the true amount of accepted tweets plotted against the sample size in Figure 3.9.

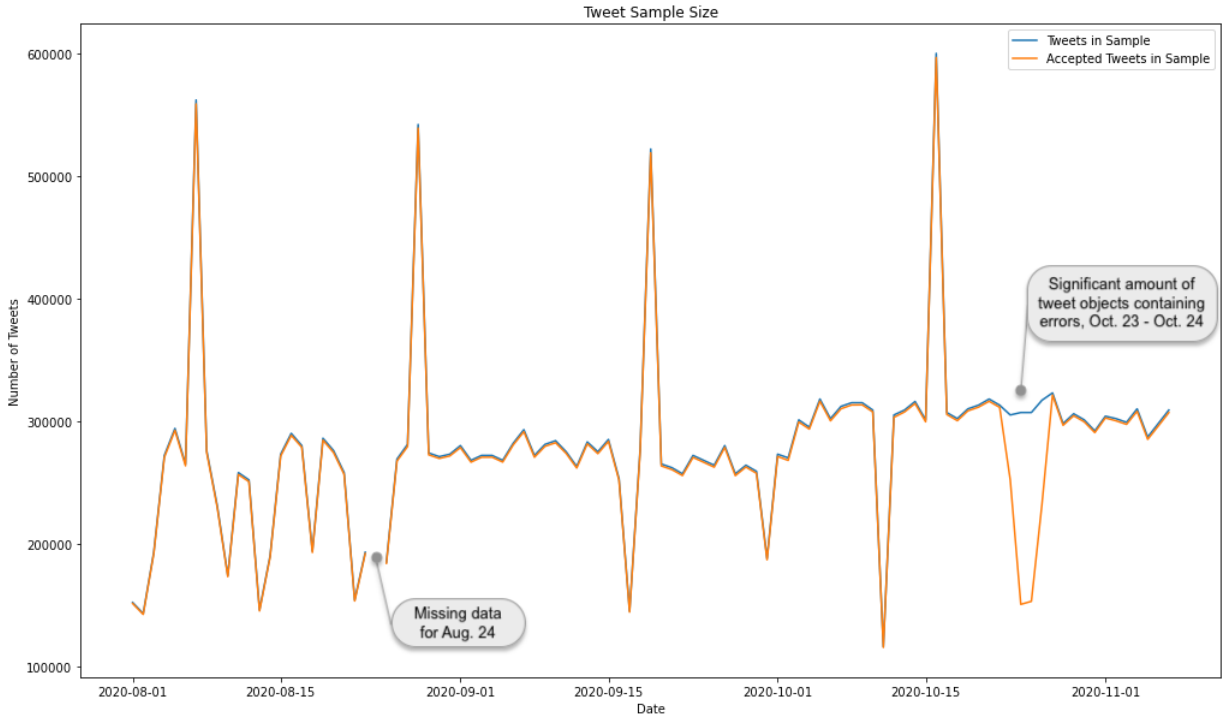


Figure 3.9 Sample Size of Valid Tweets

The amount of accepted tweets in each sample is just slightly below the true size of the sample. This shows that every sample contains a handful of tweets that contain errors blocking the tweet from being processed correctly. The only surprising dates are October 23rd and 24th. Even though a normal amount of tweets were collected, a large amount contained errors. These two days of errors are attributed to a bug within our request. On October 22nd, we noticed that Twitter automatically returns the tweet text truncated to 140 characters (a remnant of the original Twitter API when 140 characters was the tweet limit). We edited the API request to return the untruncated version on October 23rd, but simultaneously introduced a minor bug that affected our thinned tweet objects. The error was caught and fixed by October 26th. The following table presents summary statistics:

	Sample Size	Accepted Tweets Size	Failed Tweets Size
days	98	98	98
mean	281,408	275,324	6084.56
std	73,112.8	74,844.9	23,783.4
min	116,000	115,383	617
25%	264,250	258,142	1,429.25
50%	280,000	274,435	1,564.5
75%	304,750	300,022	1,739.5
max	600,000	596,600	156,519

Table 3.5: Summary Statistics by Tweet Validity

Next, let us look at how our sample breaks down between original tweets, replies, and retweets. Figure 3.10 shows this composition. The daily data is shown overlaid with a seven day simple moving average (SMA). (From here on out, we do not show the gap for the missing data on August 24th for visual readability).

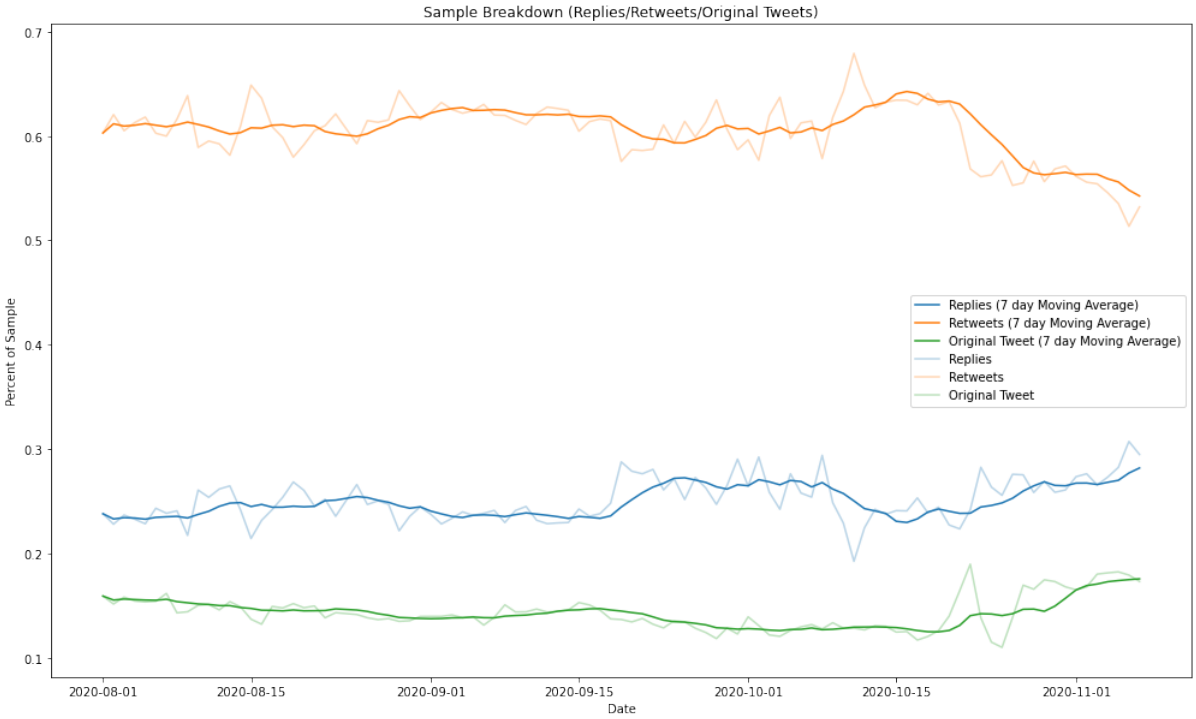


Figure 3.10 Sample Breakdown by Tweet Type

Retweets are the most common type of tweet followed by replies and original tweets. Retweets account for approximately 60% of a daily sample. Replies account for approximately 23%, and original tweets account for approximately 17%

of a daily sample. It is interesting that the amount of retweets decreases from around October 20th through the election, and the amount of replies and original tweets increases. This suggests a potential increase in engaged public discourse. Instead of just sharing someone else's tweet, people are either replying or authoring their own tweets more. The summary statistics of the retweet, reply, and original tweet counts per day are displayed in Table 3.6.

Before focusing on the candidate-related tweets, let us look at the average amount of hashtags and mentions per tweet. The following plots show the mean number of hashtags per tweet in each sample (Figure 3.11) and the mean number of mentions per tweet in each sample along with the respective seven day SMA (Figure 3.12).

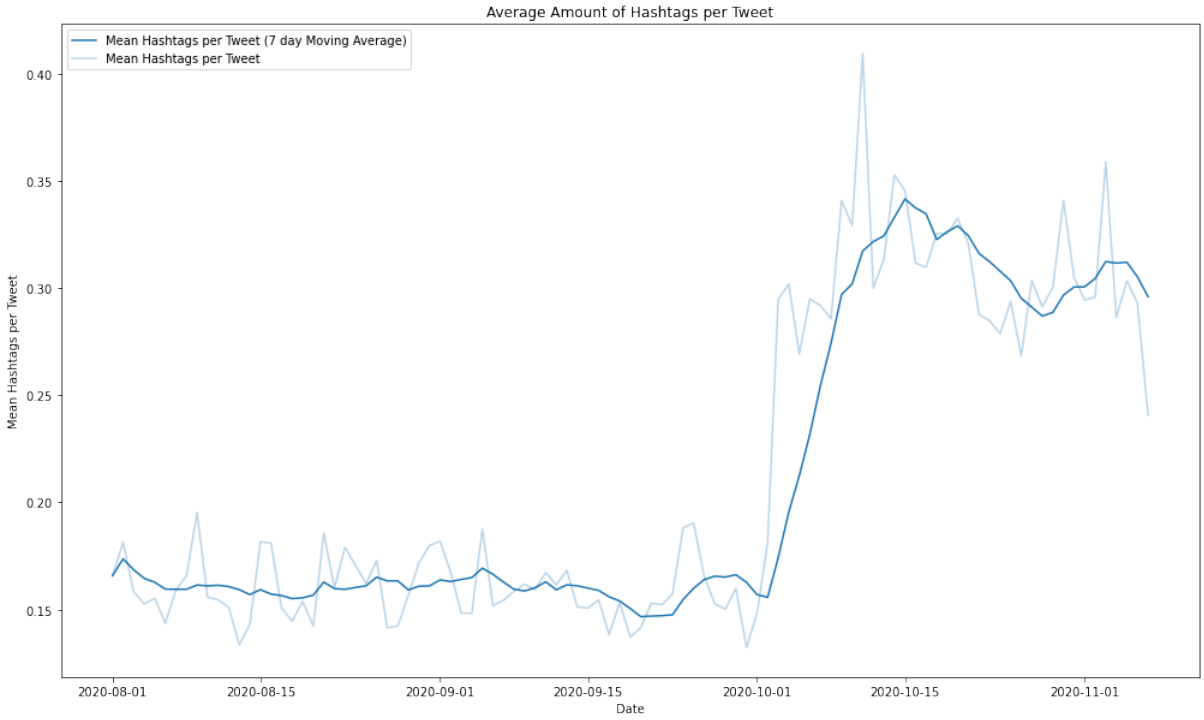


Figure 3.11 Average Amount of Hashtags per Tweet

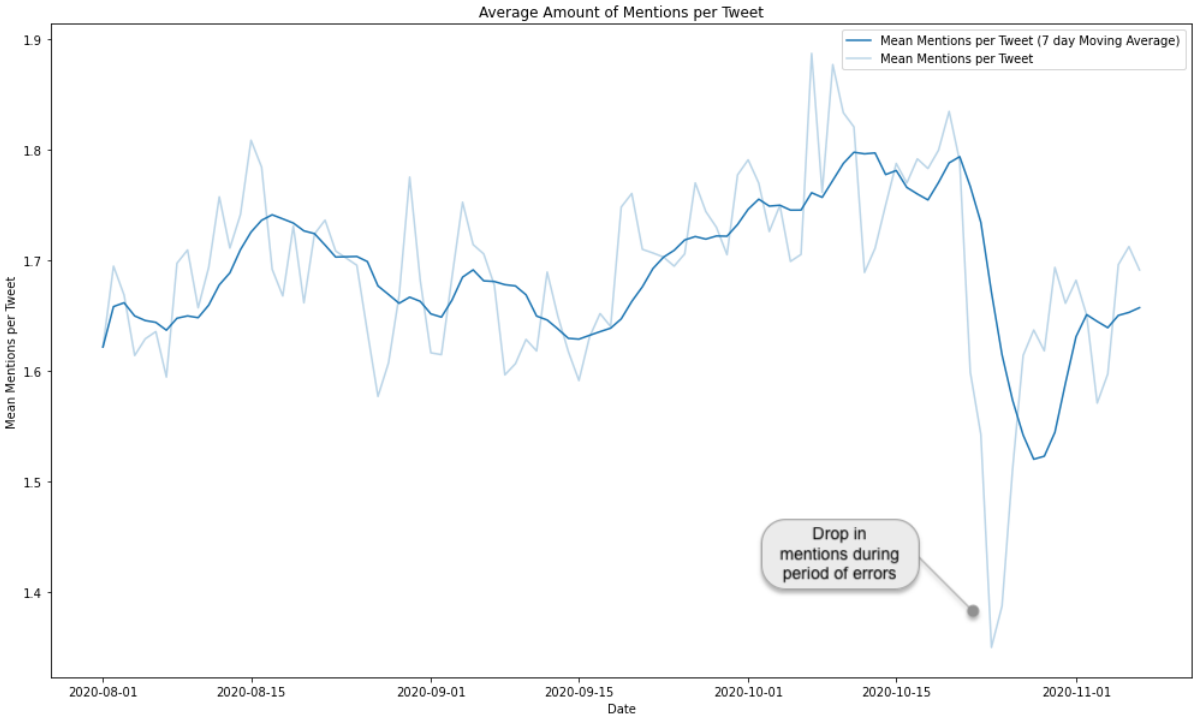


Figure 3.12 Average Amount of Mentions per Tweet

Up until October, a tweet on average contains approximately 0.16 hashtags. This shows that hashtags are not very common among our tweets. However, once October hits, the mean increases to the range between 0.30 to 0.35 hashtags per tweet. This increase is interesting and highlights a potential upsurge in hashtags during the month prior to the election. We see that mentions are more prevalent per tweet. On average, a tweet contains approximately 1.62 to 1.80 mentions. While this appears to suggest that mentions are widely used, we also note that tweets with mentions typically include several handles, and replies automatically tack on the handle of the user that one is replying to. The summary statistics below show the sheer magnitude of mentions compared to hashtags.

	Retweets	Replies	Original Tweets	Hashtags	Mentions
days	98	98	98	98	98
mean	169,791	70,843.2	40,244.2	59,871.4	466,010
std	44,679.3	19,718.2	11,385.3	28,939.3	130,594
min	78,766	22,325	14,909	19,373	203,124
25%	157,189	63,599.8	35,698.8	40,148.8	436,749
50%	170,048	69,972	38,505	46,112.5	462,123
75%	178,932	77,810.2	42,079.2	88,432.5	508,465
max	380,420	150,200	90,890	185,918	1,056,060

Table 3.6: Summary Statistics for Tweet Types and Attributes

3.2.2 Candidate Specific Data

Now that we have explored the datasets at large, let us narrow our focus and look at tweets related to Biden and Trump. Candidate-related tweet objects take two forms:

1. **Candidate Interaction Tweets** which are retweets of, replies to, or original tweets by @JoeBiden or @realDonaldTrump
2. **Candidate Keyword Tweets** which are tweets that contain the keyword "biden" or "trump"

The following table shows the summary statistics for both Candidate Interaction Tweets and Candidate Keyword Tweets per daily sample.

	Biden Interaction Tweets: **Replies**	Trump Interaction Tweets: **Replies**	Biden Interaction Tweets: **Retweets**	Trump Interaction Tweets: **Retweets**	**Biden Total Interaction Tweets**	**Trump Total Interaction Tweets**	**Biden Keyword Tweets**	**Trump Keyword Tweets**
days	98	98	98	98	98	98	98	98
mean	435.867	2003.73	502	3247.69	937.867	5251.43	18,909.3	51,919.3
std	246.494	1212.29	452.77	2479.78	592.016	3329	11,509.5	18,100
min	47	2	0	0	69	2	4221	18,227
25%	245.25	1222.5	247	1481.5	546	2904.25	11,040.8	40,615.2
50%	399.5	1766.5	401.5	2695	838	4707	14,009.5	49539
75%	552	2523	576.75	4680.25	1225.5	6740.75	26,072.2	61,064.2
max	1046	8300	2984	11,274	3349	19,574	62,010	122,724

Table 3.7: Summary Statistics for Candidate Specific Data

First, looking at the total Candidate Interaction Tweets, we see that Trump's tweets have more interaction. To determine if Trump's tweets have significantly

more interaction, we perform a two-sample difference in means Welch's t-test (due to unequal variance) at an $\alpha = 0.05$ significance level. The test hypotheses are:

$$H_0 : \mu_{ti} - \mu_{bi} = 0$$
$$H_a : \mu_{ti} - \mu_{bi} \neq 0$$

where μ_{ti} is the mean amount of Trump interaction tweets and μ_{bi} is the mean amount of Biden interaction tweets. The resulting 95% confidence interval is $[3636, 4991]$ with a test statistic of $t_{df:103} = 12.63$ and p-value of $p < 0.0001$. Thus, since 0 is not included in the confidence interval and $p < 0.0001$, we reject the null hypothesis. We conclude that on average, Trump's tweets had significantly more interaction than Biden's tweets during our selected date range (which we infer from the confidence interval). In a similar manner, Welch's t-test comparing the two-sample difference in means for amount of Candidate Keyword Tweets finds that on average, Trump had significantly more keyword mentions than Biden during our selected date range (95% CI $[28732, 37288]$, $t_{df:164} = 15.24$, $p < 0.0001$).

We now provide the summary statistics for the number of unique candidate tweets we pick up in our sample. That is, Table 3.8 shows data regarding tweets specifically authored by each candidate. We may gather a reference to the tweet via a reply, retweet, or original tweet.

	Biden Authored Tweets	**Trump Authored Tweets**
days	98	98
mean	14.602	31.102
std	8.53947	15.9358
min	5	2
25%	9	21
50%	12	28
75%	17.75	36.75
max	49	81

Table 3.8: Summary Statistics for Candidate Authored Tweets

A two-sample difference in means Welch's t-test comparing the mean amount of

candidate authored tweets finds that on average, Trump authored significantly more tweets than Biden during our selected date range (95% CI $[12.89, 20.11]$, $t_{df=148} = 9.03$, $p < 0.0001$).

In summary, we have provided an overview of the candidate specific data that we use in subsequent chapters. Preliminary data analysis shows traits such as the prevalence of mentions over hashtags and the relative popularity of retweets versus replies and original tweets. The candidate specific data introduces relevant summary statistics for variables that we analyze. This subsection also shows the increased interaction that users have with Trump's Twitter account in comparison to Biden's account. This difference begins to exhibit the contrast in the candidates' social presence. The next chapter describes the methods in which we use this selected data.

Chapter 4

Methods

This chapter explains the methods we use to analyze our data. We describe our methodology to provide common ground to understand the analysis of the candidates' social presence, campaign momentum, and effect of misinformation in the next chapter. First, we discuss how we preprocess the data. Then, we detail the sentiment analysis methods used. We then describe ego networks and conclude the chapter by outlining the indicators we use to determine momentum. A description of the methods used to handle the data can be found in Appendix B.1.

4.1 Data Preprocessing

The extraction and preprocessing methods are slightly different for each analysis, but do share some common ground. The main analyses we cover in Chapter 5 are on the candidates' social presence, the campaign momentum, and the effect of misinformation. In each of these analyses, we extract the text, the type of tweet (retweet vs. reply vs. original tweet), the date we collect the tweet, if the tweet is a Candidate Interaction Tweet (see Section 3.2.2), or if the tweet is a Candidate Keyword Tweet (see Section 3.2.2).

In the social presence section, we analyze the Candidate Interaction Tweets. To extract the Candidate Interaction Tweets and related data, we first determine if the user is Biden or Trump, or if the tweet is a reply or retweet of a Biden or

Trump tweet. We then extract the following:

Description	Data Type
Tweet ID of the original Biden or Trump tweet	`int`
Candidate that the interaction is with	`string`, "Biden" or "Trump"
Text of the tweet	`string`
Number of in-sample retweets of the tweet	`int`
Number of in-sample replies to the tweet	`int`
Total number of in-sample references (original tweet, retweet of, reply to)	`int`
Overall number of favorites	`int`
Date we first see a reference to the tweet	`datetime`
Date we last see a reference to the tweet	`datetime`
Whether or not the opponent is mentioned	`bool`

Table 4.1: Candidate Interaction Tweets Data Dictionary

If the reference to the candidate tweet is a reply, the thinned object does not include the original tweet's text, so the text of the tweet is only available if the tweet is original or a retweet. The overall number of favorites updates with each subsequent retweet or reply. An original tweet will have zero favorites since it has just been authored. A retweet or reply captures the tweet at a specific moment in time. Thus, the favorites update with each subsequent snapshot. To determine whether or not the opponent is mentioned, we see if the lowercased text contains the opponent's last name. After the data is extracted, we comb back through the candidate tweets and update any subtler mentions we may have missed (e.g. "Joe Hiden'" from Trump and "Mr. President" from Biden). These are the Candidate Interaction Tweets that we use throughout the analysis.

The last part of the social presence section broadens the scope to examine all tweets that contain specific hashtags. Specifically, we extract the hashtags from any tweet that has one or more of the following hashtags: #biden, #trump, #bidenharris2020, or #trump2020. We also note the connections between hashtags within a tweet, i.e., if a tweet contains #biden, #covid19, and #vote, we would keep track of three connections: biden-covid19, biden-vote, covid19-vote.

We explain the use of these edges in Section 4.3.

The effect of misinformation analysis uses the same Candidate Interaction Tweets. The campaign momentum analysis instead uses the Candidate Keyword Tweets surrounding specific dates. We extract any text that contains any variation of "biden" or "trump." Overall, each analysis focuses on the texts from a subset of tweets. We now detail the preprocessing of these texts.

The main similarity in the preprocessing is the normalization of a tweet's text. We preprocess the text to improve our sentiment analysis methods as discussed in Section 2.2. The two main text modifications are converting emojis to descriptions and converting links to a single token. In order to grasp meaning from emojis, we convert the emoji symbols into their unicode description. We use the `emoji` library in Python. For example, is converted to ":beaming_face_with_smiling_eyes:" from which we can then extract sentiment (e.g., "beaming" and "smiling" convey positive sentiment). Two of the methods detailed in Section 4.2 require the removal of underscores in order to understand the words separately. Thus, we further remove underscores from emoji descriptions.

We convert any link to the token "URL." This single token indicates that a link is present in the tweet while removing the noise associated with links—which is helpful since they are frequently a jumble of unmeaningful letters. In the same vein, Section 2.2 of the literature review also recommends replacing any user mention with a single token. We choose to leave the user mentions intact since we have a targeted, political sample based on user handles where we hypothesize that the mention of some handles may inform the sentiment of the tweet. It is an open question whether sentiment analysis accuracy is significantly influenced by the use of a single token for user mentions.

We use this text preprocessing method in each of the sentiment analyses. We detail our sentiment analysis methods in the following section.

4.2 Sentiment Analysis

In our analyses, we use sentiment classifications to gain insight into the engagement styles and campaign styles of the two candidates. First, we look at the sentiment of the candidates' tweets overall which helps us grasp the ways in which each candidate engages with their audience. The candidates' tweet sentiments also help us infer the tone in which they talk about their opponent. We investigate this more in Section 5.1.3. Additionally, we examine the sentiment of the Candidate Keyword Tweets to observe campaign momentum surrounding specific events. The following subsections detail the three sentiment classifiers that we use to complete these analyses.

4.2.1 SentiStrength and AFINN

The first methods we describe are dictionary-based classifiers. SentiStrength and AFINN are both widely used, dictionary-based, pre-trained models that extract sentiment from English text. As described in Section 2.2, SentiStrength classifies the sentiment and strength of the sentiment based on a bounded scale. The tool consists of two different classifiers. The *scale* model classifies text with a singular sentiment score on the [-4, 4] scale where [-4, -1] indicates negative sentiments, 0 is neutral, and [1, 4] labels positive sentiments. The *dual* model extracts the strength of both the positive and negative sentiments present in the text. For every input text, this dual classifier returns two scores: one positive and one negative. A positive score is between [2, 5] and a negative score is between [-5, -2]. A positive or negative score of 1 is synonymous to no positive or negative attributes, respectively. Thus, a score of (-1, 1) is a neutral score.

The SentiStrength library requires a commercial Java file and a directory of text files that we acquired for academic use from The University of Wolverhampton's Professor Thelwall (who is the main researcher behind SentiStrength). We set the paths to these directories using the `sentistrength` Python wrapper. Once set, the SentiStrength tool is usable via the `getSentiment` method. The dual classifier

is the most interesting application of SentiStrength in our analyses. It allows us to examine the Candidate Interaction Tweets to see if the candidates publish tweets with mixed sentiment.

For the other social presence and momentum analyses, we generally use the AFINN sentiment classifier. The AFINN library is a wordlist-based approach for sentiment analysis. Unlike SentiStrength, AFINN does not bound the sentiment score. Instead, the model bounds each word within the text between [-5, 5]. The individual word scores are then aggregated to produce the overall sentiment score. To use the classifier, we import the Python `afinn` library and use the `score` method. We use this classifier to investigate campaign momentum via a Sentiment Indicator which we further describe in Section 4.4. AFINN is less computationally intensive than our other methods so we use it when we have a larger subset of tweets to classify (i.e., the Candidate Keyword Tweets).

Since SentiStrength and AFINN are both dictionary-based models, no model training is needed. Preprocessing the text and initializing the model objects is the only necessary setup. As mentioned in Section 4.1, neither of these models understand a string of words that contain an underscore delimiter. Therefore, we need to further preprocess the emoji tokens to remove the underscores. For example, 😄 is first converted to ":beaming_face_with_smiling_eyes:" and then converted to ":beaming face with smiling eyes:" which SentiStrength and AFINN can both interpret.

We compare the accuracies of these models in Section 4.2.3. Before the model comparison, we present our third sentiment analysis model.

4.2.2 BERT

As discussed in Section 2.2, Google's Bidirectional Encoder Representations and Transformers (BERT) model is the current state-of-the-art language model. BERT is a machine learning model designed to understand short phrases. Unlike dictionary-based methods, BERT attempts to understand the context of a phrase. This model

can perform a variety of tasks, but we construct a BERT-based model to classify the sentiment of tweets. The class labels are -1 for negative, 0 for neutral, and 1 for positive. We use a SemEval (Semantic Evaluation) dataset constructed of tweets from 2017 that are classified with equivalent class labels. The dataset has 49,570 classified tweets. We observe class imbalance (see Figure 4.1), but we do not fix the imbalance since we would lose a decent portion of the data.

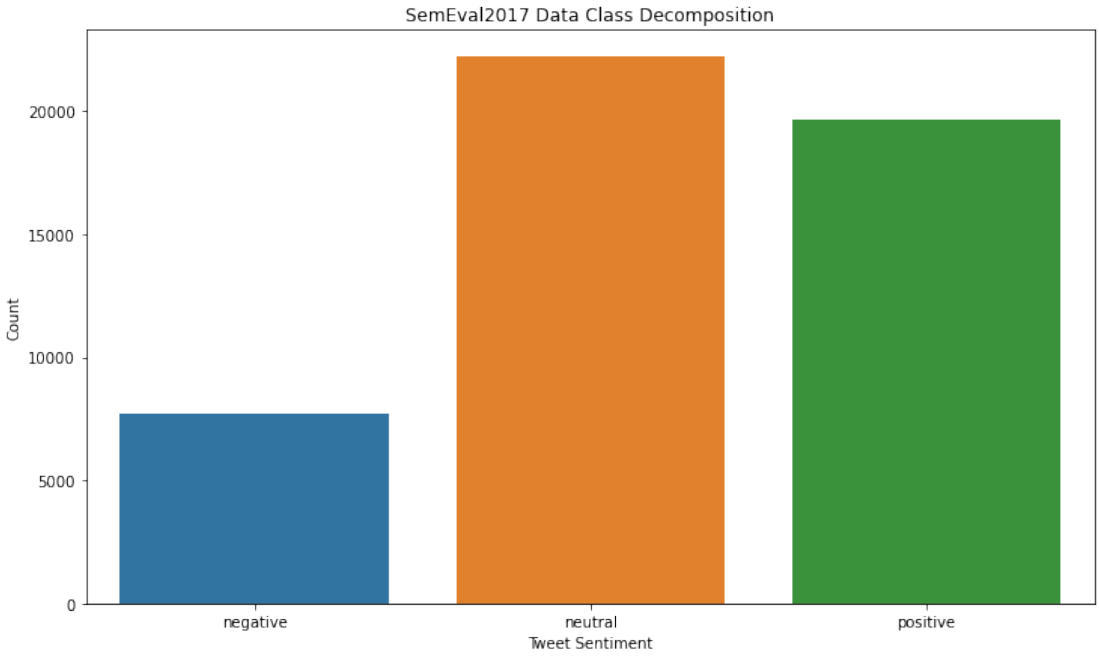


Figure 4.1 SemEval Class Decomposition

Before splitting the data, we preprocess the tweets. The SemEval dataset does not include emojis, but our actual data does, so we add them into a handful of tweets based on emoji sentiment previously extracted from tweets [13]. The emoji sentiment data provides a sentiment decomposition for 751 emoticons. It also provides the average location where each emoji appears in a tweet. Based on this breakdown, we add the relative frequencies into our tweets. For example, given that emoji A is found in 6,036 positive tweets, 109 neutral tweets, and 121 negative tweets, we add emoji A into a corresponding amount of tweets in the SemEval data. We add the emoji at its average location in a tweet. We then preprocess the tweets according to Section 4.1.

After preprocessing, we split the data into training, validation, and test sets (keep-

ing the class imbalance consistent throughout the split). The training set is 66% of the full data set with 33,210 labeled tweets. The validation and test sets are each 16.5% of the full data set such that they each have 8,180 labeled tweets. We then tokenize the data so that the input is in the correct form for the BERT model. We load the "bert-base-uncased" pre-trained tokenizer from the transformers library. For each input text, the tokenizer adds a unique start and end token, introduces padding so that each text is of equal length, and creates an attention mask so that the model understands where the true data is located (an array with 1s indicating real data and 0s indicating padding). In order to choose a max sequence length, we first tokenize each text in the training set with BERT's absolute max length of 512 tokens. Then, we plot the distribution of token lengths and select a max length of 100 tokens since the vast majority of the tokenized tweets are less than 100 tokens. We set up a data loader to tokenize and batch our data (with a batch size of 32).

Next, we construct the sentiment classifier. We load the pre-trained "bert-base-uncased" model from the transformers library. The pre-trained BERT model becomes the initial "layer" of our sentiment classifier. We then add a dropout layer and a fully-connected linear output layer. We train our model using the AdamW optimizer. The hyperparameter values are selected according to the suggestions outlined in the original BERT paper. We choose a learning rate of 2e-5, a batch size of 32, and we train for 3 epochs. After 3 epochs, the training accuracy and validation accuracy begin to diverge (see Appendix, Figure B.1).

Next, we evaluate and compare the three sentiment models.

4.2.3 Model Comparison

We investigate the accuracy of each model based on the testing set from the SemEval data. We present the confusion matrix for each model followed by the classification report. The main statistics of interest are the precision and recall scores. For each given label, the precision score is the fraction of correctly classi-

fied tweets over the total number of tweets predicted to be that label. This score corresponds with the columns in the confusion matrix (i.e. $\frac{\text{true label}}{\sum \text{column}}$). For each given label, the recall score is the fraction of correctly classified tweets over the total number of tweets truly in that category. This score corresponds with the rows in the confusion matrix (i.e. $\frac{\text{true label}}{\sum \text{row}}$). Starting with BERT, the confusion matrix is:

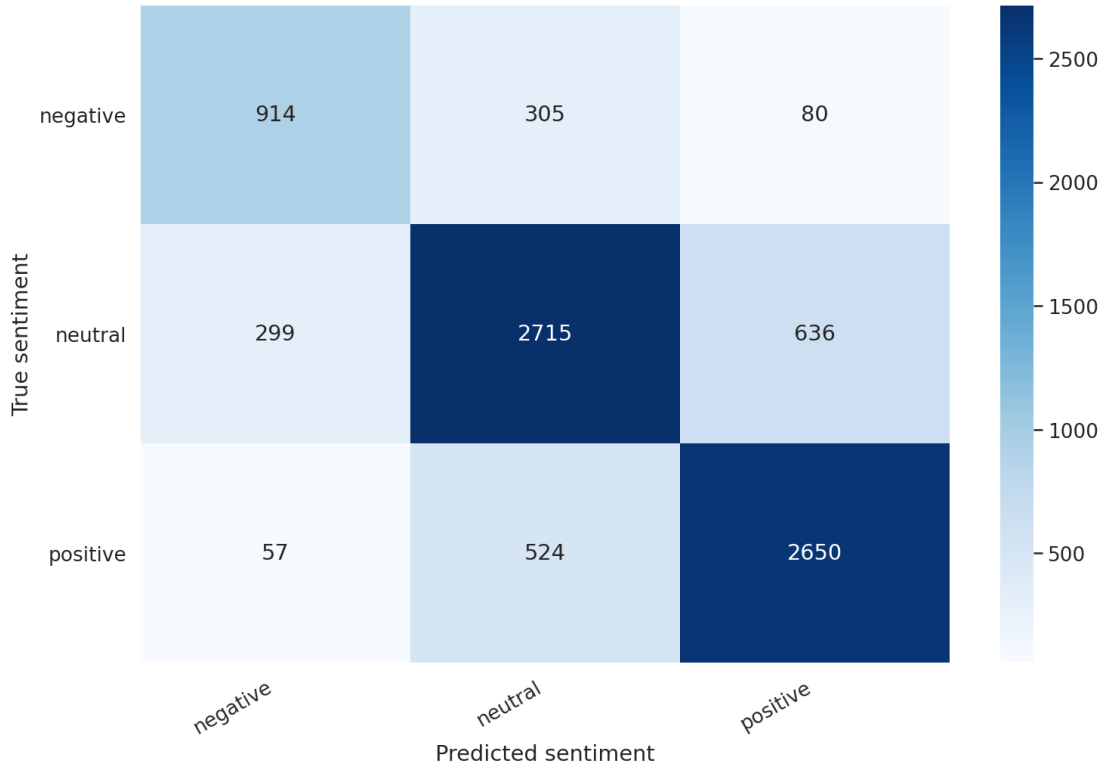


Figure 4.2 BERT Confusion Matrix

We see that the BERT-based sentiment classifier performs relatively well. The model seems to face the most difficulty with negative tweets; it confuses negative tweets as neutral more frequently than positive, and it confuses positive tweets as neutral more frequently than negative. It confuses neutral tweets as positive more frequently than negative. The model's classification report is shown in Figure 4.3.

	precision	recall	f1-score	support
negative	0.72	0.70	0.71	1299
neutral	0.77	0.74	0.75	3650
positive	0.79	0.82	0.80	3231
accuracy			0.77	8180
macro avg	0.76	0.76	0.76	8180
weighted avg	0.77	0.77	0.77	8180

Figure 4.3 BERT Classification Report

From the classification report, we confirm that the BERT model does perform relatively well since the precision and recall for each category is equal to or above 0.70. The model finds negative tweets are the most difficult to classify and positive tweets the easiest to classify.

Next, we look at the confusion matrix and classification report for the SentiStrength model.

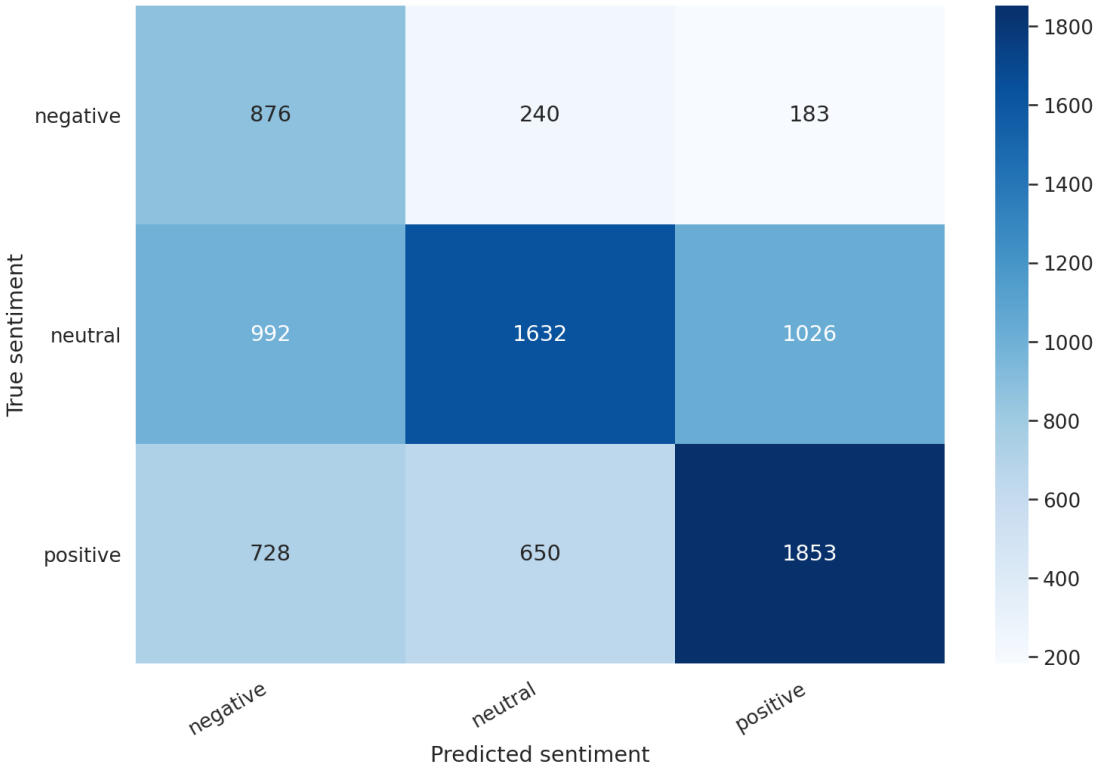


Figure 4.4 SentiStrength Confusion Matrix

	precision	recall	f1-score	support
negative	0.34	0.67	0.45	1299
neutral	0.65	0.45	0.53	3650
positive	0.61	0.57	0.59	3231
accuracy			0.53	8180
macro avg	0.53	0.56	0.52	8180
weighted avg	0.58	0.53	0.54	8180

Figure 4.5 SentiStrength Classification Report

We observe more classification problems relative to the BERT model. SentiStrength has decent recall for negative tweets, but it does not have great precision. We can see this in the confusion matrix because many of the predicted negative labels are not truly negative. In fact, out of the predicted negative labels, a greater majority are neutral than negative. This precision score shows that SentiStrength is not better than random for negative tweets. SentiStrength's precision for neutral tweets is decent at 0.65, but its recall is lacking at 0.45 meaning that over 50% of the truly neutral tweets were labeled incorrectly. SentiStrength is decent at classifying positive tweets, but it is still not amazing. Overall, an accuracy of 0.53 pales in comparison to the 0.77 overall BERT model accuracy.

Lastly, we examine the confusion matrix and classification report for the AFINN model.

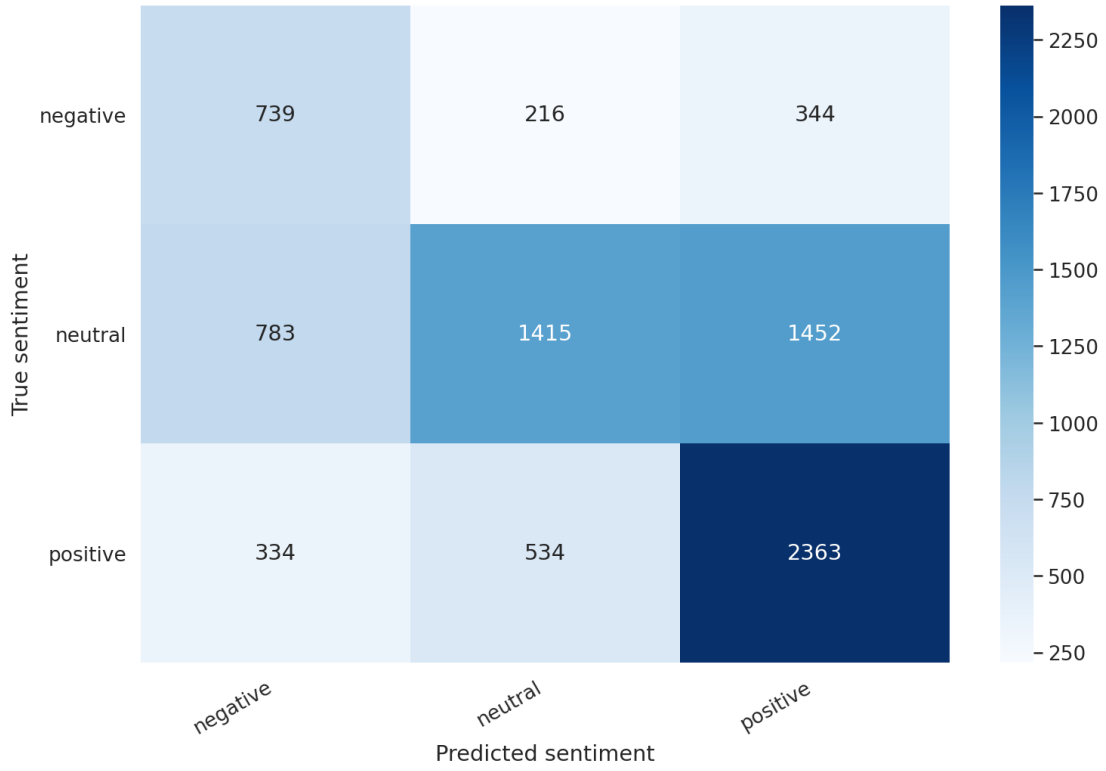


Figure 4.6 AFINN Confusion Matrix

	precision	recall	f1-score	support
negative	0.40	0.57	0.47	1299
neutral	0.65	0.39	0.49	3650
positive	0.57	0.73	0.64	3231
accuracy			0.55	8180
macro avg	0.54	0.56	0.53	8180
weighted avg	0.58	0.55	0.54	8180

Figure 4.7 AFINN Classification Report

Like SentiStrength, we see that the AFINN classifier struggles. Each of the precision scores are better than random, but not by much for the negative class. The recall is decent for negative and positive tweets, but AFINN struggles to correctly classify all of the truly neutral tweets. Within the truly neutral tweets, AFINN classifies them as positive more often than neutral. Again, we see that the BERT model outshines this dictionary-based model.

In comparison, the BERT model is by far the best of the three classifiers. This makes sense since the BERT model attempts to understand context while the dictionary-based methods only judge sentiment via isolated word choice. BERT

outperforms SentiStrength and AFINN in terms of precision and recall for all labels. Additionally, when BERT misclassifies a tweet, it more frequently chooses the label that is one away (i.e., if a truly negative tweet is misclassified, BERT more frequently selects neutral than positive). This comparison shows that we should trust our BERT classifications more than the AFINN and SentiStrength classifications. The model evaluations also demonstrate that we should take the dictionary-based sentiment classifications with a grain of salt since they only provide mediocre predictions. In our analyses, we try to use BERT when possible. However, we do use AFINN and SentiStrength for certain tasks especially since BERT is computationally expensive when processing a large amount of tweets.

4.3 Ego Networks

We use ego networks at the end of the social presence analysis to understand how Twitter users discuss the two candidates. An ego network is a graph based on a central "ego" node. All of the other nodes are directly connected to the ego. We build ego networks out of hashtags to visualize the sentiment and associated topics surrounding the candidates. Each node in the network is a hashtag. We select two neutral candidate ego nodes: #biden and #trump. We also investigate two supporting ego nodes: #bidenharris2020 and #trump2020. Nodes and edges are added to the graph when a tweet contains one of the candidate ego nodes. All hashtags present in the tweet are then connected by edges. For example, if a tweet contains #biden, #covid19, and #vote, we add the three hashtags as nodes and three undirected edges: (biden, covid19), (biden, vote), (covid19, vote). The sizes of the nodes are relative to the frequencies at which they appear.

After collecting all the nodes and connections from August 1st to November 7th, we select the top 95% of nodes based on their size. This removes unpopular and niche hashtags. We logarithmically transform the node sizes since the distribution of node sizes is right-skewed (see Appendix, Figures B.2 and B.3). This logarithmic transformation does not entirely remove the right-skew, but it allows us to visu-

ally compare relative sizes within the network. We then manually categorize the hashtags into five categories: supporting, against, important, general, and other. If a hashtag is categorized as "supporting" or "against," then the hashtag can have more than one category (i.e. supporting Biden, against Trump). However, if a hashtag is categorized as "important," "general," or "other," it cannot have another associated category. To remove any association bias during categorization, we hide the central ego nodes connected to the hashtags. Once categorized, we use Gephi to create the corresponding ego networks shown in the analyses.

4.4 Momentum Strategies

The next methods detail momentum strategies used to investigate campaign momentum and the effect of misinformation. These momentum indicators help us observe relative campaign momentum surrounding specific events. First, we discuss the Sentiment Indicator. Then, we describe the Curvature Indicator.

The Sentiment Indicator (SI) compares the candidates' tweets via sentiment count using the AFINN model. The total sentiment count is the sum of all sentiment scores within a time period. Thus, to judge the momentum surrounding a specific event, we aggregate sentiment scores for each of the Candidate Keyword Tweets that occur during the date range of the event. A higher sentiment count indicates that a campaign has greater positive momentum.

The Curvature Indicator (CI) compares the momentum of different candidates by measuring the way in which a chosen tweet attribute changes. The tweet attribute we monitor for the CI is the daily number of tweets published by the candidates. We use the CI to determine if a misinformation story has a significant association with a change in momentum. To do so, we approximate the second derivative at each day and fit a series of nested linear models to test the effect of a misinformation story. Specifically, the numerically approximated second derivative is our response variable (y). The predictors are date (X_d), a binary indicator of

pre or post misinformation story (X_m), and an interaction term between the date and the binary indicator ($X_d X_m$). The reduced model is:

$$y = \beta_0 + \beta_d X_d + \beta_m X_m + \epsilon_i$$

The second model is:

$$y = \beta_0 + \beta_d X_d + \beta_m X_m + \beta_{d:m} X_d X_m + \epsilon_i$$

To test whether there is a significant change in the second derivative (and thus, a significant change in momentum), we perform a partial F test. The partial F test compares these two nested models via an ANOVA. If the p-value is less than $\alpha = 0.05$, the test is significant, and we can conclude that the interaction term between day and the pre-post indicator adds significant explanatory power to the model. This conclusion also signifies that the misinformation event has some significant association with a change in momentum.

In Section 2.3, we also discussed the use of Rate of Change (ROC) and Relative Strength Indicators (RSI). The CI essentially investigates the same changes as these indicators since they all measure the change in the momentum's direction. Thus, we only use the CI in our analysis.

In the next chapter, we apply these methodologies and present our findings about the behaviors of the presidential candidates on Twitter.

Chapter 5

Results and Discussion

In this chapter, we present and discuss the results from our analysis. First, we look at the candidates' social presence by analyzing their respective Twitter accounts and Candidate Keyword Tweets. We investigate their social presence by looking at temporal characteristics, engagement style, campaign style, and how they appear in discussion. Temporal characteristics show how the candidates use Twitter and how others interact with their tweets from August 1, 2020 to November 7, 2020. Engagement style aims to uncover the sentiment, emotions, and common themes used by the candidates to engage their audience. Campaign style examines how the candidates refer to one another and the effect an opposing mention has on user interaction. These three subsections are focused on the candidates' Twitter accounts, giving a directed focus to our analysis. To widen the scope, we also look at how the candidates appear in discussion by other users.

The next section looks at the efficacy of campaign momentum indicators surrounding events meaningful to the election. Specifically, we observe the momentum at the Democratic National Convention (DNC), the Republican National Convention (RNC), the first presidential debate, Trump's COVID-19 diagnosis announcement, the vice presidential debate, the second presidential debate, and election day.

The last section uses the momentum indicators to investigate the effect of misinformation. We select three misinformation events to examine. Then, we look at the momentum changes indicated by the Sentiment Indicator (SI) and Curvature

Indicator (CI). We conclude the chapter by discussing implications of the misinformation results.

5.1 Social Presence

In order to examine the candidates' social presence, we principally consider the Candidate Interaction Tweets (see Section 3.2.2). Narrowing the scope to these tweets removes a layer of uncertainty in trying to figure out if a tweet supports or opposes the mentioned candidate. We know that a tweet from Biden's account supports Biden, and a tweet from Trump's account supports Trump. Exploring the candidates' tweets also allows us to gain a better sense of campaign messaging, style, and momentum since we are examining the primary source. However, reducing the scope means that we disregard a vast amount of data, losing a range of opinions by Twitter users. Section 5.1.4 attempts to counteract this issue by investigating how the rest of Twitter discusses the candidates.

5.1.1 Temporal Characteristics

This subsection analyzes the temporal trends associated with the candidates' tweeting behaviors. We first examine the daily number of tweets authored by each account. Figure 5.1 shows the daily number of tweets in our sample from Biden and from Trump as well as the corresponding seven day simple moving averages (SMA). Several events are highlighted.

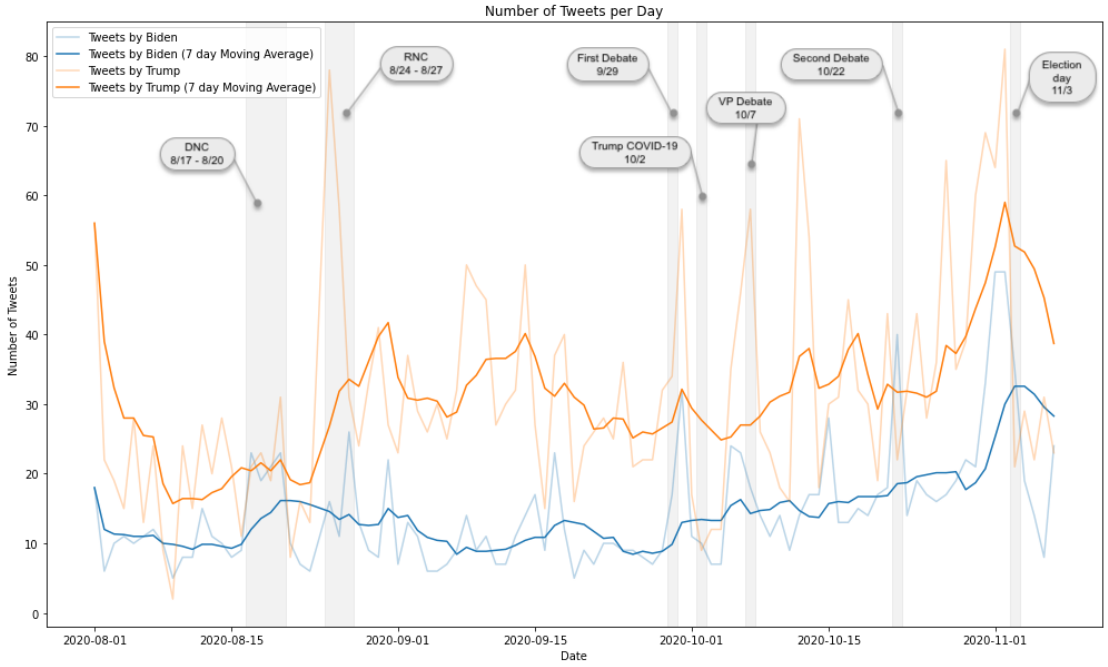


Figure 5.1 Number of Tweets per Day by Candidate

First, we note an initial spike at the beginning of our date range (August 1st). This first data point overcounts the true daily tweets because it accumulates tweets authored prior to August 1st due to our data summarization method. The data after this first point is valid.

Examining the daily tweets trend via the SMAs, we notice that Trump typically tweets more than Biden in this period of time since his trend is always greater than Biden's trend. Trump's trend line is also more sporadic than Biden's—we see larger spikes in the granular data that makes his trend line jumpier. In mid-August, Trump averages 17 to 22 tweets daily, his lowest average in our date range. From Late August to mid-October, Trump averages between 25 and 40 tweets a day. On the other hand, Biden's trend line is fairly stable from August 1st to October 1st, where he averages about 10 to 15 tweets picked up in our sample (fewer than Trump's lowest average). While these numbers are estimates from our roughly representative sample of Twitter, they closely align with the true number of daily tweets by each candidate which we compared for a handful of random days. Once October hits, both Trump and Biden's accounts see a steady increase in the amount of daily tweets up through the election. These trends show

that Trump's tweeting behavior is more active than Biden's. We also note that both candidates increase their tweeting behavior in the month prior to the election.

Looking at the granular data, we can interpret certain spikes and trends based on surrounding events. During the DNC (the earliest event highlighted), we see an increase in Biden's tweets. Biden averages approximately 20 tweets a day through the duration of this event in support of him and his party. This is one of the few locations where we see a roughly equivalent number of daily tweets between Trump and Biden. The RNC (the next event) shows a massive spike in Trump's daily tweets. On the second day of the RNC (recall that we are missing data for August 24th), Trump tweets approximately 78 times. Trump's daily levels dip after the RNC, but remain at an increased average, suggesting a more active tweeting behavior after the official kickoff to the election season.

The next event that we examine is the first presidential debate on September 29th. While we do not observe a staggering increase in tweets during the actual day of the debate, we do see a spike in both Trump and Biden's daily tweets on the following day. This dual spike highlights increased tweeting behavior following an event that had major press. In context, this increase likely suggests that the candidates wanted to capitalize on the extra attention media brought to their campaigns. They can reach a larger audience to promote their plans and fix any blunders from the prior night's debate.

Trump tweeted about his COVID-19 diagnosis on October 2nd. He then greatly reduced his tweeting in the days immediately following. We see this highlighted in Figure 5.1 as his daily tweet count drops. Biden's tweet count also drops. We see both of their tweet counts rise a few days later as the vice presidential debate approaches. Trump shows a peak in daily tweets during the day of the debate while Biden exhibits a peak the day before.

The second to last event is the second presidential debate (originally scheduled as the third). Trump's daily amount of tweets peaks the day before and the day after

the debate. Biden's tweets peak during the day of the debate. This is another day where Biden has more tweets than Trump. Unlike the other events where we generally see the trend in tweeting levels decrease, we notice an increase in both of the candidates' tweeting behaviors following the debate up through the election. The days immediately prior to the election show a very large increase in daily tweets. Both Trump and Biden hit their maximum amount of tweets in a given day (Trump at 80, Biden at 48). We then see a fairly drastic decrease in both of their tweets on and after election day.

Overall, these granular trends show the candidates' tendency to increase their tweeting behaviors in the days surrounding an important event.

Next, we examine user interaction with candidates' tweets. We observe user interaction via replies, retweets, and favorites. Measured replies and retweets are not equivalent to the true number of replies and retweets. Instead, they are based on the replies and retweets we pick up in our sample. Since our sample is roughly representative of a true day on Twitter (see Section 3.1.2), we assume that the relative levels of replies and retweets within our sample is sufficiently representative of the truth. Favorites are inherently different because they do not trigger a new tweet object. Thus, we collect roughly the correct amount of favorites per tweet. For analysis, we look at the daily proportion of candidate tweets that are replies and retweets. These plots should roughly be inverses since replies and retweets are the main form of tweets collected (see Figure 3.10). Then, we look at the average number of favorites by day.

The following figures show the SMA of average proportion of replies to candidate tweets by day (Figure 5.2) and the SMA of average proportion of retweets to candidate tweets by day (Figure 5.3). We look at the seven day SMA because the daily data is quite erratic.

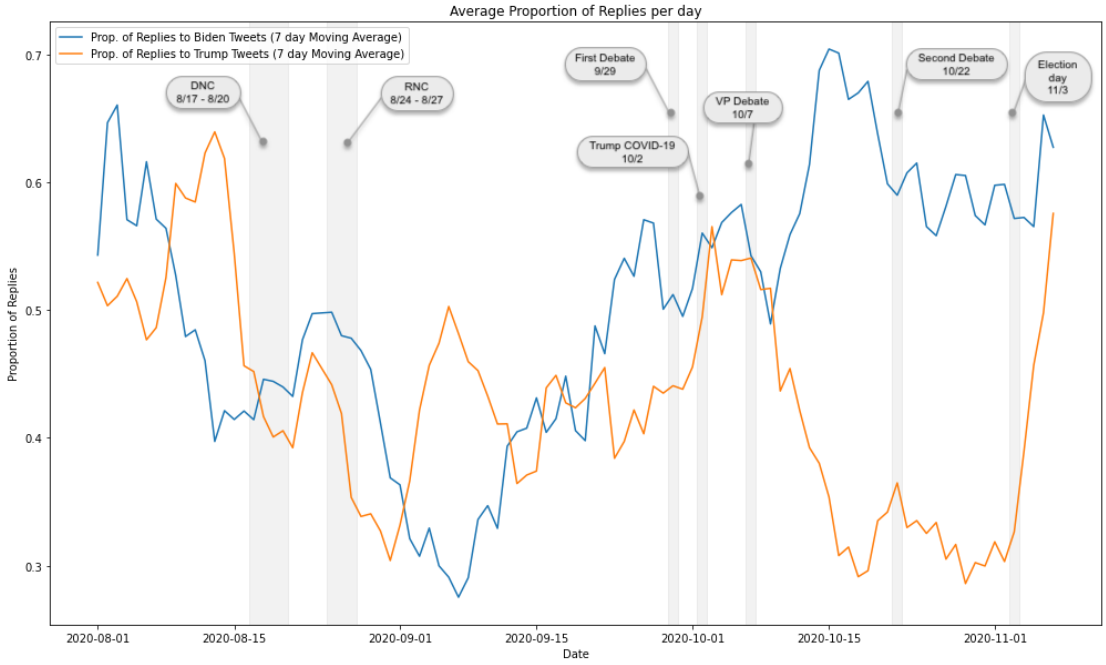


Figure 5.2 Average Proportion of Replies per Day by Candidate

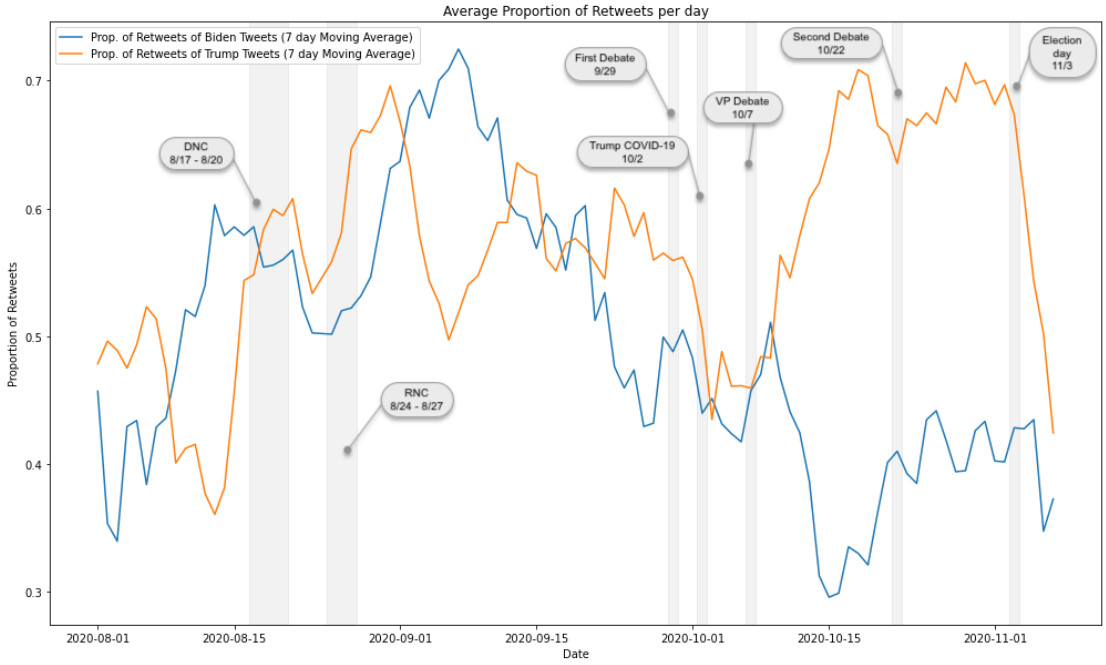


Figure 5.3 Average Proportion of Retweets per Day by Candidate

As expected, figures 5.2 and 5.3 are essentially inverses of one another. Within each figure, it is interesting to note that the Trump and Biden trends are also frequently inverses. When one candidate has a high amount of replies/retweets, the opponent sees a low amount of replies/retweets. The only date range that does not clearly show this relationship is roughly August 23rd to September 1st

(near the RNC). The combination of these two relationships suggests that when one candidate has a greater proportion of replies, their opponent has a greater proportion of retweets (and vice versa).

A possible interpretation of this interesting relationship could be that when a divisive event occurs, both candidates tweet about it, causing one side to interact in support and the opposing side to interact in opposition. This interpretation assumes that a retweet and a reply carry statements of support and disagreement, respectively. We do not further investigate this correlation in this thesis.

A second area of interest occurs from October 8th to November 3rd. This ranges from one day after the vice presidential debate to election day. During this chunk of time, we see Biden receive more replies and Trump receiving more retweets. This period of stark interaction difference correlates with the candidates' increased tweeting behavior prior to the election. Overall, when the candidates increase their tweeting, we see more user interaction via reply for Biden and via retweet for Trump.

5.1.2 Engagement Style

We now examine the content of the candidates' tweets in order to better understand their engagement style. We analyze the sentiment, emotions, and common themes that the candidates employ to capture their audience's attention.

As described in the Methods chapter (Section 4.2), we use three different sentiment classifiers to analyze the sentiment polarity of the candidates' tweets. Section 4.2 also shows that the BERT model performs best on test data (especially for negative tweets). We cannot determine how well the models perform on our candidate data since it is unlabeled, but we can compare the polarity breakdowns from each classifier. Figure 5.4 shows this decomposition.

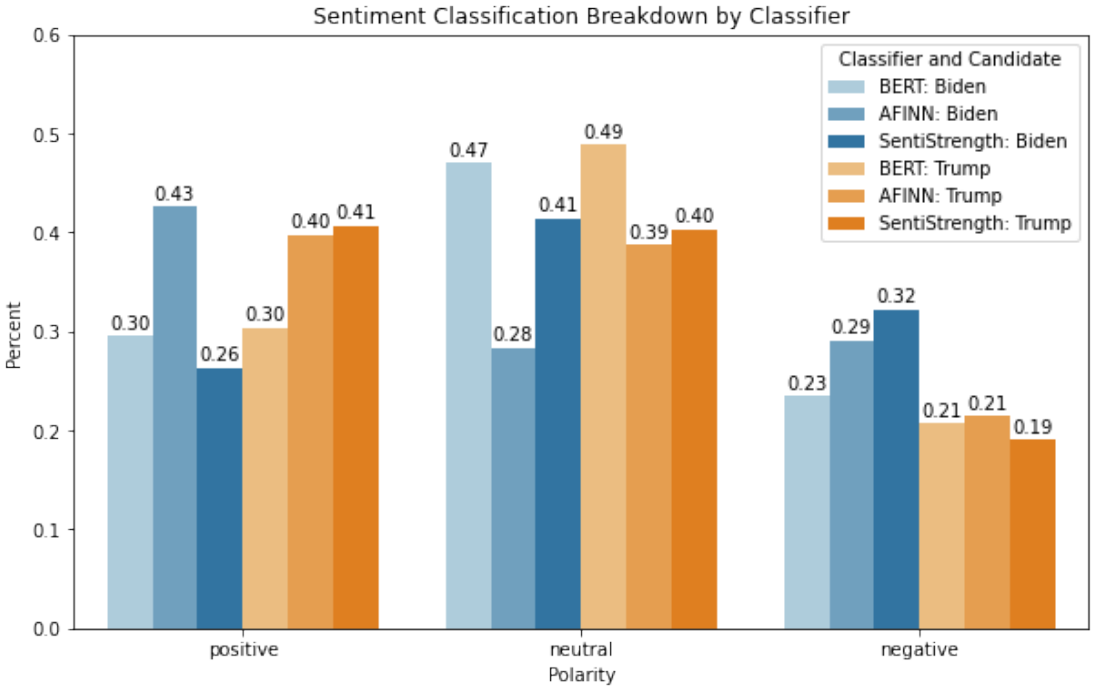


Figure 5.4 Sentiment Classification Breakdown by Classifier

Comparing the BERT, AFINN, and SentiStrength classifiers, we see that BERT seems the most conservative in its predictions since it is predicting neutral almost 50% of the time for both Biden and Trump. On Trump's tweets, AFINN and SentiStrength are fairly similar in their classification breakdown. This is not the case for Biden's tweets. Instead, we see AFINN classify 43% of Biden's tweets as positive while SentiStrength only classifies 26% as positive.

For Biden's tweets, BERT, AFINN, and SentiStrength only match on 13.17% (116 tweets) of positive tweets, 14.19% (125) of neutral tweets, and 13.62% (120) of negative tweets. This leaves 59.02% (520) of Biden's tweets with mixed classifications. For Trump's tweets, the three classifiers agree on 23.73% (445) of positive tweets, 28.75% (539) of neutral tweets, and 9.97% (187) of negative tweets—thus, leaving 37.55% (704) of Trump's tweets with mixed classifications. This discrepancy between models highlights the difficulty of sentiment analysis on short texts. We also notice that Trump's tweets are slightly easier to classify compared to Biden's tweets since the classifiers agree more frequently for Trump versus Biden.

The large percent of tweets with mixed classifications inspires us to use the Sen-

tiStrength dual algorithm which can predict if a text shows mixed sentiment in addition to the traditional positive, neutral, and negative polarity classification. This breakdown is shown as follows:

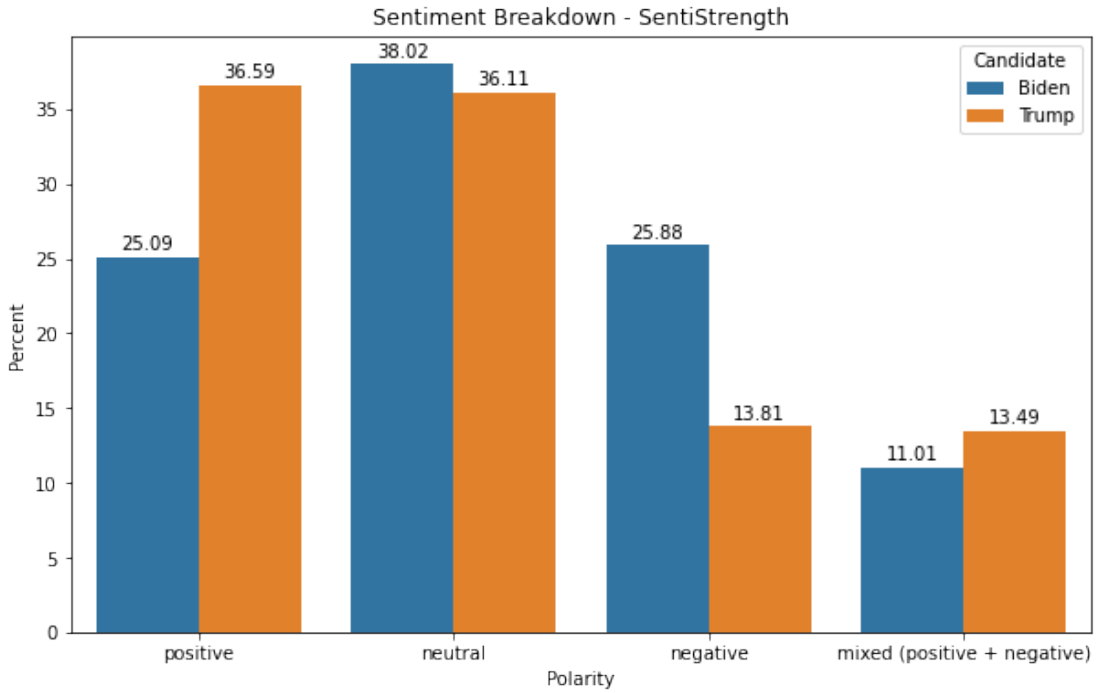


Figure 5.5 SentiStrength Dual Sentiment Breakdown

We see SentiStrength predict 11.01% of Biden's tweets and 13.49% of Trump's tweets as displaying mixed sentiment (both positive and negative). Comparing the SentiStrength classifications in Figure 5.4 with Figure 5.5, we see that the majority of mixed sentiment tweets end up classified as negative for both Trump and Biden since the negative category is the one with the biggest difference. We also notice that via SentiStrength, Trump has more positive tweets while Biden has more negative tweets.

In order to better understand the difficulties these algorithms face in classifying sentiment polarity, we present examples. Recall that the AFINN algorithm does not have a bounded classification score range like BERT and SentiStrength. BERT classifies with a -1 for negative, a 0 for neutral, and a 1 for positive. SentiStrength classifies on a range of $[-5, 5]$ where negative scores indicate negative polarity, 0 is neutral, and positive scores signal positive polarity. Thus, we look at the most

positive and most negative tweets as classified by AFINN. The first 10 are the tweets classified as most positive (five for Biden, five for Trump). The second 10 are the tweets classified as most negative.

Candidate	Tweet	AFINN	BERT	Senti Strength
Biden	Receiving the Medal of Freedom from my friend @BarackObama was a great, great honor. I'm proud to have dedicated my... https://t.co/lbUazDNc5q	18	1	3
Biden	Folks, the people of this nation have spoken. They have delivered us a clear victory. A convincing victory. A victory for "We the People."	10	1	0
Biden	John McCain was a good man, and a good friend. He made this country better. I miss him dearly. #DemConvention https://t.co/TZ4cRY5fz4	10	1	0
Biden	Happy birthday to my brother and friend, @BarackObama! Serving alongside you for eight years was the greatest honor... https://t.co/EwRiuqBwdW	9	1	2
Biden	I want to help restore hope in this country — and I have a bold vision for how, together, we can do that. https://t.co/ZdW23fhad2	9	1	2

Table 5.1: Biden Most Positive Tweets (by AFINN)

Candidate	Tweet	AFINN	BERT	Senti Strength
Trump	If Biden Wins, CHINA WINS. If Biden Wins, the mob WINS. If Biden Wins, the rioters, anarchists, arsonists, and flag... https://t.co/irRdJa0bqi	20	-1	-1
Trump	This election is a simple choice: If Biden Wins, China Wins. When WE Win, YOU WIN, Pennsylvania WINS, and America W... https://t.co/mlDW1i3nSD	20	0	0
Trump	For half a century, Joe Biden has been outsourcing your jobs, opening your borders, and sacrificing American blood and treasure in endless foreign wars. Joe Biden is a corrupt politician— If Biden Wins, China Wins. When We Win, Florida Wins—and America Wins! https://t.co/UdqQqmxvL7	19	-1	-1
Trump	Joe Biden is a corrupt politician—the Biden family is looking more & more like a crime family. They took millions of $$ from the Chinese Communists, while Vice President Biden shipped Michigan jobs to China. If Biden wins, China Wins. When we win, Michigan wins, and AMERICA wins! https://t.co/YpnjH8838X	16	-1	-1
Trump	I am asking you to go VOTE for your favorite president, we still have work to do! Let's WIN, WIN, WIN! https://t.co/ysLmOS689V	14	1	2

Table 5.2: Trump Most Positive Tweets (by AFINN)

In these 10 tweets, we see classification discrepancy. For Biden, BERT and AFINN agree, but SentiStrength classifies two of the above tweets as neutral. Reading these tweets ourselves shows that the classification really could go in either a positive or neutral direction. For Trump's top AFINN tweets, we see greater inconsistency. BERT only agrees with one of the positive classifications. The BERT classifications more closely align with the true sentiment of the tweets as four of Trump's top five "positive" tweets appear to be negative. We notice that the repetition of "win" seems to boost the AFINN score while BERT can better interpret the context within the tweet. We also observe a stylistic difference between these top Biden and Trump tweets. Biden's tweets talk about his positive connections with popular bipartisan members (John McCain and Barack Obama). His tweets also discuss his accomplishments and goals for the country. On the other hand, Trump's top tweets are all directed at attacking Biden, spamming the word "win," and inspiring fear.

Now, we look at the 10 tweets classified as most negative.

Candidate	Tweet	AFINN	BERT	Senti Strength
Biden	It's long past time we take action to end the scourge of gun violence in America. As president, I'll ban assault weapons and high-capacity magazines, implement universal background checks, and enact other common-sense reforms to end our gun violence epidemic.	-12	-1	-3
Biden	Jill and I are keeping the French people in our prayers following the horrific terror attack in Nice — which targeted innocents in a house of worship. A Biden-Harris administration will work with our allies and partners to prevent extremist violence in all forms.	-10	-1	-2
Biden	Iran's cruel execution of Navid Afkari is a travesty. No country should arrest, torture, or execute peaceful protes... https://t.co/nCDv90Bytw	-10	-1	-3
Biden	It didn't have to be this bad — and as @MichelleObama said, "Our commander in chief, sadly, has been missing in action." It's unforgivable. https://t.co/RIWEck8d3m	-10	-1	-1
Biden	Today marks 22 years since Matthew Shepard passed away after a brutal anti-gay hate crime. Over two decades later,... https://t.co/ZcUEuSe7Vv	-10	0	-3

Table 5.3: Biden Most Negative Tweets (by AFINN)

Candidate	Tweet	AFINN	BERT	Senti Strength
Trump	Failed RINO former Governor Rick Snyder of Michigan was a disaster with respect to the Flint Water CATASTROPHE, and a very bad Governor overall. He hurt so many people with his gross incompetence. He reminds me of Sleepy Joe!	-14	-1	-1
Trump	Totally Negative China Virus Reports. Hit it early and hard. Fake News is devastated. They are very bad (and sick!) people!	-13	-1	-3
Trump	Our prayers are with the people of Vienna after yet another vile act of terrorism in Europe. These evil attacks against innocent people must stop. The U.S. stands with Austria, France, and all of Europe in the fight against terrorists, including radical Islamic terrorists.	-13	-1	-2
Trump	Cases up because we TEST, TEST, TEST. A Fake News Media Conspiracy. Many young people who heal very fast. 99.9%. Corrupt Media conspiracy at all time high. On November 4th., topic will totally change. VOTE!	-12	-1	-1
Trump	Joe Biden's plan to BAN FRACKING is an Economic DEATH SENTENCE for Pennsylvania. A vote for Biden is a vote to Ban Fracking, Outlaw Mining, explode energy costs, and totally DESTROY Pennsylvania. Get out tomorrow and VOTE #TrumpPence2020! https://t.co/niFJw558En	-11	-1	-1

Table 5.4: Trump Most Negative Tweets (by AFINN)

We only see one discrepancy in the top, negatively classified tweets. BERT classifies Biden's fifth tweet as neutral while the other algorithms classify the tweet as negative. BERT's classification better aligns with the sentiment of the tweet since the tweet is honoring someone who passed away. The words "brutal," "hate," and "crime" likely boosted the negative scores for AFINN and SentiStrength. Reading through the tweets, we see that Trump's are easier to place as negative tweets. They attack Biden, the media, and the Michigan governor. Biden's tweets instead focus on heavy topics (such as violence, terrorism, and hate crimes) which is likely the reason dictionary-based methods classify them as negative. Only one of Biden's top negative tweets attacks Trump. For both candidates, we see that a message condemning terrorism is negatively classified. This makes sense because of the word choice used, but the overall meaning makes it hard to clearly sort the tweet's polarity.

In a similar vein, we investigate the top words used by each candidate. We look at the most common words (excluding stopwords) in their tweets overall and in their positive, neutral, and negative tweets. The following tables present the 15 most frequently used words of each candidate, categorized by tweet sentiment polarity. The words are in order (i.e., the first word in the cell is the most common).

Sentiment Polarity	Common Among All Classifiers	Additional
Overall	trump, president, people, donald, vote, time, kamalaharris, american, america, nation, country, us, one, election, day	
Positive	president, people, america, kamalaharris, country, time, make, back, better, us, american, one, vote	**BERT**: nation, build **AFINN**: donald, election, help **SentiStrength**: care, like, love, promise, support, truth
Neutral	trump, president, vote, donald, people, kamalaharris, time, election, day, make, get, us, today, country	**BERT**: american **AFINN**: america, together **SentiStrength**: one
Negative	trump, president, donald, people, years, americans, time, failed, violence, crisis, america	**BERT**: american, white, pandemic **AFINN**: country, one **SentiStrength**: us

Table 5.5: Biden's Frequently Used Words

In positive tweets originating from Biden's account, we see the frequent use of his campaign slogan, "Build Back Better." The word "vote" is common overall and especially in positive and neutral tweets which conveys an important theme of Biden's campaign (and of the election in general). We see that Biden's tweets often mention Trump, no matter the polarity. Yet, we see that "trump" and "donald" are most common among Biden's negative tweets. This suggests negative messaging surrounding Trump which we further investigate in Section 5.1.3. Biden's negative tweets also commonly mention "violence," "crisis," "pandemic," and "failed." Next, we analyze Trump's most frequent words.

Sentiment Polarity	Common Among All Classifiers	Additional
Overall	great, people, biden, trump, maga, joe, vote, president, thank, news, time, get, america, would, american	
Positive	great, maga, thank, people, vote, american, america, new, state, big, trump	**BERT**: way, enjoy, see, one **AFINN**: likes, win **SentiStrength**: time, pennsylvania, foxnews
Neutral	trump, biden, president, people, joe, via, maga, election, fake, get, great, poll, news	**BERT**: vote, realdonaldtrump **AFINN**: energy, many, see, sign, today, would **SentiStrength**: big, democrats, state
Negative	biden, people, joe, would, fake, vote, news, country, never, energy, bad, election	**BERT**: time, like, want **AFINN**: get, media, president **SentiStrength**: american, corrupt

Table 5.6: Trump's Frequently Used Words

Like Biden, we see Trump's campaign slogan "maga" and the urge to "vote." Other common words from Trump's tweets overall are "great," "thank," and "news." Unlike Biden, Trump's own name is one of his top words in every category. In his positive tweets, we see the use of "great" and "maga" again, but we also see the addition of "foxnews." In his neutral tweets, we see words more closely directed at the election such as "poll," "democrats," and "state." His neutral tweets also show Trump's frequent use of "fake" which is likely related to his claims of "fake news." The use of "fake" and "news" becomes more prominent in his negative tweets along with "media" and "corrupt." We also see "biden" as the most frequent keyword in his negative tweets.

To further understand the engagement style of the candidates' tweets, we look at their emotional composition. We use the NRC lexicon which associates over 14,000 English words with eight emotions: fear, anger, disgust, surprise, sadness, anticipation, trust, and joy [27]. These eight emotions are derived from Plutchik's wheel of emotions. The results in Figure 5.6 show the breakdown of emotions displayed in the candidates' tweets where the mean emotion score represents the expected occurrence of an emotion per tweet.

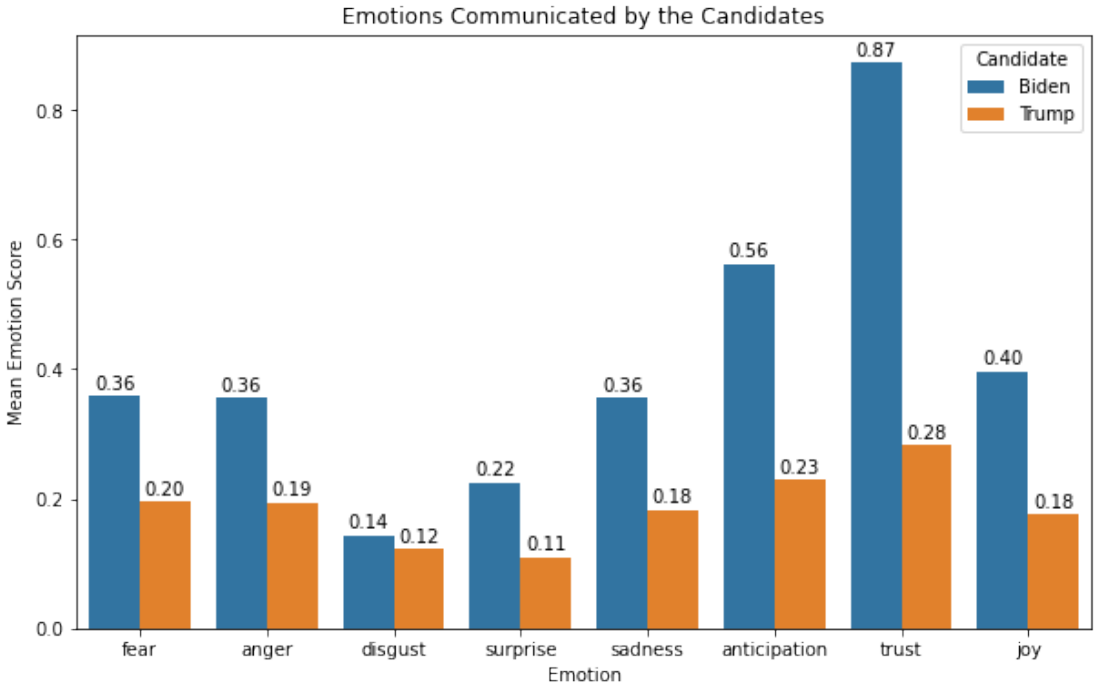


Figure 5.6 Tweet Emotional Composition by Candidate

Overall, we see that Biden's tweets are more emotionally charged. Biden communicates trust far more than any other emotion. Second to trust is anticipation, followed by joy. Trump's top emotion is also trust, but the difference in mean score between trust and the other emotions is much less pronounced. Like Biden, we see anticipation as Trump's second most communicated emotion. Unlike Biden, fear is Trump's third most communicated emotion. Joy is one of Trump's least communicated emotions. From examining the emotional breakdown, it appears that Biden more frequently communicates emotionally charged messages of trust, anticipation, and joy. Trump's communication style is less emotionally charged, but more active (as we saw in Section 5.1.1).

The analyses so far detail the candidates' engagement styles based on sentiment and emotion. We can also gauge engagement via user interactions based on tweet polarity. The following observations of user reaction use the BERT classifications for sentiment since it is the most accurate classifier on the test data. Figure 5.7 shows how the average amount of favorites per tweet changes based on tweet sentiment.

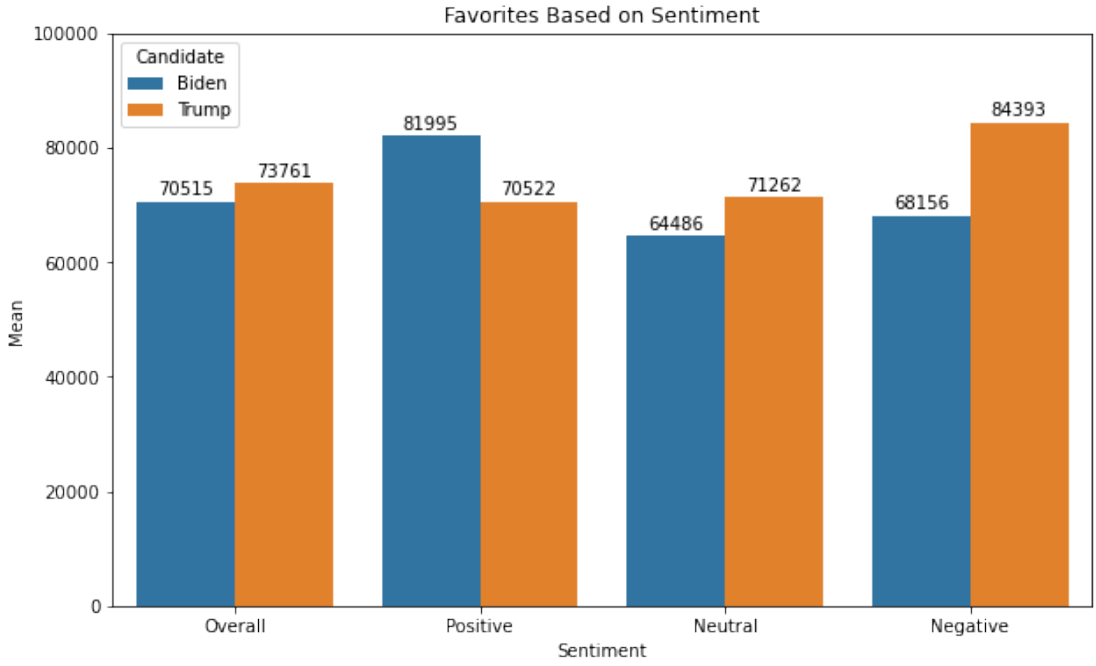


Figure 5.7 Favorites by Sentiment

The decomposition of favorites shows that Biden has increased favorites on positive tweets while Trump sees increased favorites on negative tweets. Even though we see an increased amount of favorites for Biden on positive tweets, the difference from the average number of favorites overall is not statistically significant via Welch's t test (95% CI $[-40889, 17930]$, $t_{df:325} = -0.7679$, $p = 0.443$). The two-sample difference in means Welch's t test between Trump's average overall favorites and his average negative favorites constructs a 95% CI of $[-18793, -2471]$ ($t_{df:638} = -2.56$, $p = 0.0107$). Thus, Trump's increase in average favorites on his negative tweets is statistically significant. We see similar conclusions by looking at Pearson's correlation coefficient between AFINN sentiment count and tweet favorite count. The correlation between AFINN sentiment count and Trump's favorite count is significant ($p = 0.0189$) with $r = -0.054$. This shows a weak, negative correlation between Trump's sentiment and favorite count, corroborating the earlier t test. We see $r = 0.071$ with $p = 0.036$ for the correlation between Biden's AFINN sentiment and favorite count. This suggests a significantly weak, positive correlation between Biden's sentiment and favorite count, highlighting the trend we observe in Figure 5.7. Additionally, a two-sided difference in means Welch t test comparing the average amount of favorites between Trump and Biden

is not significant, suggesting that we do not have enough evidence to claim that either Trump or Biden receives more favorites on their tweets overall (95% CI $[-7238, 13730]$, $t_{df:1169} = 0.6083$, $p = 0.543$).

Now, let's observe the breakdown for replies and retweets. Recall that the number of favorites is approximately the correct amount of favorites per tweet, but replies and retweets are based on our sample. Figure 5.8 shows the average amount of replies per tweet and Figure 5.9 shows the average amount of retweets per tweet.

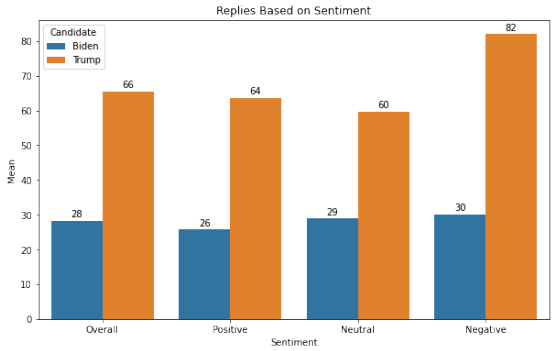


Figure 5.8 Replies by Sentiment

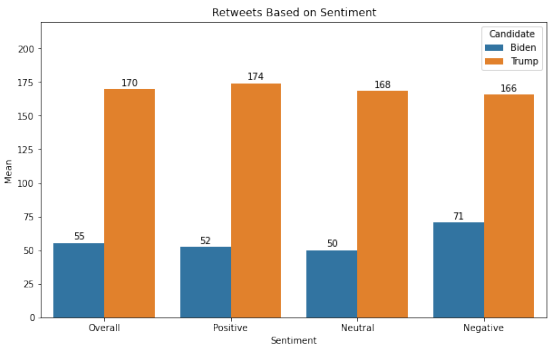


Figure 5.9 Retweets by Sentiment

Biden and Trump both receive more replies on negative tweets and fewer replies on positive tweets. Neither of these differences are significantly different from the average overall amount of replies. We see that Biden receives more retweets on negative tweets while Trump receives more retweets on positive tweets. Again, neither of these differences are statistically significant. What we do observe is that Trump significantly receives an increased amount of average replies (95% CI $[26.9, 47.7]$, $t_{df:2256} = 7.058$, $p < 0.0001$) and average retweets (95% CI $[95, 133]$, $t_{df:2256} = 11.72$, $p < 0.0001$) in a two-sample difference of means Welch's t test compared with Biden's average replies and average retweets.

Overall, we see a significant increase in favorites on Trump's tweets when they are negative, but we do not observe any other significant difference based on sentiment. We also find that Trump has significantly more replies and retweets than Biden which verifies our user interaction findings from Section 3.2. Additionally, we distinguish Trump and Biden's engagement styles via their emotional content.

Biden more frequently communicates with emotional messages. He especially conveys trust, anticipation, and joy. While Trump's tweets are less emotional, he is more active in tweeting.

5.1.3 Campaign Style

To analyze the candidates' campaign styles, we look at how they mention their opponent. We identify an opponent mention in several ways. First, we look at how many times the opponent's name appears (e.g. "Trump" or "Biden"). Since this is not the only way that the candidates refer to one another, we also count mentions of related nicknames or proper titles (e.g. "Joe Hiden'" from Trump and "Mr. President" from Biden). In total, 28.15% (248) of Biden's tweets mention Trump and 9.49% (178) of Trump's tweets mention Biden. Already, we see that Biden directly addresses his opponent more often than Trump. Let's analyze this subset of tweets to attain a better understanding of how each candidate talks about their opponent.

Figure 5.10 shows the decomposition of the tweets mentioning an opponent based on sentiment polarity.

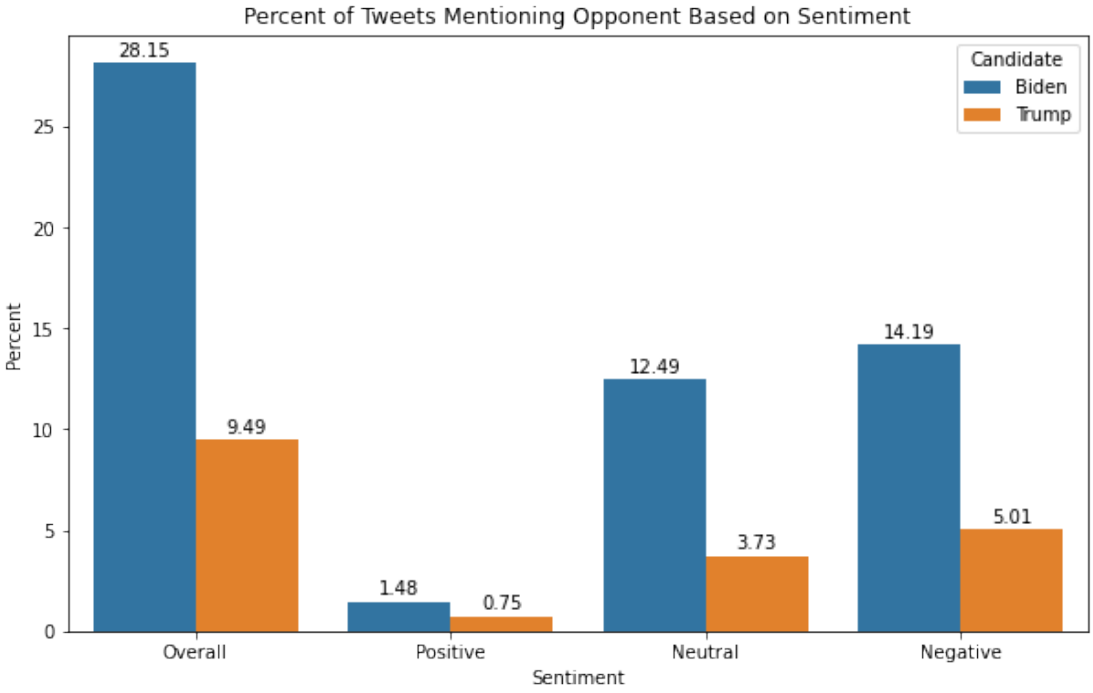


Figure 5.10 Sentiment Decomposition When Mentioning Opponent

Immediately, we notice that Biden has a greater percent of his tweets mentioning Trump over all polarities. We also observe that Biden and Trump rarely mention their opponent in a positive light. To understand the context behind these tweets, we look at their emotional composition in Figure 5.11.

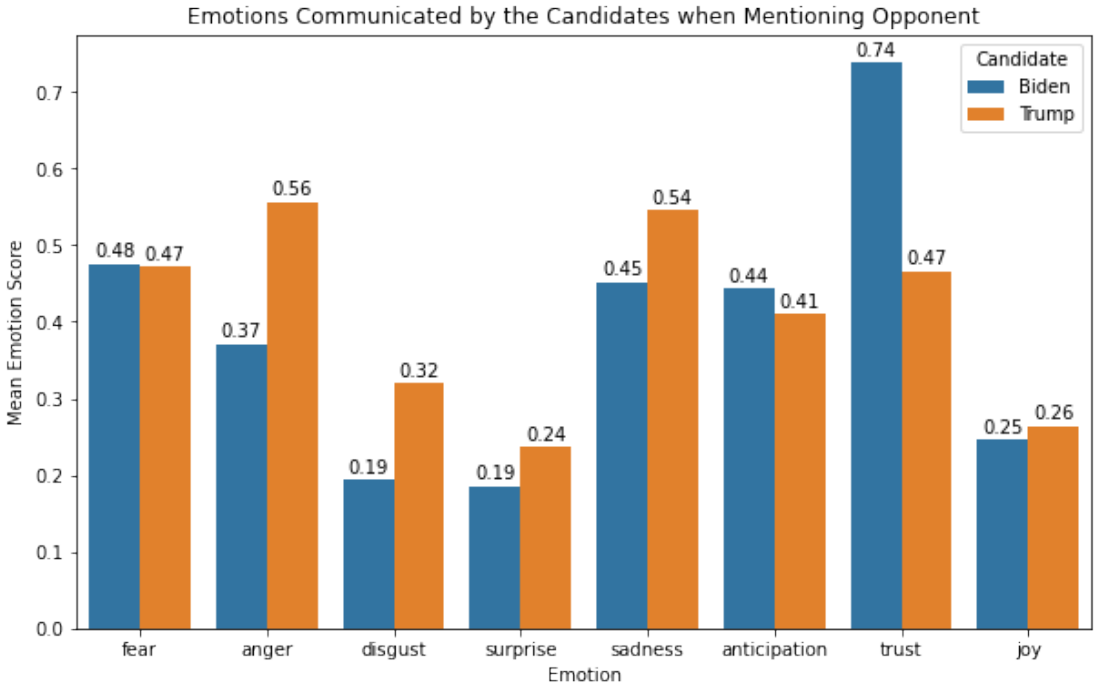


Figure 5.11 Tweet Emotional Composition When Mentioning Opponent

We see a stark difference from Figure 5.6 which shows the emotions communicated by the candidates overall. The difference is that Trump's emotional content is much more on par with Biden's when Trump is mentioning Biden. Interestingly, the top emotions present when Trump mentions Biden are anger, sadness, and fear—all more negative emotions. Biden's top emotions also change from trust, anticipation, and joy to trust, fear, and sadness. These changes suggest a negative campaign style for both Trump and Biden since their tweets are emotionally charged with fear, sadness, and anger.

User interaction with tweets mentioning the opponent also provides insight into the campaign style. Figure 5.12 compares the average amount of favorites overall to the average amount of favorites when the opponent is mentioned.

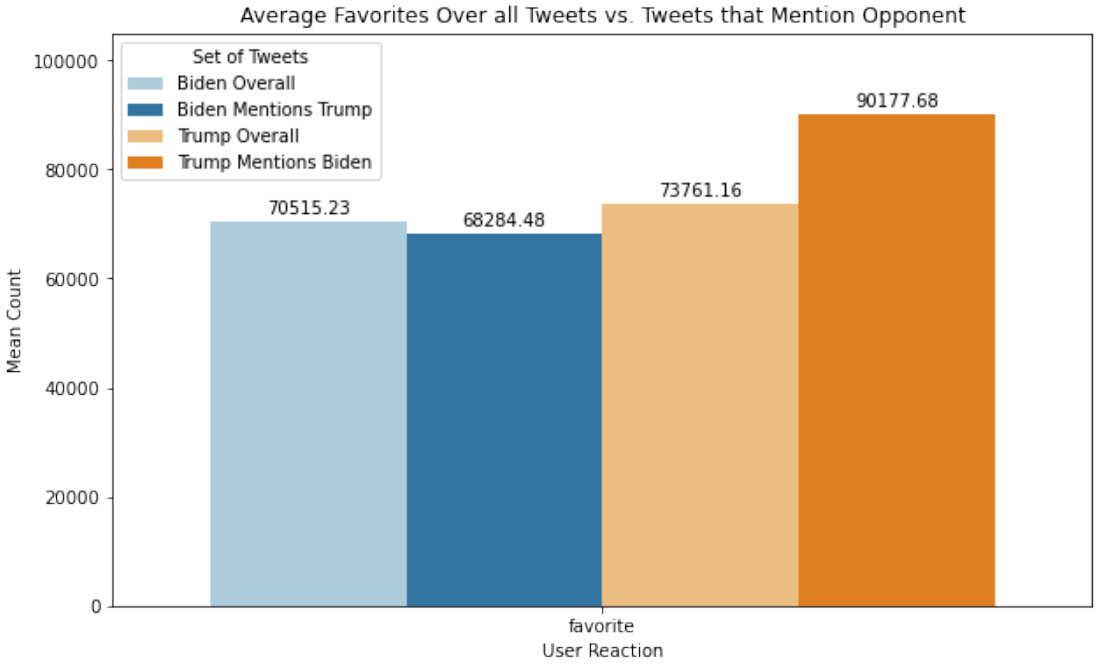


Figure 5.12 Average Change in Favorites When Mentioning Opponent

Biden's favorites do not significantly decrease when he mentions Trump even though we see a difference above. On the other hand, Trump's favorites significantly increase when he mentions Biden ($\mu_{\text{overall}} - \mu_{\text{mentions B}}$: 95% CI $[-26409, -6424]$, $t_{df:245} = -3.24$, $p = 0.0014$). This suggests that Trump's base may be more captivated by his negative campaign style than Biden's base is by his negative approach. We can further investigate this hypothesis by seeing if the replies and/or retweets significantly increase for Trump when he mentions Biden.

Figure 5.13 compares the average amount of replies and retweets.

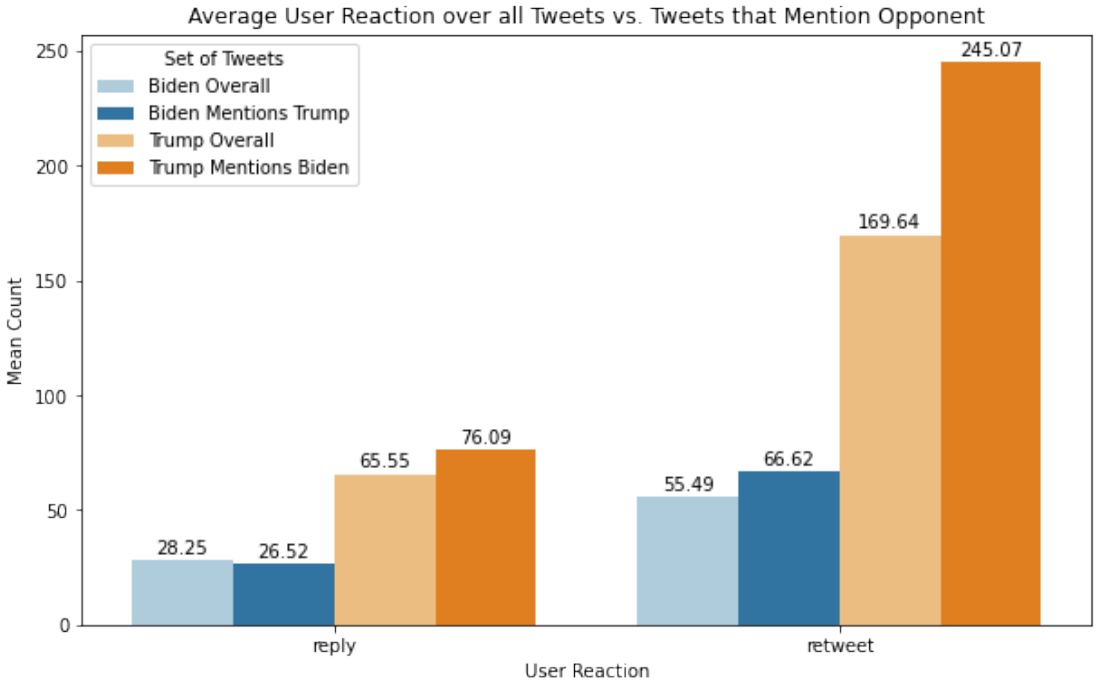


Figure 5.13 Average Change in User Reaction When Mentioning Opponent

As hypothesized, we see an increase in user interaction with Trump's tweets when he mentions Biden. The average amount of replies do not significantly increase for Trump and do not significantly decrease for Biden. However, the average amount of retweets significantly increases for Trump when he mentions Biden ($\mu_{\text{overall}} - \mu_{\text{mentions B}}$: 95% CI $[-145, -5.56]$, $t_{df:201} = -2.13$, $p = 0.0345$). Biden's average retweets do not significantly increase when he mentions Trump. This evidence further supports the proposition that Trump's negative campaign style is more effective in sparking user interaction than Biden's attempted negative campaign style.

Thus, we observe that both candidates attempt to employ a negative campaign style, but Trump's negative campaign is more effective in drawing user interaction.

5.1.4 Candidates in Discussion

This subsection widens the scope to tweets including either #biden or #trump in order to understand how Biden and Trump are discussed on Twitter. The previous subsections explore the social presence of the candidates via their tweets. This

does not give the full picture of how Twitter views each candidate. We visually examine the candidates in discussion on Twitter by constructing a categorized hashtag ego network. See Section 4.3 for details on ego networks.

Before presenting the ego networks, we want to recall the daily usage of hashtags in Section 3.2.1. Hashtags are not as popular as they once were which is evident in Figure 3.11 (shown again below as Figure 5.14).

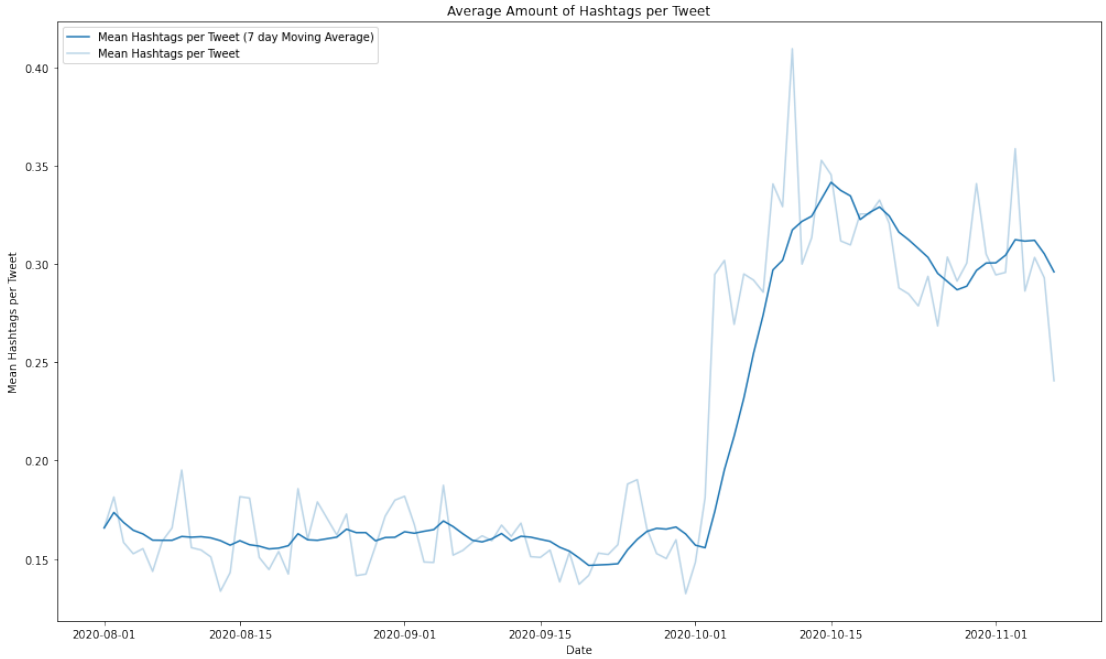


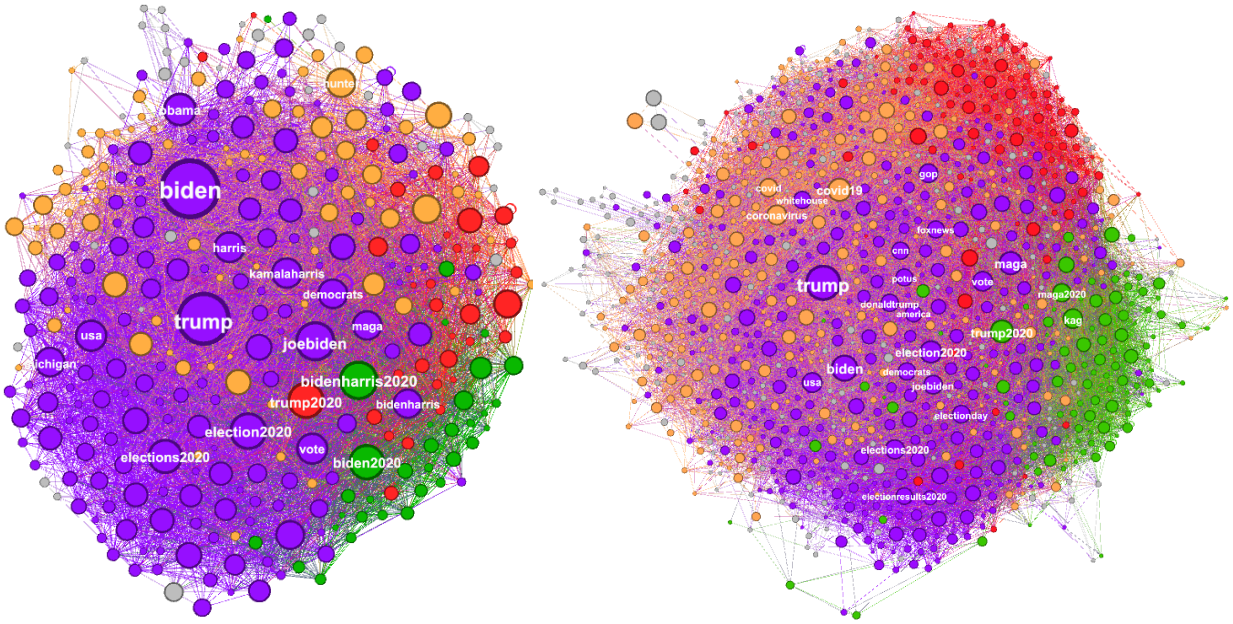
Figure 5.14 Average Amount of Hashtags per Tweet

From August 1st to October 1st, a tweet is expected to have approximately 0.1625 hashtags. Hashtag use then almost doubles during the month prior to the election. This aligns with the period of increased tweeting by the candidates visible in Figure 5.1.

Even though hashtag use is low, they still help us capture a snapshot of the candidates in discussion on Twitter since they specifically tag a tweet. After selecting the top 95% of hashtags to display (see Appendix C.1), determining their node size based on number of occurrences, and taking the logarithm of their sizes to better perceive relative size distribution (Appendix B.3), we categorize them into five groups:

- **Supporting**: Hashtags directly support a candidate (e.g. #bidenharris2020, #trump2020, #votehimout, #fourmoreyears)
- **Against**: Hashtags directly oppose a candidate (e.g. #votehimout, #bidencrimefamily)
- **Important**: Hashtags that refer to important topics raised during the election (e.g. #covid19, #hunter, #trumphascovid)
- **General**: Hashtags that carry general information about the 2020 election. These include media names, party names, important dates, etc. (e.g. #trump, #biden, #vote, #elections2020, #cnn)
- **Other**: Hashtags that do not fall into the other categories (e.g. #sundaythoughts, #follow, #wednesdaywisdom)

Once categorized, we construct the following ego networks (with the above colors for categories). The first two are centered around the general hashtags of "biden" and "trump." Biden's corresponding network has 339 nodes and 7253 edges (network density ≈ 0.127). Trump's corresponding network has 1091 nodes and 28,387 edges (network density ≈ 0.048). These two hashtags are "general" hashtags so we expect the networks to display a snapshot of mixed opinions from all sides.



(a) #Biden Network **(b)** #Trump Network

Figure 5.15 Candidate Ego Networks

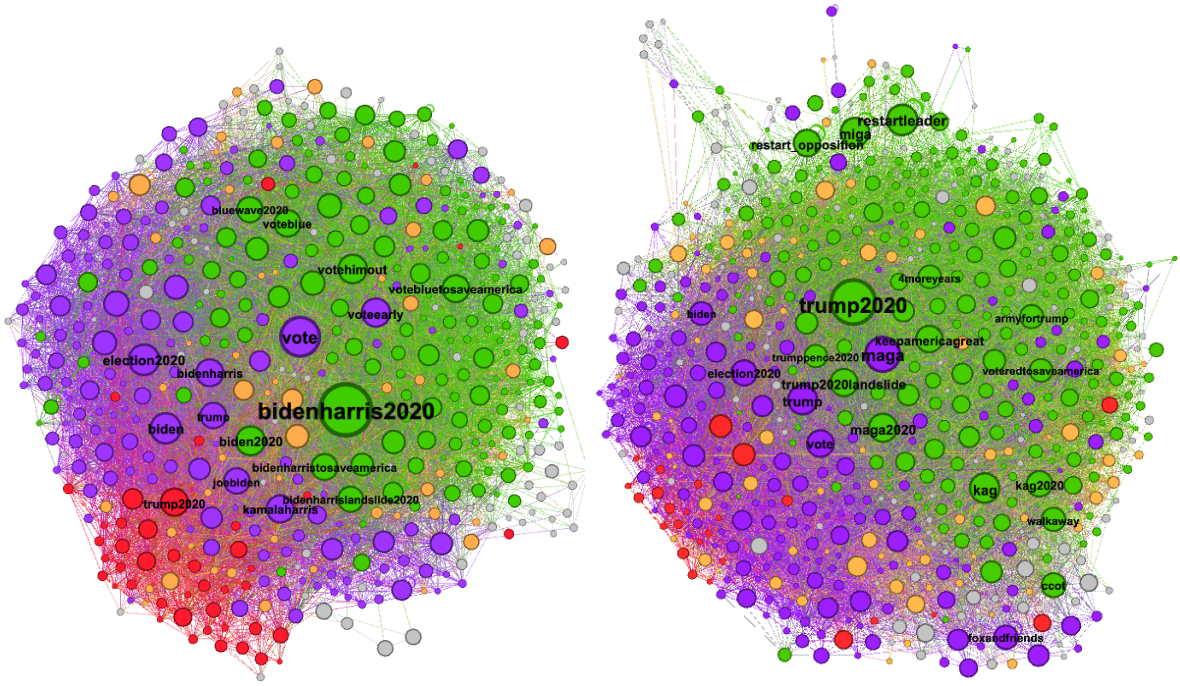
Immediately, we observe that Biden's network is more saturated with purple (i.e. "general" hashtags), and Trump's network is much larger. Based on relative size, Biden's network consists of 50.53% "general" hashtags while Trump's network consists of 33.47%. This difference suggests that in discussion on Twitter, people more often link general election information to Biden than to Trump. General election information typically relates to voting and information about election day. Thus, the greater proportion of general information tags in Biden's network highlights his campaign's goal to get people out to vote.

Trump's large network demonstrates the sheer size of his social reach. He appears in discussion often on Twitter which we see from the size and mess of his network. Approximately 20.26% of his network is attributed to "other" hashtags which only make up 8.21% of Biden's network. These hashtags are more random and niche, showing Trump's on Twitter since as his name infiltrates a spread of communities and topics.

While the "supporting" and "against" categories may look bigger on Trump's side, they are fairly similar between the two networks. Biden's network consists

of 9.4% "supporting" and 9.63% "against" compared to 9.94% "supporting" and 9.46% "against" for Trump. The slight difference suggests that Trump is discussed in a more positive light. Finally, we see that people tie the candidates to important topics roughly a quarter of the time (22.22% for Biden, 26.86% for Trump). The "important" hashtags contain contentious topics such as "COVID," "abortion," and "Amy Coney Barrett" as well as popular stories that plagued the election cycle such as "Hunter," "Burisma," and "fake news." It makes sense that Trump is commonly discussed in relation to these important stories since his responses and thoughts on these topics as president are widely scrutinized.

The second set of ego networks are centered around the supporting hashtags of "bidenharris2020" and "trump2020." These two campaign hashtags are selected because they are the most popular campaign-related hashtags in our data. The Biden-related network has 490 nodes and 11,093 edges (network density ≈ 0.093). The Trump-related network has 639 nodes and 16,302 edges (network density ≈ 0.08). These two hashtags are "supporting" hashtags so we expect the networks to display a snapshot of their base.



(a) #BidenHarris2020 Network **(b)** #Trump2020 Network

Figure 5.16 Campaign Ego Networks

Since these two hashtags directly support each campaign, we see that the networks are more saturated with "supporting" hashtags. Trump's network consists of 39.62% "supporting" and 3.87% "against" compared to Biden's 33.44% "supporting" and 8.82% "against." The greater amount of supporting and lesser amount of against for Trump suggests that users more commonly tag "trump2020" in Trump-supporting content than "bidenharris2020" is tagged in Biden-supporting content. Similarly to the first two networks, we see that Biden's network is more focused on discussing "general" election information at 32.14% compared to Trump's 28.63%.

Table 5.7 provides a summary of the relative sizes of the hashtag categories seen in the above ego networks. The maximum value in each category is bolded.

Ego	**Supporting (%)**	**Against (%)**	**General (%)**	**Important (%)**	**Other (%)**
Biden	9.4	**9.63**	**50.53**	22.22	8.21
Trump	9.94	9.46	33.47	**26.86**	**20.26**
BidenHarris2020	33.44	8.82	32.14	10.72	14.88
Trump2020	**39.62**	3.87	28.63	11.69	16.19

Table 5.7: Summary Hashtag Categories (by Node Size)

Overall, Figures 5.15 and 5.16 and Table 5.7 show that the candidates are discussed differently on Twitter. General election information is more commonly tied to Biden-related tags while "important" and "other" tags are more commonly linked to Trump.

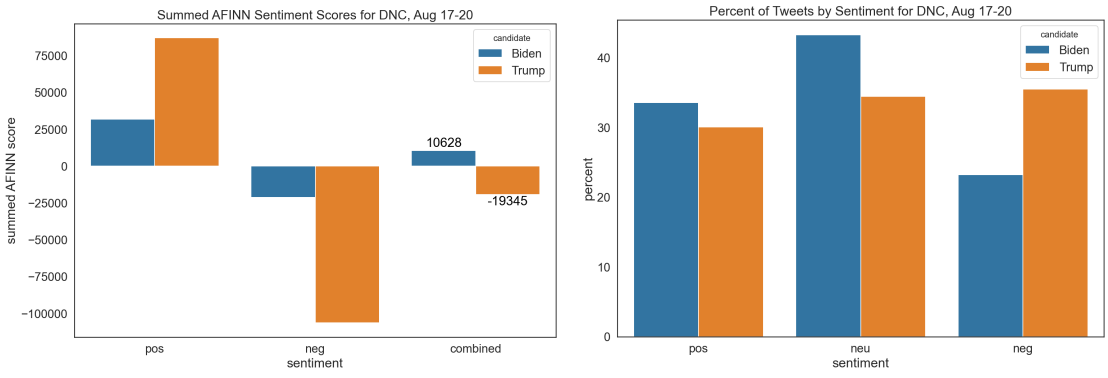
5.2 Campaign Momentum

To investigate campaign momentum, we now consider the Candidate Keyword Tweets (see Section 3.2.2) at specific events. This data widens the scope of our analysis, adding a layer of uncertainty because our sentiment analysis does not denote if a tweet is simultaneously positive for one candidate and negative for the other. Another drawback is that we use the AFINN library for sentiment analysis even though BERT is far more accurate. We use AFINN because its classifications have an unbounded range so tweets are weighted by the extremity of their

sentiment. Additionally, AFINN is computationally not intensive since it is an uncomplex dictionary-based method.

We analyze the momentum at seven events using the Sentiment Indicator (see Methods, Section 4.4). Recall that the SI aggregates the positive and negative scores such that the candidate with the higher sentiment count corresponds to the campaign with the positive momentum. The seven events are the same as discussed in the Social Presence section: the DNC, the RNC, the first presidential debate, Trump's COVID-19 diagnosis announcement, the vice presidential debate, the second presidential debate, and election day.

First, looking at the DNC, we expect that Biden should have positive momentum in comparison to Trump because the DNC is the first event truly kicking off and supporting Biden as the official Democratic nominee. We analyze the sentiment of Candidate Keyword Tweets from August 17th to August 20th. Figure 5.17a shows the AFINN score breakdown by sentiment polarity and the combined sentiment count, i.e. the SI. Figure 5.17b then shows the sentiment polarity breakdown by percentage.



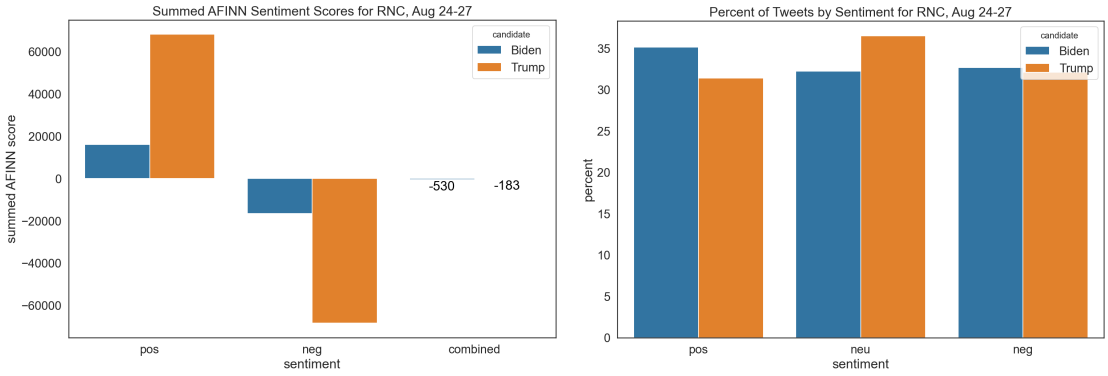
(a) SI **(b)** Sentiment Breakdown

Figure 5.17 Campaign Momentum at the DNC

The SI indicates that Biden has greater campaign momentum during the DNC since the sentiment count for the Biden keyword is greater than the sentiment count for the Trump keyword ($SI_b = 10,628 > SI_t = -19,345$). We notice that Trump has a greater amount of aggregated positive scores and aggregated nega-

tive scores. This suggests that there are more tweets mentioning Trump which agrees with the analysis done in this thesis thus far. By percentage breakdown, we see that the Biden keyword is tied with a greater percent of positive and neutral classifications. These results help validate the use of the SI.

Next, we look at the RNC. Similarly to the DNC, we expect Trump to have greater momentum since this is the event to kickoff and officially support his re-election campaign as the Republican nominee. We analyze the sentiment of Candidate Keyword Tweets from August 24th to August 27th. Figure 5.18a shows the AFINN score breakdown by sentiment polarity and the SI. Figure 5.18b then shows the sentiment polarity breakdown by percentage.

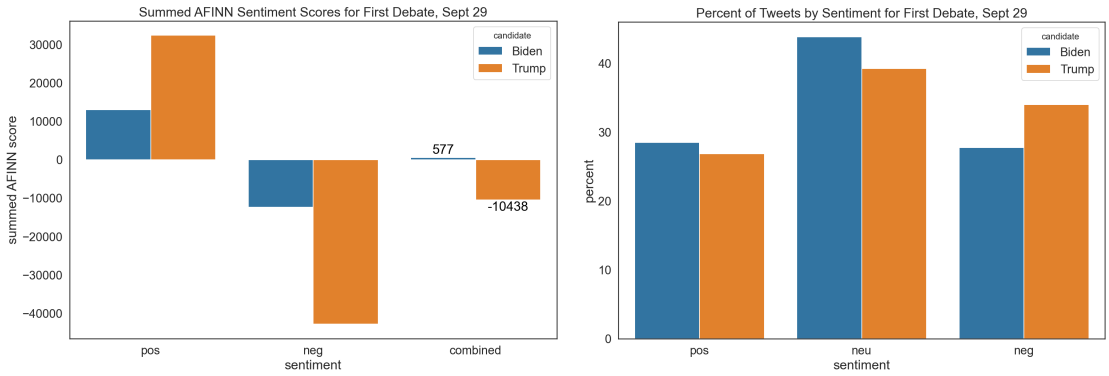


(a) SI **(b)** Sentiment Breakdown

Figure 5.18 Campaign Momentum at the RNC

As hypothesized, Trump has the greater momentum during the RNC ($SI_b = -530 < SI_t = -183$). Interestingly, the difference between Biden and Trump's SI is much closer than it was at the DNC. We again see that there are more tweets mentioning Trump since his aggregate positive scores and aggregate negative scores are approximately four times the corresponding aggregate scores for Biden. Figure 5.18b shows that there are more Biden-related positive tweets by percent, yet his overall SI is negative. This suggests that the negative tweets mentioning Biden had a greater magnitude on average compared to the positive tweets. Analyzing these two results together, even though the SI indicates positive momentum for Trump, the greater percent of positive Biden tweets and the small difference in SI suggest a better-than-expected outcome for Biden.

We now analyze the Candidate Keyword Tweets on the first presidential debate, September 29th. Neither Trump nor Biden had an outstanding performance, but Trump blundered and interrupted an unusual amount, causing many reactions on Twitter. Figure 5.19 shows the AFINN score breakdown, SI, and sentiment polarity breakdown by percentage.

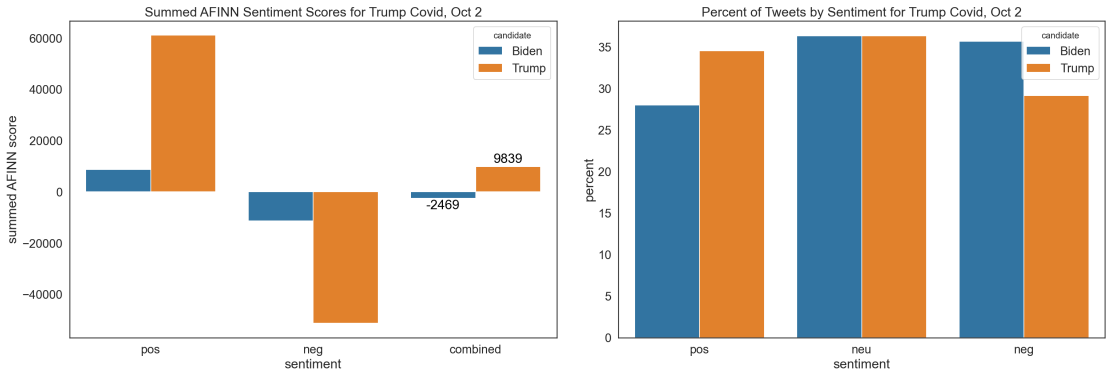


(a) SI **(b)** Sentiment Breakdown

Figure 5.19 Campaign Momentum at the First Presidential Debate

The SI suggests that Biden has the positive momentum over Trump after the first debate ($SI_b = 577 > SI_t = -10,438$). The small, positive score of the SI for Biden compared to the large, negative score for Trump indicates that instead of Biden having a blowout performance, Trump performed poorly. Figure 5.19b supports this conclusion since Trump has a greater percent of negative tweets than positive tweets. He also has a greater percent of negative tweets in comparison to Biden.

The next event is Trump's COVID-19 diagnosis, occurring on October 2nd. We suspect little momentum indications for Biden since this is event focuses on Trump, and Biden refrained from commenting on Trump's diagnosis. The SI will likely suggest that Trump has positive momentum since many messages of sympathy and recovery were directed at him. Figure 5.20 shows the AFINN score breakdown, SI, and sentiment polarity breakdown by percentage.

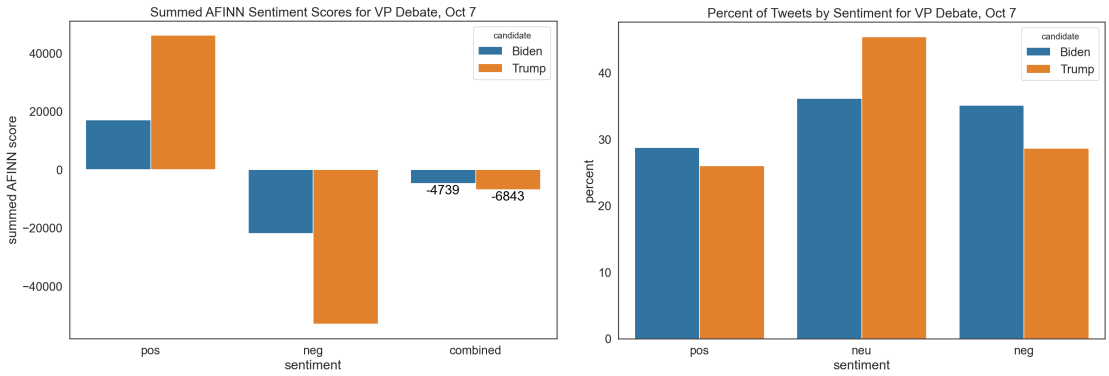


(a) SI **(b)** Sentiment Breakdown

Figure 5.20 Campaign Momentum at Trump's COVID-19 Diagnosis

As expected, we see a greater SI for Trump ($SI_b = -2469 < SI_t = 9839$). Interestingly, this is the first positive SI for Trump over the four events we have investigated. We see a greater percent of positive tweets directed at Trump, while a greater amount of negative tweets are directed at Biden in Figure 5.20b.

The fifth event we observe is the vice presidential debate, occurring on October 7th. Figure 5.21 shows the AFINN score breakdown, SI, and sentiment polarity breakdown by percentage.



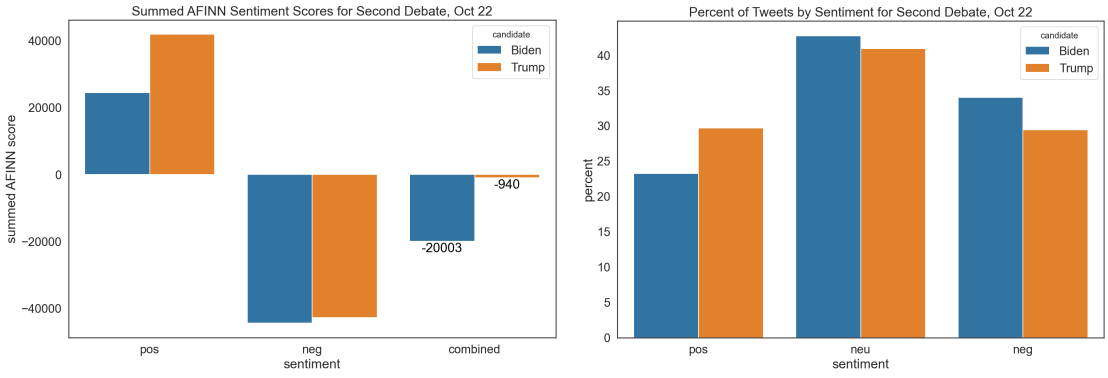
(a) SI **(b)** Sentiment Breakdown

Figure 5.21 Campaign Momentum at the Vice Presidential Debate

The SI suggests that Biden's campaign gains the momentum from the vice presidential debate ($SI_b = -4739 > SI_t = -6843$). Again, even though we see a greater percentage of negative tweets for Biden, the aggregated sentiment count shows that the average magnitude of Trump's negative tweets is larger. Thus,

we see a slight momentum lead for Biden, but we note a greater proportion of negative tweets for both candidates.

Next, we study the second presidential debate, occurring on October 22nd. Trump had a much better showing at the second debate in comparison with the first. Figure 5.22 shows the AFINN score breakdown, SI, and sentiment polarity breakdown by percentage.



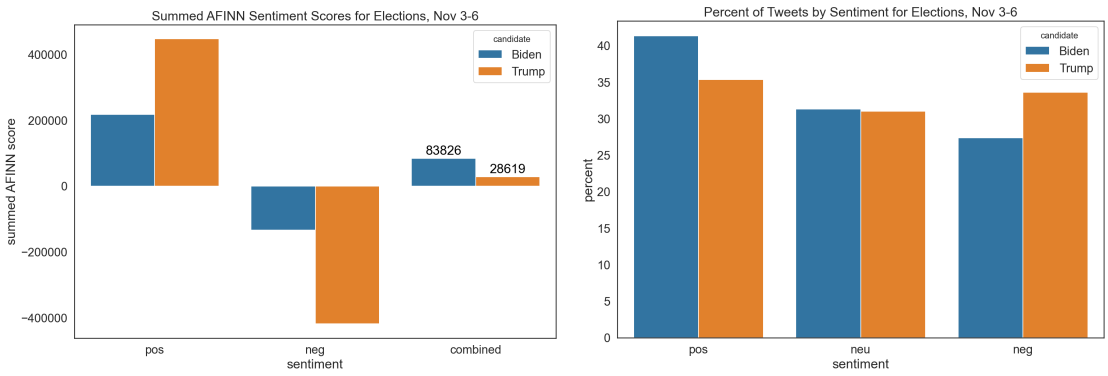
(a) SI **(b)** Sentiment Breakdown

Figure 5.22 Campaign Momentum at the Second Presidential Debate

The SI indicates that Trump has the greater momentum following the second debate ($\text{SI}_b = -20,003 < \text{SI}_t = -940$). The sentiment score breakdown in Figure 5.22a is fairly different from the breakdowns we have seen thus far. In prior sentiment score decompositions, Biden's aggregate positive and negative counts are lesser in magnitude. Here, the magnitude of Biden's aggregate negative sentiment score is greater than Trump. Explaining this difference needs to be further investigated. Potential explanations could be either a large number of people mentioning Trump and Biden in the same tweet (since the aggregate negative scores are similar) or some topic causing Twitter to erupt with negative emotions directed at Biden (such as Biden's comment about eventually moving away from fossil fuels).

Finally, we look at campaign momentum during the election, prior to the finalized decision on November 7th. This means we look at the sentiment from November 3rd to November 6th. Since Biden won the election on the 7th, we suspect to see positive campaign momentum for his campaign. Figure 5.23 shows the AFINN

score breakdown, SI, and sentiment polarity breakdown by percentage.



(a) SI **(b)** Sentiment Breakdown

Figure 5.23 Campaign Momentum at the Election

As suspected, the SI shows greater momentum for Biden's campaign ($SI_b = 83,826 > SI_t = 28,619$). Figure 5.23b also shows a greater percentage of positive tweets for Biden in comparison to neutral and negative tweets. Trump also has a greater percentage of positive tweets, but the difference between the percent of positive and negative tweets is not as drastic as the same difference for Biden. These results highlight the outcome of the election which ended in a Biden win.

Overall, we see Biden carry greater momentum four times out of seven. We see the momentum flip between campaigns. Out of the seven events, disregarding who has the greater momentum, Biden has a positive SI only three times, and Trump has a positive SI only twice. This suggests that the candidates are often discussed negatively which we see in the candidate ego networks in Figure 5.15. Since the SI verifies the majority of our hypotheses, we conclude that the SI is an effective measure of campaign momentum. We use the SI in the next section to understand how change in momentum may be associated with misinformation events.

5.3 Effect of Misinformation

To investigate the effect of misinformation on campaign momentum, we use the CI and an adaptation to the SI. We do not use ROC or RSI since the CI essentially

summarizes the same momentum changes. The SI via sentiment count is best used to compare momentum at one event. To observe the associated effect of misinformation, we compare two weeks' worth of data surrounding a misinformation event. Therefore, we adapt the SI by comparing how the average percentage of negative tweets changes from one week to the next. Overall, we examine the momentum change via change in sentiment and change in daily number of tweets by comparing the SI and CI constructed from a week's worth of Candidate Keyword Data before the event to the SI and CI constructed from a week's worth of said data after the event for three misinformation case studies.

5.3.1 Wisconsin Ballots

The first misinformation story is about absentee ballots found in a ditch in Wisconsin on September 24th. This story highlights a popular topic of misinformation during the 2020 presidential election: mail-in voting fraud. While three trays of mail were truly found in a ditch in Wisconsin, the sheriff reported that only "several" ballots were included. An election official later clarified that no Wisconsin ballots were found, but there could have been ballots from other states. These facts starkly contrast the claims made by Trump that "ballots in Wisconsin are being dumped in rivers or creeks" [21]. Thus, this story facilitated the spread of misinformation regarding the security of mail-in voting and voting fraud.

We start by analyzing the change in percent of negative tweets. Hypothetically, this misinformation story should benefit Trump since he targets mail-in voting and general voting fraud throughout the election. Figure 5.24 shows the change in average percent of negative tweets for Biden and Trump prior to and after the misinformation story.

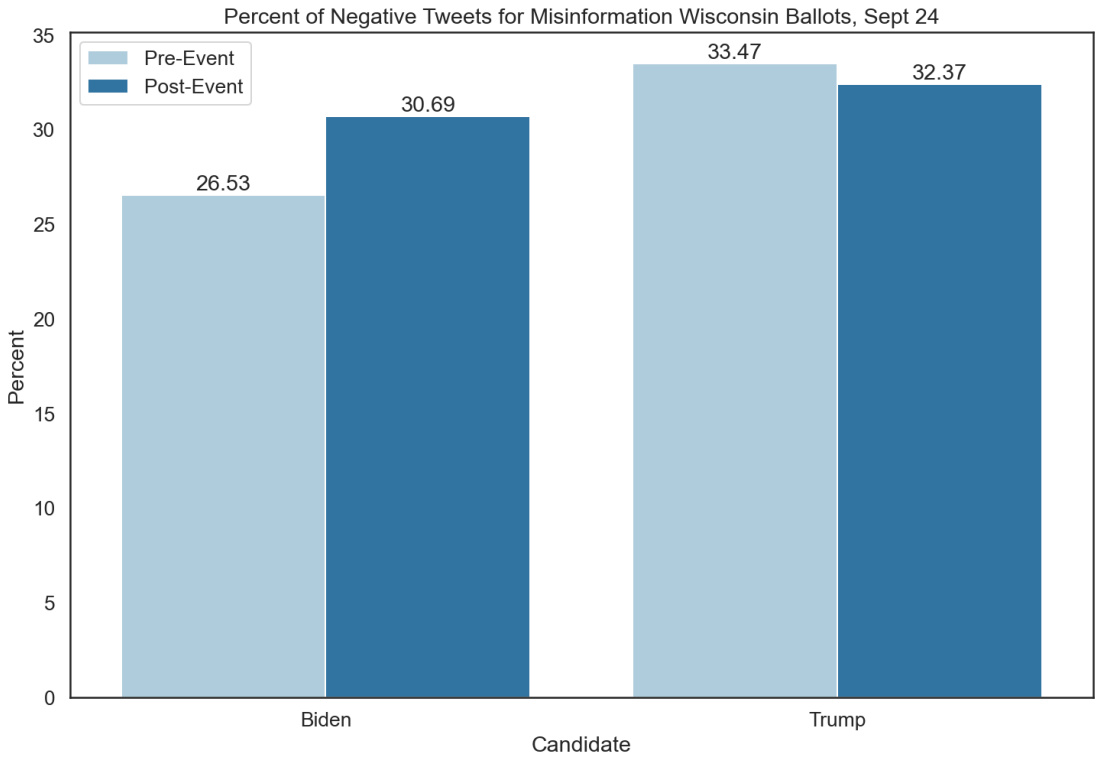
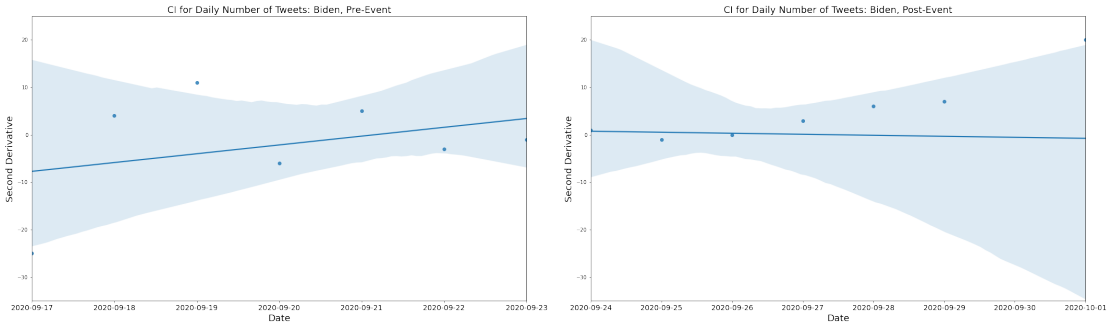


Figure 5.24 Change in Percent of Negative Tweets for Wisconsin Ballots Story

We see that the percent of negative tweets mentioning Biden increases after the misinformation event while the percent of negative tweets mentioning Trump decreases. A two-proportion z-test comparing the percent of negative tweets pre-event to the percent of negative tweets post-event for each candidate shows that both of these statements are significant. For Biden, the 95% confidence interval is [-0.046, -0.037], suggesting that the true difference between sample proportions (pre - post) is less than 0, leading us to reject the null hypothesis and conclude that the proportion of negative tweets related to Biden significantly increases after the misinformation event ($z = -18.49$, $p < 0.0001$). For Trump, the 95% confidence interval is [0.0081, 0.0139], suggesting that the true difference between sample proportions (pre - post) is greater than 0, leading us to reject the null hypothesis and conclude that the proportion of negative tweets related to Trump significantly decreases after the misinformation event ($z = 7.44$, $p < 0.0001$). Since this is not a causal study, we cannot conclude that the misinformation event is the reason why these momentum changes occur. However, we can conclude an association between these significant momentum changes and the misinformation event.

Before moving to the second misinformation event, let's look at the CI for this case study. The tweet attribute we analyze via the CI is the number of daily tweets. Figure 5.25a shows the CI from the week before the event and Figure 5.25b shows the CI from the week after the event for Biden-related tweets. The linear regression determines the CI momentum trend while the points plotted are the approximated second derivatives per day of interest (see Methods, Section 4.4).

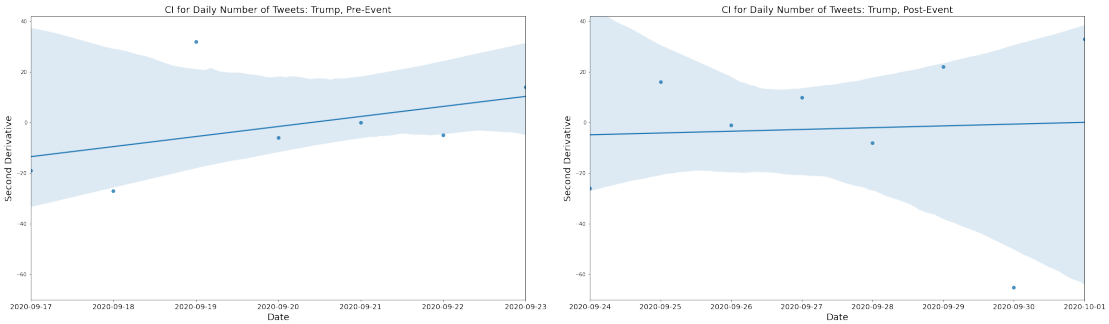


(a) Pre-Event **(b)** Post-Event

Figure 5.25 CI for Biden-Related Tweets, Wisconsin Ballots

Like the SI, we see that before the misinformation event, Biden's momentum was positive according to the CI's positive trend in Figure 5.25a. After the misinformation event, the CI has a slightly negative trend, indicating that Biden has negative momentum following the Wisconsin ballots story. To test if this relationship is significant, we perform a partial F test as described in Section 4.4. The F statistic is $F = 2.021$ with a p-value of $p = 0.1855$. Since $p > \alpha = 0.05$, the added interaction term between day and the pre-post indicator does not significantly improve the fit of the linear model, thus suggesting that there is not a significant change in momentum based on daily tweet count.

For Trump-related tweets, Figure 5.26a shows the CI from the week before the event and Figure 5.26b shows the CI from the week after the event.



(a) Pre-Event **(b)** Post-Event

Figure 5.26 CI for Trump-Related Tweets, Wisconsin Ballots

Visually, the CI trend for Trump seems to support the SI. Trump's momentum surrounding the misinformation event remains positive both before and after the story breaks. We perform a partial F test to determine if a significant change in momentum exists. The F statistic is $F = 1.289$ with a p-value of $p = 0.2826$. Since $p > \alpha = 0.05$, the added interaction term between day and the pre-post indicator does not significantly improve the fit of the linear model, thus indicating that there is not a significant change in momentum based on daily tweet count. While this insignificance does not show an association between targeted misinformation and an *increase* in positive momentum for Trump's campaign, it does demonstrate that the misinformation event does not associate with a significant *decrease* in Trump's momentum.

5.3.2 Hunter Biden New York Post Article

Next, we investigate our second misinformation story, this one regarding a New York Post article published on October 14th about Hunter Biden. The article claims that Hunter Biden set up a meeting between his father (Joe Biden) and a Ukranian businessman. This story appears to have been set up as an "October surprise" by the Trump campaign since the information in the article relies on data supposedly found on a hard drive provided to the Post by Trump's lawyer, Rudy Giuliani. The emails cited in the article lack corresponding metadata such as a timestamp or message ID number. Thus, the authenticity of the emails could not be determined. Twitter, along with Facebook, decided to restrict the sharing

of the article due to violations of their personal information policies and their hacked materials policies. This caused outrage on Trump's side of Twitter due to a feeling of censorship [17].

Figure 5.27 shows the change in average percent of negative tweets for Biden and Trump prior to and after this second misinformation story.

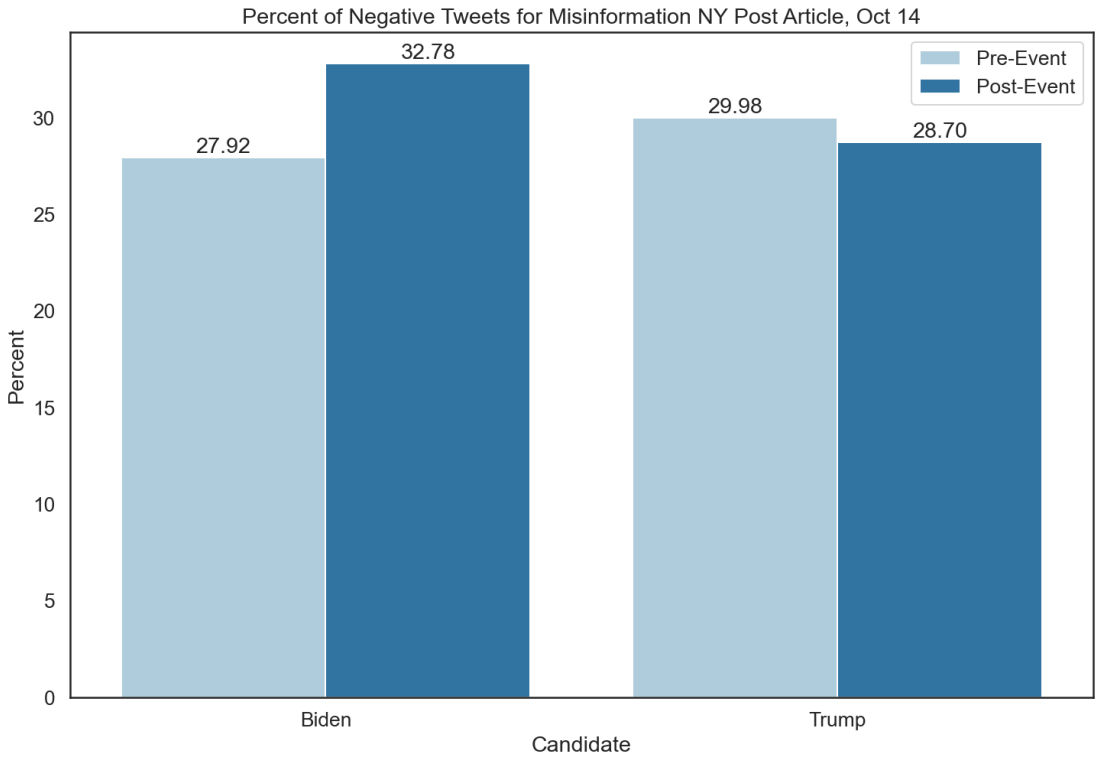
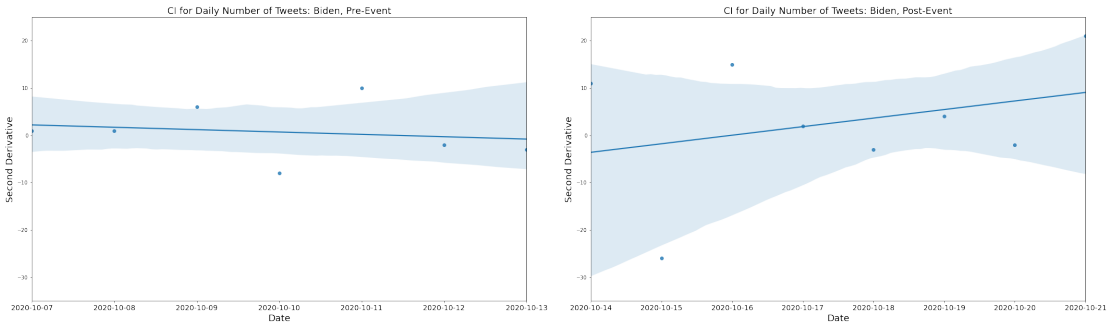


Figure 5.27 Change in Percent of Negative Tweets for NY Post Article

As with the first misinformation event, we see that the proportion of Biden-related negative tweets increases after the story, and the proportion of Trump-related negative tweets decreases after the story. A two-proportion z-test comparing the percent of negative tweets pre-event to the percent of negative tweets post-event for each candidate shows that both of these statements are significant. For Biden, the 95% confidence interval is [-0.052, -0.046], suggesting that the true difference between sample proportions (pre - post) is less than 0, leading us to reject the null hypothesis and conclude that the proportion of negative tweets related to Biden significantly increases after the misinformation event ($z = -31.31$, $p < 0.0001$). For Trump, the 95% confidence interval is [0.0108, 0.0148], suggesting that the

true difference between sample proportions (pre - post) is greater than 0, leading us to reject the null hypothesis and conclude that the proportion of negative tweets related to Trump significantly decreases after the misinformation event ($z = 12.67$, $p < 0.0001$). Again, we conclude that an association exists between these significant momentum changes and the misinformation event.

Now we investigate the CI for this case study. Figure 5.28a shows the CI from the week before the event and Figure 5.28b shows the CI from the week after the event for Biden-related tweets.



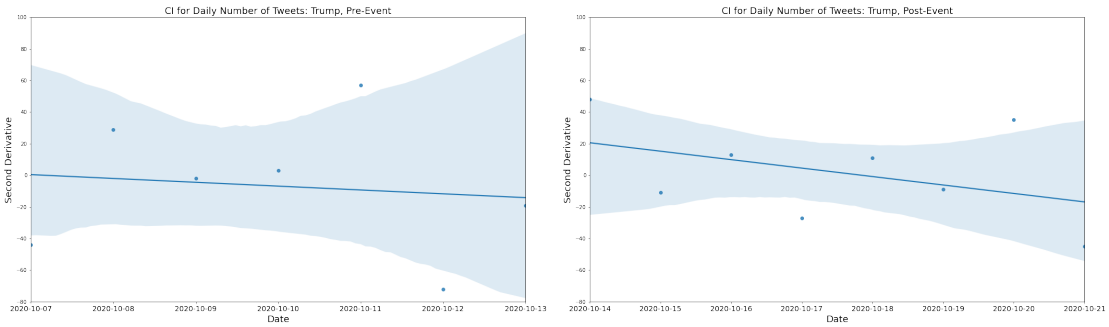
(a) Pre-Event **(b)** Post-Event

Figure 5.28 CI for Biden-Related Tweets, NY Post Article

Visually, the CI does not agree with the findings of the SI. Instead, we see that Biden's momentum pre-event was negative and becomes positive post-event. We perform a partial F test to determine if a significant change in momentum exists. The F statistic is $F = 0.656$ with a p-value of $p = 0.4351$. Since $p > \alpha = 0.05$, the added interaction term between day and the pre-post indicator does not significantly improve the fit of the linear model, thus indicating that there is not a significant change in momentum based on daily tweet count. Other explanatory variables are potentially influencing the CI trend such as the number of days until the election. (Recall that the candidates exhibit increased tweeting behavior for the month prior to the election as highlighted in Section 5.1.1). For this case in particular, Twitter's article-sharing restriction may also influence the momentum change surrounding this story.

For Trump-related tweets, Figure 5.29a shows the CI from the week before the

event and Figure 5.29b shows the CI from the week after the event.



(a) Pre-Event **(b)** Post-Event

Figure 5.29 CI for Trump-Related Tweets, NY Post Article

The CI trend for Trump also does not support the SI. Trump's momentum surrounding the misinformation event remains negative both before and after the story breaks. Again, other explanatory variables are potentially influencing the CI trend. A partial F test results in an F statistic of $F = 0.094$ with a p-value of $p = 0.7643$. Since $p > \alpha = 0.05$, the added interaction term between day and the pre-post indicator does not significantly improve the fit of the linear model, thus indicating that there is not a significant change in momentum based on daily tweet count.

5.3.3 Stop the Steal

The final misinformation event that we consider is the "Stop the Steal" movement. This event differs from the other two because it is a longer-living movement without a clear beginning whereas the first two events can be tied to an initial day. In order to retain comparable results, we tie the movement to election day, November 3rd. The "Stop the Steal" movement is a conservative-backed protest to overturn the election results due to misinformed claims of "widespread election fraud" [45]. We expect suspicious results in determining the sole effect of the misinformation movement since the date range conflates the beginning of the movement and the widespread positivity directed at Biden for winning the election (see Section 5.2). However, we still investigate the SI and CI.

Figure 5.30 shows the change in average percent of negative tweets for Biden and Trump prior to and after this third misinformation event.

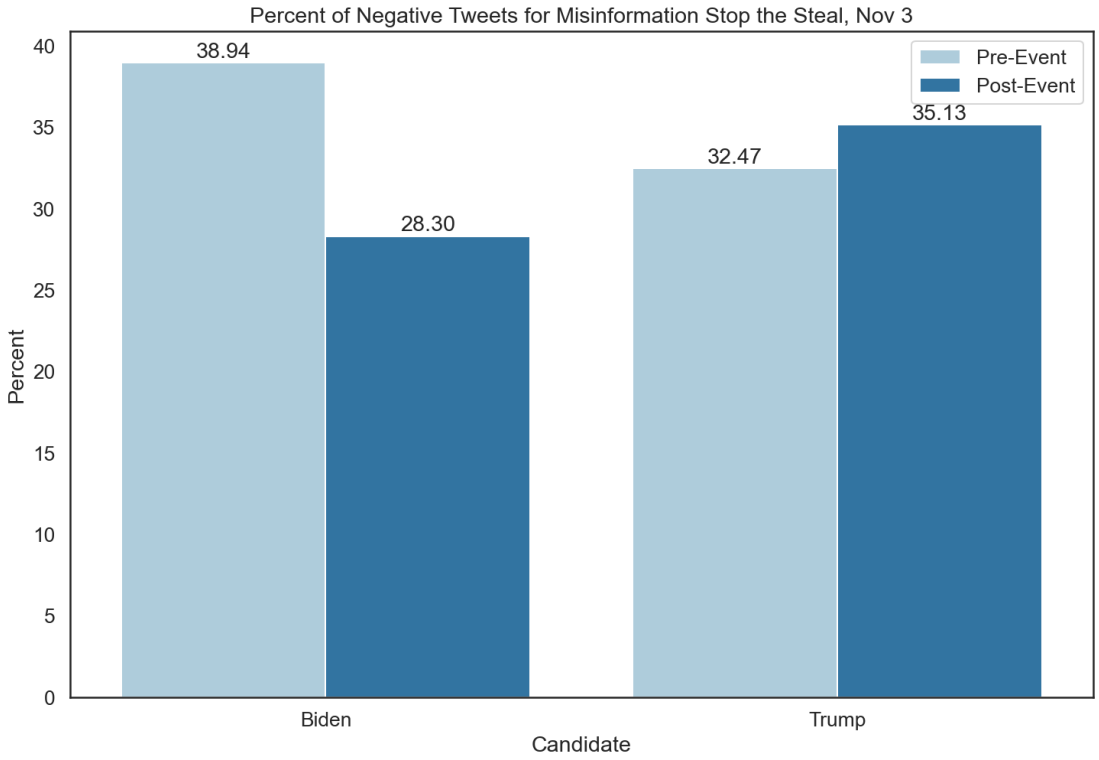
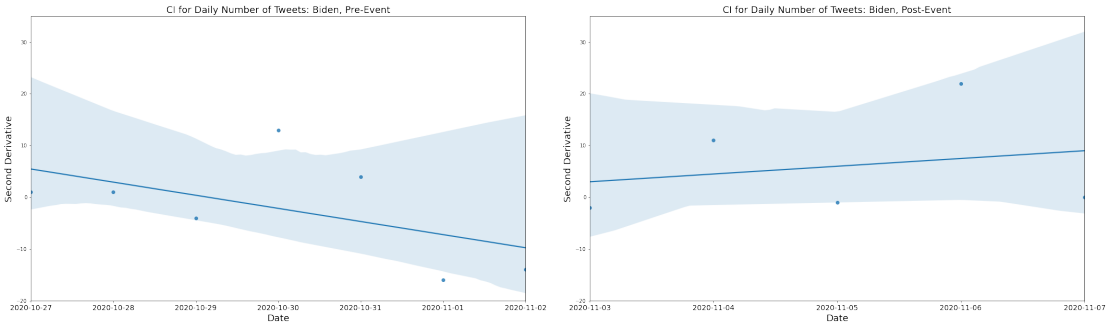


Figure 5.30 Change in Percent of Negative Tweets for Stop the Steal

Unlike the other two events, we see that the proportion of Biden-related negative tweets decreases after the beginning of the movement, and the proportion of Trump-related negative tweets increases. A two-proportion z-test comparing the percent of negative tweets pre-event to the percent of negative tweets post-event for each candidate shows that both of these statements are significant. For Biden, the 95% confidence interval is [0.104, 0.109], suggesting that the true difference between sample proportions (pre - post) is greater than 0, leading us to reject the null hypothesis and conclude that the proportion of negative tweets related to Biden significantly decreases after the misinformation event ($z = 82.52$, $p < 0.0001$). For Trump, the 95% confidence interval is [-0.0284, -0.0248], suggesting that the true difference between sample proportions (pre - post) is less than 0, leading us to reject the null hypothesis and conclude that the proportion of negative tweets related to Trump significantly increases after the misinformation event ($z = -29.16$,

$p < 0.0001$). As suggested earlier, this data likely reflects the sentiment and corresponding momentum from the election instead of the misinformation event.

For the CI, Figure 5.31a shows the CI from the week before the event and Figure 5.31b shows the CI from the week after the event for Biden-related tweets.

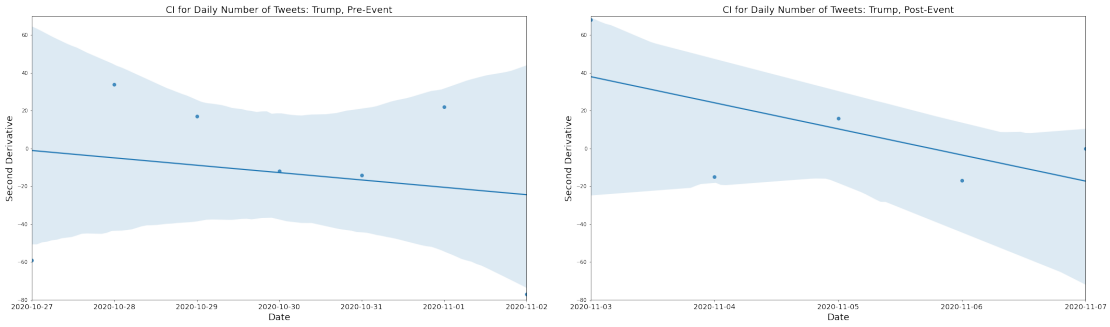


(a) Pre-Event **(b)** Post-Event

Figure 5.31 CI for Biden-Related Tweets, Stop the Steal

The CI trend supports the conclusion from the SI: namely, Biden gains positive momentum following the election (or start of the misinformation movement). Testing this significance via a partial F test produces an F statistic of $F = 3.339$ with a p-value of $p = 0.1103$. Since $p > \alpha = 0.05$, the added interaction term between day and the pre-post indicator does not significantly improve the fit of the linear model, thus indicating that there is not a significant change in momentum based on daily tweet count. This insignificance shows that the visual trend is not as strong as it may appear. As in the previous CI trend analyses, other explanatory variables are likely influencing the momentum—in this case, likely the outcome of the election.

For Trump-related tweets, Figure 5.32a shows the CI from the week before the event and Figure 5.32b shows the CI from the week after the start of the movement.



(a) Pre-Event **(b)** Post-Event

Figure 5.32 CI for Trump-Related Tweets, Stop the Steal

The CI visually corroborates the SI outcome that Trump's campaign has negative momentum following the election. We perform a partial F test to determine if a significant change in momentum exists. The F statistic is $F = 0.825$ with a p-value of $p = 0.3939$. Since $p > \alpha = 0.05$, the added interaction term between day and the pre-post indicator does not significantly improve the fit of the linear model, thus indicating that there is not a significant change in momentum based on daily tweet count. This insignificance shows that no significant momentum change (neither positive nor negative) occurs for Trump's campaign suggesting that his momentum remains negative. Trump's negative momentum potentially reveals the *need* for a successful misinformation movement to boost his momentum.

5.3.4 Discussion

From these three case studies, we see that misinformation is likely positively associated with a change in campaign momentum. The first example of mail-in voting fraud plays on the misleading claims from the Trump campaign attempting to discredit absentee voting during a global pandemic. The SI shows that this story likely associates with the desired effect of hurting Biden's campaign momentum and amplifying Trump's momentum. On the other hand, the CI demonstrates an insignificant change in momentum for both campaigns. This CI result does not disqualify the SI outcome since the CI judges momentum based on the number of daily tweets authored by the candidates while the SI determines momentum based on the sentiment surrounding the candidates.

The second example regarding the NY Post article about Hunter Biden suggests both desired and undesired associated effects. The SI supports the idea that the misinformation story associates with negative campaign momentum for Biden and increased momentum for Trump. However, the CI again shows that neither campaign experienced a significant change in momentum. This highlights a possible association with Twitter restricting the sharing of the article. This association should be further researched to understand the effect Twitter can have on restricting misinformation spread.

The third example does not show any intended success from the "Stop the Steal" misinformation movement. Instead, the SI and CI are both likely associated with data supporting Biden's campaign since he won the election during the examined time period. The drastic decrease in Trump's momentum potentially highlights the need for the "Stop the Steal" movement in order to reinvigorate Trump's momentum.

This chapter investigated the candidates' social presence on Twitter and momentum changes tied to their campaigns. The next chapter concludes the thesis by summarizing key takeaways and proposing future research directions.

Chapter 6

Conclusion

6.1 Takeaways

This thesis explores the 2020 U.S. presidential candidates' social presence on Twitter, their campaign momentum, and the effect of misinformation. Three limitations present are: the sample does not include all of the candidate tweets and it is only roughly representative of a day on Twitter, the dictionary-based sentiment models used are not very accurate, and the observational nature of the study means that our conclusions are not causal. Thus, the following conclusions are presented with caution.

We find that both candidates tend to increase their tweeting behavior in the days surrounding a notable event. The trends also show that Trump's tweeting behavior is more active. Additionally, both candidates increase their activity in the month prior to the election. We note that in this period of time before the election, Biden sees more user interaction via reply while Trump sees more interaction via retweet.

Examining their engagement style, we see that Biden communicates with emotionally charged messages of trust, anticipation, and joy. Trump's tweets are far less emotionally charged. From investigating their most positive and most negative tweets, we find that Biden discusses his political connections, accomplishments, and goals for the country. Trump's tweets comparatively inspire fear and attack Biden. This agrees with our campaign style findings since Trump's tweets portray

a negative style when mentioning Biden. His tweets mentioning Biden are composed of fear, sadness, and anger. Increased user interaction with these tweets suggests that his negative campaign style resonates with his base. Biden's tweets mentioning Trump are more negative in nature, but trust is still the top emotion present, thus showing a slight negative campaign style. Unlike Trump, Biden's user interactions do not significantly change when he mentions Trump. This suggests that Trump's base is more engaged by a negative campaign style relative to Biden's base.

In regards to campaign momentum, we find that Biden has slightly more instances of positive campaign momentum in comparison to Trump's campaign. The events that we investigate demonstrate the efficacy of the Sentiment Indicator (SI) as a measure of campaign momentum. We then use the SI and Curvature Indicator (CI) to uncover associations between misinformation and changes in campaign momentum. Overall, we see an association between targeted misinformation events and changes in campaign momentum. Specifically, we see an example of mail-in voting fraud misinformation associated with the desired effect of hurting Biden's campaign momentum and amplifying Trump's momentum. The other two misinformation stories exhibit known covariates (Twitter's influence in the second and the election in the third) so the association is not as clear as the first misinformation story.

6.2 Future Directions

We propose several paths for future research. First, in the realm of sentiment analysis with Twitter data, future work should continue researching the BERT model and other contextually-based models. The BERT model demonstrates good precision and recall scores for every category suggesting that this model should be used for sentiment analysis of Twitter. Part of this future work should focus on data preprocessing to investigate whether or not a single token for user mentions improves accuracy. We also propose that researchers construct their own dataset

of labeled tweets from the subset of Twitter they are interested in studying. This path additionally suggests the comparison of models trained on different sets of Twitter data (i.e., from different time periods and from different spheres of Twitter).

In the context of understanding social presence on Twitter, we suggest future research into the meaning of a retweet versus a reply. Does a retweet indicate support and a reply indicate disagreement? Future work can help uncover this correlation. Understanding the specific meanings of retweets and replies would help better interpret user interactions with notable figures on Twitter.

Lastly, on the theme of misinformation, we propose future research into the effect Twitter has on restricting misinformation spread. The Hunter Biden New York Post article presents an interesting case study because Twitter restricted the sharing of the article which correlates with insignificant results in momentum change, suggesting that the targeted misinformation story did not have the intended effect. This correlation highlights the need for an exploration into the effects Twitter may have on reducing misinformation spread. Does restricting the spread of an article, hashtag, or keyword reduce the spread and influence of misinformation on Twitter?

6.3 Conclusion

Overall, this thesis demonstrates the use of Twitter as a data source to investigate the social presence of influential members. Through a political lens, one can research the temporal characteristics, engagement style, and campaign style that a candidate employs. The resulting information can influence a candidate's campaign by identifying themes and tones that best resonate with their audience. An opponent can also use the same research to better understand the flaws present in both campaigns and to devise methods to reach an opposing base. In today's world, misinformation is one tactic to increase the momentum of one campaign while hurting the momentum of an opposing campaign. This thesis shows that

analyzing Twitter can help discover these momentum shifts which may help campaigns mitigate the effect of misinformation.

References

[1] Ameeta Agrawal and Aijun An. "Unsupervised Emotion Detection from Text Using Semantic and Syntactic Relations". In: *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 01*. WI-IAT '12. USA: IEEE Computer Society, Dec. 2012, pp. 346–353. ISBN: 9780769548807. (Visited on 02/17/2021).

[2] Giulio Angiani et al. "A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter". In: *KDWeb*. 2016.

[3] Michael Arrington. *Odeo Releases Twttr*. en-US. URL: https://social.techcrunch.com/2006/07/15/is-twttr-interesting/ (visited on 02/18/2021).

[4] Noureddine Azzouza, Karima Akli-Astouati, and R. Ibrahim. "TwitterBERT: Framework for Twitter Sentiment Analysis Based on Pre-trained Language Model Representations". In: *IRICT*. 2019. DOI: 10.1007/978-3-030-33582-3_41.

[5] Philip Bump. "Analysis — At least five people close to Trump engaged with Russian Twitter trolls from 2015 to 2017". en-US. In: *Washington Post* (). ISSN: 0190-8286. URL: https://www.washingtonpost.com/news/politics/wp/2017/11/02/at-least-five-people-close-to-trump-engaged-with-russian-twitter-trolls-from-2015-to-2017/ (visited on 02/18/2021).

[6] *Changelog*. en. URL: https://developer.twitter.com/en/updates/changelog (visited on 03/08/2021).

[7] Kevin Collier. *Twitter Was Warned About This Russian Troll Account But Refused To Take It Down*. en. Oct. 2017. URL: https://www.buzzfeednews.com/article/kevincollier/twitter-was-warned-repeatedly-about-this-fake-account-run (visited on 02/18/2021).

[8] Colin Crowell. *Our approach to bots and misinformation*. en_us. June 2017. URL: https://blog.twitter.com/en_us/topics/company/2017/Our-Approach-Bots-Misinformation.html (visited on 02/18/2021).

[9] Robert Darnton. *The True History of Fake News*. en. URL: https://www.nybooks.com/daily/2017/02/13/the-true-history-of-fake-news/ (visited on 02/18/2021).

[10] *Definitions*. en. URL: https://mediamanipulation.org/definitions (visited on 02/18/2021).

[11] *Disinformation, 'Fake News' and Influence Campaigns on Twitter*. en-US. URL: https://knightfoundation.org/reports/disinformation-fake-news-and-influence-campaigns-on-twitter/ (visited on 02/18/2021).

[12] Joan Donovan. *Thank you for posting: Smoking's lessons for regulating social media*. en. URL: https://www.technologyreview.com/2020/10/05/1009231/social-media-facebook-tobacco-secondhand-smoke/ (visited on 02/18/2021).

[13] *Emoji sentiment data*. en. URL: https://kaggle.com/thomasseleck/emoji-sentiment-data (visited on 03/21/2021).

[14] *Frequently Asked Questions*. publisher: Defending Democracy Together. URL: https://russiatweets.com/page/faq (visited on 02/18/2021).

[15] Daniel Funke and Amy Sherman. *Trump did not ask supporters for $421 million to help him recover from COVID-19*. en-US. Oct. 2020. URL: https://www.politifact.com/factchecks/2020/oct/02/tweets/trump-did-not-ask-supporters-421-million-help-him-/ (visited on 02/18/2021).

[16] Z. Jianqiang and G. Xiaolin. "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis". In: *IEEE Access* 5 (2017), pp. 2870–2879. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2017.2672677.

[17] Tom Kertscher. *A look behind the NY Post headline about Joe Biden and a Ukraine meeting*. en-US. Oct. 2020. URL: https://www.politifact.com/article/2020/oct/15/look-behind-ny-post-headline-about-joe-biden-and-u/ (visited on 03/16/2021).

[18] Ema Kušen and Mark Strembeck. "Politics, sentiments, and misinformation: An analysis of the Twitter discussion on the 2016 Austrian Presidential Elections". en. In: *Online Social Networks and Media* 5 (Mar. 2018), pp. 37–50. ISSN: 24686964. DOI: 10.1016/j.osnem.2017.12.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S2468696417301088 (visited on 02/18/2021).

[19] Kalev Leetaru. *Is Twitter's Spritzer Stream Really A Nearly Perfect 1% Sample Of Its Firehose?* en. Feb. 2019. URL: https://www.forbes.com/sites/kalevleetaru/2019/02/27/is-twitters-spritzer-stream-really-a-nearly-perfect-1-sample-of-its-firehose/ (visited on 09/16/2020).

[20] *List of most-followed Twitter accounts*. en. Page Version ID: 1006919089. Feb. 2021. URL: https://en.wikipedia.org/w/index.php?title=List_of_most-followed_Twitter_accounts&oldid=1006919089 (visited on 02/18/2021).

[21] Eric Litke. *Trump whiffs describing Wisconsin ballot case*. en-US. Oct. 2020. URL: https://www.politifact.com/factchecks/2020/oct/09/donald-trump/trump-whiffs-describing-wisconsin-ballot-case/ (visited on 03/16/2021).

[22] Jane Mayer. *How Russia Helped Swing the Election for Trump*. en-us. URL: https://www.newyorker.com/magazine/2018/10/01/how-russia-helped-to-swing-the-election-for-trump (visited on 02/18/2021).

[23] *Misinformation Has Created a New World Disorder*. en. DOI: 10.1038/scientificamerican0919-88. URL: https://www.scientificamerican.com/article/misinformation-has-created-a-new-world-disorder/ (visited on 02/18/2021).

[24] D. Brendan Nagle and Stanley Mayer Burstein, eds. *The Ancient World: Readings in Social and Cultural History*. 4th ed. OCLC: ocn268789164. New York: Prentice Hall, 2010. ISBN: 9780205691876.

[25] Pandu Nayak. *Understanding searches better than ever before*. en. Oct. 2019. URL: https://blog.google/products/search/search-language-understanding-bert/ (visited on 02/18/2021).

[26] Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen. "BERTweet: A pre-trained language model for English Tweets". In: *arXiv:2005.10200 [cs]* (Oct. 2020). arXiv: 2005.10200. URL: http://arxiv.org/abs/2005.10200 (visited on 02/18/2021).

[27] *NRC Emotion Lexicon*. URL: https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm (visited on 03/16/2021).

[28] Ciara O'Rourke. *Viewers mistake shadow on Biden's neck for a mask*. en-US. Oct. 2020. URL: https://www.politifact.com/factchecks/2020/oct/08/viral-image/viewers-mistake-shadow-bidens-neck-mask/ (visited on 02/18/2021).

[29] Sarah Perez. *Twitter is building 'Birdwatch,' a system to fight misinformation by adding more context to tweets*. en-US. URL: https://social.techcrunch.com/2020/10/02/twitter-is-building-birdwatch-a-system-to-fight-misinformation-by-adding-more-context-to-tweets/ (visited on 02/18/2021).

[30] *Permanent suspension of @realDonaldTrump*. en_us. URL: https://blog.twitter.com/en_us/topics/company/2020/suspension.html (visited on 03/25/2021).

[31] Marco Pota et al. "A Subword-Based Deep Learning Approach for Sentiment Analysis of Political Tweets". In: *2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA)*. Krakow: IEEE, May 2018, pp. 651–656. ISBN: 9781538653951. DOI: 10.1109/WAINA.2018.00162. URL: https://ieeexplore.ieee.org/document/8418147/ (visited on 02/18/2021).

[32] Marco Pota et al. "An Effective BERT-Based Pipeline for Twitter Sentiment Analysis: A Case Study in Italian". en. In: *Sensors* 21.1 (Jan. 2021), p. 133. DOI: 10.3390/s21010133. URL: https://www.mdpi.com/1424-8220/21/1/133 (visited on 02/18/2021).

[33] Procopius. *The Secret History*. Trans. by Richard Atwater. University of Michigan Press, 1961. URL: https://sourcebooks.fordham.edu/basis/procop-anec.asp (visited on 02/18/2021).

[34] *Q3'20 Investor Fact Sheet*. 2020. URL: https://s22.q4cdn.com/826641620/files/doc_financials/2020/q3/Q3_20__InvestorFactSheet.pdf.

[35] *Report of the Select Committee on Intelligence United States Senate on Russian Active Measures Campaigns and Interference in the 2016 U.S. Election*. Tech. rep. Volume 2: Russia's Use of Social Media with Additional Views. URL: https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf.

[36] Aliza Rosen. *Tweeting Made Easier*. en_us. Nov. 2017. URL: https://blog.twitter.com/en_us/topics/product/2017/tweetingmadeeasier.html (visited on 02/18/2021).

[37] Yoel Roth and Noah Pickles. *Updating our approach to misleading information*. en_us. May 2020. URL: https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html (visited on 02/18/2021).

[38] David Sarno. *Twitter creator Jack Dorsey illuminates the site's founding document. Part I*. en-US. Feb. 2009. URL: https://latimesblogs.latimes.com/technology/2009/02/twitter-creator.html (visited on 02/18/2021).

[39] Katie Sehl. *Top Twitter Demographics That Matter to Social Media Marketers*. en-US. May 2020. URL: https://blog.hootsuite.com/twitter-demographics/ (visited on 02/18/2021).

[40] Elisa Shearer and Elizabeth Grieco. *Americans Are Wary of the Role Social Media Sites Play in Delivering the News*. en-US. Oct. 2019. URL: https://www.journalism.org/2019/10/02/americans-are-wary-of-the-

role-social-media-sites-play-in-delivering-the-news/ (visited on 02/18/2021).

[41] Yuke Shi et al. *Using Social Media and News Sentiment Data to Construct a Momentum Strategy*. en. SSRN Scholarly Paper ID 3406075. Rochester, NY: Social Science Research Network, Sept. 2017. URL: https://papers.ssrn.com/abstract=3406075 (visited on 02/18/2021).

[42] P. W. Singer and Emerson T. Brooking. *Likewar: the weaponization of social media*. Boston: Eamon Dolan/Houghton Mifflin Harcourt, 2018. ISBN: 9781328695758.

[43] Kit Smith. *60 Incredible and Interesting Twitter Stats and Statistics*. en-US. Jan. 2020. URL: https://www.brandwatch.com/blog/twitter-stats-and-statistics/ (visited on 03/08/2021).

[44] Mike Thelwall et al. "Sentiment strength detection in short informal text". In: *Journal of the American Society for Information Science and Technology* 61.12 (Dec. 2010), pp. 2544–2558. ISSN: 1532-2882.

[45] Catherine Triomphe. *Explaining 'Stop the Steal', Trump supporters' viral offensive to discredit the election*. en. Nov. 2020. URL: https://www.ctvnews.ca/world/america-votes/explaining-stop-the-steal-trump-supporters-viral-offensive-to-discredit-the-election-1.5177864 (visited on 03/16/2021).

[46] *Twitter is building 'Birdwatch,' a system to fight misinformation by adding more context to tweets*. en-US. URL: https://social.techcrunch.com/2020/10/02/twitter-is-building-birdwatch-a-system-to-fight-misinformation-by-adding-more-context-to-tweets/ (visited on 02/17/2021).

[47] Twitter Safety. *Improved account security during the 2020 US election*. en_us. Sept. 2020. URL: https://blog.twitter.com/en_us/topics/company/2020/Improved-Account-Security-2020-US-Election.html (visited on 02/18/2021).

[48] Twitter Safety. *Strengthening our approach to deliberate attempts to mislead voters*. en_us. Apr. 2019. URL: https://blog.twitter.com/en_us/topics/company/2019/strengthening-our-approach-to-deliberate-attempts-to-mislead-vot.html (visited on 02/18/2021).

[49] *Updating our approach to misleading information*. en_us. URL: https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html (visited on 02/18/2021).

[50] Soroush Vosoughi, Deb Roy, and Sinan Aral. "The spread of true and false news online". en. In: *Science* 359.6380 (Mar. 2018), pp. 1146–1151. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aap9559. URL: https://www.sciencemag.org/lookup/doi/10.1126/science.aap9559 (visited on 02/18/2021).

[51] Claire Wardle. "Misinformation has created a new world disorder". In: 321 (2019), pp. 88–93.

[52] D. Watts et al. "Tweet sentiment as proxy for political campaign momentum". In: *2016 IEEE International Conference on Big Data (Big Data)*. Dec. 2016, pp. 2475–2484. DOI: 10.1109/BigData.2016.7840885.

Appendix A

Data

A.1 Tweet Object: Extended Entities Expanded


```
▼ extended_entities {1}
  ▼ media [1]
    ▼ 0  {10}
        id   : 1257878957707452400
        id_str : 1257878957707452418
      ▼ indices [2]
          0  : 28
          1  : 51
        media_url : http://pbs.twimg.com/media/EXTiP1yXYAIOEyF.jpg
        media_url_https : https://pbs.twimg.com/media/EXTiP1yXYAIOEyF.jpg
        url  : https://t.co/O6tAvOwtmS
        display_url : pic.twitter.com/O6tAvOwtmS
        expanded_url : https://twitter.com/TheWiseMan11/status/1257878959573868544/photo/1
        type : photo
      ▼ sizes {4}
        ▼ thumb {3}
            w    : 150
            h    : 150
            resize : crop
        ▼ small {3}
            w    : 564
            h    : 680
            resize : fit
        ▼ large {3}
            w    : 674
            h    : 812
            resize : fit
        ▼ medium {3}
            w    : 674
            h    : 812
            resize : fit
```

A.2 Tweet Object: Entities Expanded


```
▾ entities {5}
  ▾ hashtags [0]
      (empty array)
  ▾ urls [0]
      (empty array)
  ▾ user_mentions [2]
    ▾ 0 {5}
        screen_name : realDonaldTrump
        name : Donald J. Trump
        id : 25073877
        id_str : 25073877
      ▾ indices [2]
          0 : 0
          1 : 16
    ▾ 1 {5}
        screen_name : honeywell
        name : Honeywell
        id : 64764487
        id_str : 64764487
      ▾ indices [2]
          0 : 17
          1 : 27
  ▾ symbols [0]
      (empty array)
  ▾ media [1]
    ▾ 0 {10}
        id : 1257878957707452400
        id_str : 1257878957707452418
      ▾ indices [2]
          0 : 28
          1 : 51
        media_url : http://pbs.twimg.com/media/EXTiP1yXYAIOEyF.jpg
        media_url_https : https://pbs.twimg.com/media/EXTiP1yXYAIOEyF.jpg
        url : https://t.co/O6tAvOwtmS
        display_url : pic.twitter.com/O6tAvOwtmS
        expanded_url : https://twitter.com/TheWiseMan11/status/1257878959573868544/photo/1
        type : photo
      ▾ sizes {4}
        ▾ thumb {3}
            w : 150
            h : 150
            resize : crop
        ▾ small {3}
            w : 564
            h : 680
            resize : fit
        ▾ large {3}
            w : 674
            h : 812
            resize : fit
        ▾ medium {3}
            w : 674
            h : 812
            resize : fit
```

A.3 User Object

▼ user {38}
id : 342972711
id_str : 342972711
name : Mark
screen_name : TheWiseMan11
location : Virginia, USA
url : null
description : null
translator_type : none
protected : false
verified : false
followers_count : 40
friends_count : 92
listed_count : 0
favourites_count : 1094
statuses_count : 5206
created_at : Tue Jul 26 21:33:20 +0000 2011
utc_offset : null
time_zone : null
geo_enabled : false
lang : null
contributors_enabled : false
is_translator : false
profile_background_color : FF6699
profile_background_image_url : http://abs.twimg.com/images/themes/theme11/bg.gif
profile_background_image_url_https : https://abs.twimg.com/images/themes/theme11/bg.gif
profile_background_tile : true
profile_link_color : B40B43
profile_sidebar_border_color : CC3366
profile_sidebar_fill_color : E5507E
profile_text_color : 362720
profile_use_background_image : true
profile_image_url : http://pbs.twimg.com/profile_images/1248356921305894916/Oi3Y-cnU_normal.jpg
profile_image_url_https : https://pbs.twimg.com/profile_images/1248356921305894916/Oi3Y-cnU_normal.jpg
default_profile : false
default_profile_image : false
following : null
follow_request_sent : null
notifications : null

A.4 List of Twitter Accounts for Filtered Stream

HalRogersKY5	TammyLeitnerNBC	MarioDB	pwgavin	MarkwayneMullin	jwpetersNYT
DrBurgessTX	PeterAlexander	repdonyoung	Dharapak	RepSarbanes	Robillard
RepTiffany	MorganRadford	RepMaxineWaters	asymmetricinfo	timburchett	freedlander
RepMikeGarcia	Xuan_Thai	BuckForColorado	terencesamuel	LoisFrankel	l_whittington
RepDanBishop	RichardEngel	ReElectCohen	politicalwire	ColinAllredTX	lauraelizdavis
RepGregMurphy	michelledubert	DebbieLesko	michaelroston	SusanBrooksIN	EmmaVOBrian
RepFredKeller	SteveKornacki	RepMikeRogersAL	ChadPergram	CollinsforGA	DaviSusan
MikeGarcia2020	jonallendc	RepBrianHiggins	edhenry	JaredHuffman	MarkLeibovich
RepRossSpano	AFrancoTX	RepJoseSerrano	ByronTau	GeorgeHolding	JournoMaggie
RepJimBaird	MCappetta	repgregwalden	ChrisStirewalt	RepAlGreenTX	cam_joseph
RepHagedorn	GadiNBC	FrankPallone	meredithshiner	BobbyScott4VA3	DavidWastell
RepCasten	SteveLouie	jahimes	mikememoli	DavidKustoff	nielslesniewski
CongressmanJVD	gabegutierrez	AyannaPressley	TomBevanRCP	MarkTakano	hillhulse
RepGregSteube	bfouhy	BradSherman	davecatanese	MikeSimpson4ID	feliciasonmez
RepKendraHorn	AliciaHastey	RepSamGraves	adamsmithtimes	bradwenstrup	scottwongDC
RepChipRoy	carolelee	RepAdamSchiff	JBennet	RepLindaSanchez	AaronBlake
RepChuyGarcia	MHause	repjohnlewis	jodikantor	RepStephenLynch	NKingofDC
RepMichaelWaltz	CalNBC	RepRaulGrijalva	jbendery	MarkPocan	Jenna_Kastan
RepLucyMcBath	AndySalsman	RepDianaDeGette	AnnGerhart	Call_Me_Dutch	mollyesque
RepGregPence	PattersonNBC	HarleyRouda	1PatriciaMurphy	RepBobbyRush	eliotnelson
Ilhan	AliVelshi	tndesjarlais	hramer	GT_TeamGT	KevinBohnCNN
RepJohnJoyce	AriMelber	PatrickMcHenry	KThomasDC	RepThompson	CarolC
RepJahanaHayes	RaineyTime	JasonSmithMO	JenniferJJacobs	Nanette4CA	TCurry_Himself
RepEdCase	tperry518	TulsiGabbard	michaelpfalcone	CongPalazzo	peterbakernyt
RepCarolMiller	Bidar411	RepMcCaul	christinawilkie	RepAdrianSmith	petermaer
RepJohnRose	jackturmanIII	spmaloney	CNNValencia	Buddy_Carter	TexasTribAbby
RepFinkenauer	kabir_here	CurtisUT	bennyjohnson	ConawayTX11	amychozick
RepMichaelGuest	cbszak	RepGusBilirakis	nationaljournal	ScottPetersSD	TheOnlyArcher
RepKirkpatrick	camiloreports	congbillposey	JFKucinich	Deb4CongressNM	mikeviqueira
RepKatiePorter	BoKnowsNews	RepSpeier	asheinin	RepWebster	ShanTravisNews
RepGilCisneros	aaronlarnavarro	TomGravesGA14	nycjim	Schneider4IL10	rachelrhartman
RepBrindisi	NicoleSganga	RepHankJohnson	rickdunham	Ann_Kirkpatrick	joehagansays
RepDMP	rikkijklieman	AndyKimNJ	KateNocera	RepZoeLofgren	IsaacDovere
RepMalinowski	ThisAlexTin	janschakowsky	stiles	RepRichmond	LisaDNews
RepDanCrenshaw	jaysunsilver	RepKenMarchant	amanbatheja	Andre4Congress	kenvogel
RepGolden	byKateSmith	SteveChabot	vplus	StenyHoyer	RebeccaRKaplan
RepGregStanton	Olivia_Gazis	MichaelJCloud	shearm	JohnSarbanes	alexanderbolton
RepTjCox	saraecook	davidcicilline	jaketapper	RepMaloney	ScottFConroy
RepCindyAxne	SalvantoCBS	KatherineClark	AHMalcolm	CheriBustos	DanaBashCNN
RepRonWright	bfmitchell	TomColeOK04	nytpolitics	RepJoeWilson	russellberman
RepJoshHarder	CaraKorte	RepGregoryMeeks	RalstonReports	RepTomReed	henrycjjackson
RepTorresSmall	ellee_watson	RepGwenMoore	CillizzaCNN	RepPeterDeFazio	DavidMDrucker
RepDebHaaland	LaCraiMitchell	KenCalvert	ktumulty	SanfordBishop	DanLothianTV
RepSylviaGarcia	adam_brew	repjimcooper	ron_fournier	RepAnnaEshoo	CarlCannon
RepPressley	PaulaReidCBS	replouiegohmert	mikeallen	RepBarbaraLee	deirdrekwalsh
RepMeuser	FaceTheNation	LeaderHoyer	georgecondon	sethmoulton	PostScottWilson
RepDavidTrone	Elaine_Quijano	gregsteube	j_strong	StevenHorsford	PoliticoCharlie
RepSherrill	weijia	VanTaylorTX	nbj914	RepJimCosta	laurenonthehill
RepUnderwood	CHueyBurns	auctnr1	JNSmall	HerreraBeutler	JessicaTaylor
RepDavids	antjuansea	DavidSchweikert	emilyrs	RepTipton	maggieNYT
RepArmstrongND	margbrennan	Andy_Levin	llerer	RepFrankLucas	ReutersZengerle
RepRiggleman	MajorCBS	tedlieu	khennessey	RepMcGovern	jasondhorowitz
repdelgado	markknoller	Duncan4Congress	marcambinder	RepYvetteClarke	sppeoples
RepMarkGreen	stevenportnoy	MarcVeasey	PoliticsReid	RepGosar	michaelcrowley
RepKimSchrier	kathrynw5	RepPerlmutter	OKnox	RepMcKinley	nataliejennings
Rep_Watkins	nancycordes	RepSchrader	jmsummers	RepJeffDuncan	jeffzeleny
RepElaineLuria	NorahODonnell	VernBuchanan	tbridis	RepKarenBass	mimms
RepAndyLevin	mollyhooper	RepPeteOlson	PoliticalTicker	RepWalberg	steinhauserNH1

RepAngieCraig	alanhe	RepKevinBrady	standupkid	RepHartzler	JessicaYellin
RepCunningham	tonydokoupil	DougLaMalfa	johnroconnor	RepSteveChabot	RyanLizza
RepJasonCrow	edokeefe	GOPLeader	thegarance	RepBillFlores	JeffYoung
RepLoriTrahan	Lynda_Tran	repbenraylujan	aliciacohn	RepMoBrooks	alexburnsNYT
RepTimmons	lexharis	krhern	kseelye	NormaJTorres	stevenjay
RepRashida	LeslieSanchez	CongressmanGT	juliemason	DesJarlaisTN04	apalmerdc
RepAOC	DanPatterson	JimLangevin	ethanklapper	RepChuck	NikolenDC
RepSusieLee	CBSNJ2012	Tom_Suozzi	PeterHamby	RepBobGibbs	mmcauliff
RepAndyKimNJ	RodneyCBSNews	JohnGaramendi	palafo	RepLarryBucshon	DanielStrauss4
RepSpanberger	SarahHuisenga	mlfudge	Shawna	AustinScottGA08	Redistrict
RepJoeNeguse	CNHorn	JeffFortenberry	MatthewKeysLive	rep_stevewomack	learyreports
RepRussFulcher	JanCBS	RepScottPerry	alt_uscbp	RepMGriffith	Maddie_Marshall
RepMaxRose	WCronkite	RepSires	ALT_uscis	RepWilson	Hadas_Gold
RepSlotkin	hagerhoo	dankildee	alt_HUD_FHEO	RepHuizenga	camanpour
RepColinAllred	AmlingAM	RepPeteKing	AltNatSecAgency	justinamash	edhornick
RepHaleyStevens	CharlieKayeNews	Jim_Jordan	AltDept_of_VA	RepRickCrawford	deucecrew
RepFletcher	ttgaughan	JudgeCarter	RoguePOTUSStaff	USRepKeating	joshgerstein
RepPeteStauber	BretHovell	AdamKinzinger	AltStateDpt	JoaquinCastrotx	pemalevy
RepEscobar	CBSNewsRadio	cathymcmorris	ALT_DOJ	RepBrendanBoyle	danbalz
RepBryanSteil	ishamdc	dinatitus	AltHomelandSec	RepDavid	mweinger
RepHarley	LuluTV	RichHudson	AltUSDA_ARS	RepMarthaRoby	PhilipRucker
RepVanTaylor	ebgreenfield	virginiafoxx	AltUS_ARC	NitaLowey	TheLloydGrove
RepDustyJohnson	LindseyBoerma	repblumenauer	AltPublicLands	RepKinzinger	amieparnes
RepBenCline	CBSPolitics	leezeldin	alt_fda	AnnLWagner	ASimendinger
RepMikeLevin	robhendin	tommcclintock	RogueNASA	gregstantonaz	Taylor_West
RepSusanWild	jacksonjk	SpeakerPelosi	ActualEPAFacts	RepBillJohnson	MichaelCBender
repkevinhern	johndickerson	michaelcburgess	altusda	RepSteveStivers	shushwalshe
RepChrisPappas	stevechaggaris	RepShimkus	altUSEPA	RepCuellar	jennifereduffy
RepMGS	Mosheh	kevinomccarthy	AltMtRainier	RepAndreCarson	SavannahGuthrie
RepJoeMorelle	jdickerson	boblatta	olympicRogueNPS	RepBenMcAdams	amyewalter
RepDeanPhillips	DanRaviv	RobWittman	altNOAA	DeFazio4Oregon	AliRogin
RepShalala	thecyclemsnbc	TeamJoeWilson	AngryNatlPark	KingforCongress	anitakumar01
RepAGonzalez	ReliableSources	chelliepingree	WHPublicPool	RepVisclosky	nathanlgonzales
RepHoulahan	MHPshow	Jim_Banks	ChanelRion	RepJudyChu	Jordanfabian
RepCloudTX	ThisWeekABC	MikeLevin	BenKennedyTV	PramilaJayapal	emtitus
RepBalderson	NowWithAlex	jackiewalorski	TocRadio	RepTomGraves	JeffreyGoldberg
Lancegooden	TheFive	RepTimRyan	Tara_Mckelvey	Foster4Congress	JillDLawrence
RepTimBurchett	BashirLive	PeteOlson	MaraLiasson	WhipClyburn	GrahamDavidA
RepWexton	davidaxelrod	SenatorLoeffler	catherine_lucey	JoeCourtneyCT	brikeilarcnn
CaptClayHiggins	David_Gergen	Risch4Idaho	svdate	EspaillatNY	DrewHampshire
RepDLesko	AnnCoulter	SenHawleyPress	rebeccaballhaus	votegregwalden	AnjeanetteDamon
JahanaHayesCT	MarkHalperin	SenatorBraun	GabbyOrr_	jerrymcnerney	brookebrower
ZoeLofgren	Lawrence	SenatorRomney	JDiamond1	RussFulcher	timgrieve
BryanSteil	VanJones68	SenBillCassidy	sarahcwestwood	LloydDoggettTX	StevenTDennis
RepConorLamb	stefcutter	SenHydeSmith	mawilner	Matsui4Congress	mikebarnicle
votemorelle	brithume	SenDougJones	vmsalama	repdavidscott	jeneps
JohnJoyceForPA	megynkelly	SenTinaSmith	ElizLanders	RepEBJ	marincogan
Armstrong_ND	CapehartJ	gillibrandny	JonLemire	NydiaVelazquez	TroyKinsey
marygayscanlon	edshow	SenJackyRosen	kaitlancollins	RepEliotEngel	hollybdc
VoteTimmons	stephenfhayes	SenJohnKennedy	KFaulders	repcleaver	samyoungman
GuestCongress	donnabrazile	SenCortezMasto	ToluseO	RepCohen	AmyAHarder
sharicedavids	Eugene_Robinson	SenKamalaHarris	cherylbolen	BobbyScott	mkraju
DanLipinski	alexwagner	RosenforNevada	jeffmason1	RepKayGranger	Elahelzadi
ElaineLuriaVA	IngrahamAngle	CatherineForNV	RichLatendresse	gracenapolitano	billkeller2014
lucymcbath	JoyAnnReid	VanHollenForMD	CostantiniWHear	larsenrick	ronlieber
VanDrewForNJ	HardballChris	SenMcSallyAZ	jcartillier	RepAlGreen	merica
MadeleineDean	finneyk	SenThomTillis	Yamiche	KindforCongress	davidjoachim
RepJohnCurtis	kilmeade	SenSasse	EliStokols	RepDavidEPrice	StephanieEbbert
DanCrenshawTX	seanhannity	SenatorRounds	JakeTurx	LangevinForRI	MattVas
LindaTSanchez	marklevinshow	SenDanSullivan	margarettalev	JohnLarsonCT	Peggynoonannyc
LoriTrahanMA	AriFleischer	sendavidperdue	AthenaCNN	RepLloydDoggett	stevebousquet
Meuser4Congress	Judgenap	SenJoniErnst	KateBennett_DC	PaulTonko	Tyrangiel
golden4congress	krauthammer	SenMurphyOffice	KatiePavlich	UptonForAllOfUs	ErinMcPike
anthonygonzalez	DanaPerino	robportman	ksenijapavlovic	RepMarciaFudge	gabrielsherman
JohnRoseforTN	LouDobbs	HawleyMO	kevincorke	TimWalberg	bethreinhard
MaxRose4NY	BillOReilly	SenBooker	ChrisMegerian	CedricRichmond	AshleyRParker
GregPenceIN	SteveDoocy	MikeLeeforUtah	kristindonnelly	DustyJohnson	shiracenter
PeteStauber	krystalball	chuckschumer	djusatoday	rosadelauro	steveholland1
SpanbergerVA07	KevinMaddenDC	braun4indiana	linda_feldmann	TeamCalvert	ktummarello
TJCoxCongress	JoeNBC	BenSasse	ShaneGoldmacher	AOC	MatthewDalyWDC
HornForCongress	murphymike	Perduesenate	betsy_klein	RepTedDeutch	Inky_Dan
RepRalphNorman	newtgingrich	joniernst	ArdenFarhi	WebsterCongress	megancarpentier
GilCisnerosCA	davidgregory	SenBrianSchatz	DaveBoyer	standwithraul	KevinKBrennan
LaurenUnderwood	billmaher	senatemajldr	KimberlyHalkett	AdamSchiff	juliebosman
DrPaulGosar	secupp	EdMarkey	Carrasquillo	RepLynch	ColleenMNelson
Axne4Congress	nickgillespie	MartinHeinrich	JenniferWishon	TeamCMR	toddzwillich
DrKimSchrier	FoxNewsSunday	SenatorRisch	johnrobertsFox	MorganGriffith	kendratypes
ElissaSlotkin	KarlRove	SenatorSinema	colvinj	JodyHice	BenLeubsdorf

Lizzie4Congress	BretBaier	SenatorBaldwin	meridithmcgraw	Steve4Kansas	jpaceDC
ConorLambPA	ezraklein	SenTedCruz	finnygo	McGovernMA	ajjaffe
RepRonEstes	MHarrisPerry	SenatorFischer	seungminkim	DavidRouzer	KObradovich
JasonCrowCO6	morningmika	SenAngusKing	hunterw	ElectConnolly	jamiedupree
HoulahanForPa	CNNSotu	SenDuckworth	AprilDRyan	RepFredUpton	PaulLewis
RepClayHiggins	alexcast	SenKevinCramer	W7VOA	Harris4Congress	stevebenen
DelgadoforNY19	maddow	RoundsforSenate	FrancoOrdonez	repdinatitus	AmyEGardner
RepRutherfordFL	rolandsmartin	SenWarren	charliespiering	DerekKilmer	suellentrop
RepMikeJohnson	greta	SenTomCotton	wpjenna	CarolynBMaloney	JamieGrayTweets
JoeCunninghamSC	michellemalkin	SenatorHassan	tamarakeithNPR	StevenPalazzo	davidshepardson
RepAnthonyBrown	foxandfriends	SheldonforRI	thomaswburr	DWStweets	jkirchick
RepDonBacon	GovMikeHuckabee	MikeCrapo	lynnsweet	collinpeterson	Alex_Roarty
RepMattGaetz	andersoncooper	PatRoberts	katierogers	barrforcongress	1bobcohn
RepDarrenSoto	DavidCornDC	AngusforMaine	CateMartel	brettguthrie	AndrewRafferty
RepGonzalez	hughhewitt	SenatorTester	KimberlyEAtkins	MikeKellyforPA	jameshohmann
Denver4VA	howardfineman	SenJackReed	jonathanvswan	KarenBassTweets	lucytimes
RepAlLawsonJr	DLoesch	MarthaMcSally	AmandaBecker	RepRonKind	AdamBKushner
RepTedBudd	chrislhayes	SenFeinstein	mj_lee	womack4congress	JoshNBCNews
RepEspaillat	MattMackowiak	SenCoonsOffice	ByChrisBrennan	EliseStefanik	CPHeinze
RepLBR	DrZoeHyde	LindseyGrahamSC	BryanLlenas	Ed4Colorado	jaconi
RepBarragan	gmleunghku	MarkWarnerVA	sarahmccammon	Liz_Cheney	ElizaRules
RepLizCheney	BogochIsaac	BobMenendezNJ	hunterschwarz	JaimeForUSRep	DavidNakamura
RepTomSuozzi	RealPressSecBot	RoyBluntMO	darrensands	ReElectHank	pmdfoster
RepAndyBiggsAZ	PreetBharara	SenJohnHoeven	cptime	Flores4Congress	abettel
RepBrianFitz	EmmaVigeland	TomCarperforDE	SecretsBedard	WilliamKeating	caphilltrish
RepRoKhanna	ByYourLogic	DianneFeinstein	joegarofoli	latta4congress	Acosta
RepArrington	PpollingNumbers	Maggie_Hassan	bluestein	DrMarkGreen4TN	pdicarlocnn
RepMcEachin	danpfeiffer	ewarren	WesleyLowery	benraylujan	jdelreal
RepCarbajal	Conflicts	TomCottonAR	mateagold	RepDwightEvans	ArletteSaenz
RepJimBanks	elizabeth_joh	McConnellPress	thisisjorge	AnnMcLaneKuster	brown_alex_
RepRooney	pwnallthethings	TammyforIL	FOX411Country	RepGaramendi	DHBerman
RepCharlieCrist	VicBergerIV	TomUdallPress	AftertheBell	AlLawsonJr	POLITICO_Steve
RepDavidKustoff	kenklippenstein	DebFischerNE	SundayHousecall	RepSchakowsky	streitfeldcnn
RepLouCorrea	ebruenig	DickDurbin	ffweekend	RepJoeCourtney	jonathanweisman
RepGallagher	PoliticsWolf	ScottforFlorida	foxnewshealth	RepPaulTonko	HotlineJosh
DrNealDunnFL2	adamjohnsonNYC	SenJohnThune	FoxBizTech	BennieGThompson	MelindaKCMO
RepJayapal	dwdavison	GrassleyPress	FoxNewsTravel	KathleenRice	mckaycoppins
RepJoshG	HeerJeet	PattyMurray	FoxBizVideo	JohnsonLeads	ZekeJMiller
RepJackBergman	meaganmday	WydenPress	fbsmallbiz	DrLarryBucshon	LarrySabato
CongressmanRaja	KuzmanovicA	SenatorEnzi	foxnewscareers	JacksonLeeTX18	SusanPage
RepBrianMast	GregJKrieg	MarshaBlackburn	foxnewsvideo	GerryConnolly	bonkapp
RepTrey	AlxThomp	SenBlumenthal	ureport	Robert_Aderholt	JakeSherman
RepPaulMitchell	KFILE	BenCardinforMD	FoxNewsPhotos	CongMikeSimpson	DavidChalian
Malinowski	Laurie_Garrett	SenatorWicker	FoxNewsLive	MikeRogersCamp	LukeRussert
RepOHalleran	LukewSavage	cindyhydesmith	GeraldoOnFox	BillPascrell	wuerker
RepRaskin	Drew_Hammill	RonWyden	TomShillueShow	WillHurd	AlexPappas
RepDrewFerguson	elizacollins1	SenatorCarper	ForbesonFox	tomemmer	perrybaconjr
RepValDemings	ddale8	SenatorDurbin	HuckabeeShow	Sewell4Congress	BFischerMartin
RepJimmyPanetta	WideAsleepNima	SenatorLeahy	America_Live	TeamRoby	JohnBerman
RepStephMurphy	Bencjacobs	SenWhitehouse	foxnewslatino	SeanCasten	DomenicoNPR
GoDutchForMD	chactivist	DougJones	SpecialReport	TedDeutch	emilykpierce
KathyCastorFL	jonlovett	SenGaryPeters	Fox411	DennyHeck	rdmurphy
VickyH4Congress	david_j_roth	SenCoryGardner	FoxNewsOpinion	TomReedCongress	mviser
RepMarshall	ChuckRocha	Sen_JoeManchin	FoxNewsLife	RajaForCongress	dannowicki
JackBergman_MI1	Birdyword	SenToddYoung	FoxNewsTech	RalphNorman	AWMooneyDC
DebbieforFL	jonfavs	SenRonJohnson	foxheadlines	foxxforcongress	DonovanSlack
voteloucorrea	PykeA	SenRubioPress	FoxNewsAutos	stevestivers	mkady
SMurphyCongress	abbydphillip	SenatorLankford	FoxBusiness	RepAdamSmith	RebeccaShabad
WarrenDavidson	woodruffbets	SenatorBennet	foxnation	mattgaetz	elizthompsn
DonnaShalala	lib_crusher	JeffMerkley	foxnewsalert	Higgins4WNY	GlennThrush
RonEstesKS	oneunderscore__	SenToomey	foxnewspolitics	RepMikeQuigley	MarkMBenjamin
ReElectNydia	ditzkoff	SenatorTimScott	foxnewsradio	DevinNunes	leslielarson
buddforcongress	joshjame	RandPaul	OutnumberedFNC	BrendaLLawrence	devindwyer
RobBishopUT	juliaioffe	BernieSanders	HQFOX	DrPhilRoe	moody
andybiggs4az	AdamWollner	BurrForSenate	JonScottFNC	RepMikeTurner	mpoindc
BrianFitzUSA	tparti	SenJohnBarrasso	AmericaNewsroom	electadamsmith	CahnEmily
davidjtrone	megtirrell	SenCapito	TeamCavuto	RepJohnLarson	GreenwireJeremy
MikeJohnson	bcburden	timkaine	HappeningNow	RepMcClintock	reidepstein
LouieGohmertTX1	brianklaas	SenBobCasey	JedediahBila	chuck4congress	jmartNYT
Vote_Davidson	jdawsey1	mazieforhawaii	JamesRosenTV	RubenGallego	SaraMurray
JohnMShimkus	spettypi	CapitoforWV	marthamaccallum	BenMcAdams	PrestonCNN
FrancisRooney	AbbyMartin	MurrayCampaign	kimguilfoyle	SteveKinglA	JesseRodriguez
CongBoyle	nycsouthpaw	LisaForSenate	ehasselbeck	ReElectWittman	michaelscherer
JimmyPanetta	cushbomb	ChrisMurphyCT	JennaLeeUSA	JuliaBrownley	RebeccaBuck
LloydSmuckerPA	sahilkapur	GaryPeters	ANHQDC	BeraForCongress	scontorno
GReschenthaler	GeoffRBennett	RonJohnsonWI	BillHemmer	BradleyByrne	katiezez
DwightEvansPA	rubycramer	SenRickScott	ShannonBream	aguilarpete	joshhafner
LisaBRochester	shujaxhaider	SenatorCantwell	HARRISFAULKNER	RepSmucker	jolingkent

RepLaHood	timkmak	JoeManchinWV	JustinFishelABC	SuzanDelBene	SteveSebelius
Mike_CA05	RobertMackey	Boozman4AR	MichelleMacTV	RepJerryNadler	Max_Fisher
JodeyArrington	ryangrim	SenatorShaheen	RobertGBeckel	Cline4Virginia	JoshuaGreen
DunnCampaign	cmarinucci	SenatorCardin	OliverLNorth	deanbphillips	adambeam
VoteForTrey	dsamuelsohn	Bob_Casey	caseystegall	USRepMikeDoyle	Philip_Elliott
GregForMontana	virgiltexas	maziehirono	pattiannbrowne	CharlieCrist	MaeveReston
TomTiffanyWI	katienotopoulos	SenMikeLee	GeraldoRivera	vgescobar	ryanbeckwith
RepTrentKelly	lachlan	tammybaldwin	MollyLineFNC	RobinLynneKelly	its_amt
TeamVisclosky	Taniel	votetimscott	HeatherChilders	DorisMatsui	felixsalmon
carbajalsalud	realdanlyons	SenAlexander	EricShawnTV	WomenforTrump	OANN
BrianMastFL	AriRabinHavt	SenStabenow	JohnStossel	GovStitt	melbournecoal
MaxineWaters	mattyglesias	SenPatRoberts	MikeEmanuelFox	IAGovernor	haeyoun
DonaldNorcross	gcaw	SenGillibrand	JaniceDean	AsaforAR	jennymedina
DonJBacon	jbouie	VoteMarsha	drmannyonFOX	BillLeeTN	johannabarr
ElectLaHood	daveweigel	SenatorTomUdall	pdoocy	Tony4WI	tannercurtis
teammoulton	alexismadrigal	ToddYoungIN	tracegallagher	GovChrisSununu	AlexaCorse
KellyforMS01	benyt	BillCassidy	RickFolbaum	NC_Governor	aduehren
RepTedLieu	tomgara	MittRomney	JudgeJeanine	GovofCO	CMcBinDC
RepPeteAguilar	QuinHillyer	brianschatz	greggutfeld	GovHolcomb	brkend
RepGrothman	FakeJenRubin	TinaSmithMN	LisWiehl	PhilMurphyNJ	TheAmaraReport
CongressmanHice	jimantle	MichaelBennet	MikeTobinFox	SteveSisolak	JeanneCummings
RepBuddyCarter	PJORourke	SenSherrodBrown	rickreichmuth	OregonGovBrown	willmauldin
RepKathleenRice	dailyherring	DrRandPaul	aacuna1	AndyBeshearKY	kristinapet
RepDebDingell	MattNaugle	MariaCantwell	GreggJarrett	GovLarryHogan	siobhanehughes
RepBonnie	johngizzi	amyklobuchar	LelandVittert	GovernorTomWolf	poverberg
RepDeSaulnier	WilliamJBennett	KamalaHarris	JulieBanderas	GovHawaii	KateMilani
RepJohnKatko	JazzShaw	KLoeffler	livedesk	GovRicketts	johndmckinnon
RepRubenGallego	DanFosterType	SenSanders	RickLeventhal	andrewcuomo	JessBravin
RepMarkWalker	JavManjarres	SenJeffMerkley	adamhousley	RoyCooperNC	randyyeipWSJ
RepBost	costareports	kevincramer	lauraingle	TomWolfPA	janetadamy
RepAlexMooney	KatrinaTrinko	JohnKennedyLA	courtneyfriel	JohnBelEdwards	wstrobel
RepRickAllen	JonahDispatch	johnthune	jonathanserrie	GovInslee	mgordonwsj
HurdOnTheHill	AmandaCarey1	PatToomey	ClaytonMorris	GovRonDeSantis	lindsaywise
RepAbraham	VinceCoglianese	SherrodBrown	andylevy	WVGovernor	TimJHanrahan
RepDonBeyer	Sam_Schulman	tedcruz	wolfblitzer	GovPhilScott	paulwsj
RepStefanik	ByronYork	ThomTillis	Kuhn	DougBurgum	WSJshutdown
RepFrenchHill	joshuamercer	SenAmyKlobuchar	campbell_brown	RonDeSantisFL	aaronzitner
RepGarretGraves	RichLowry	LamarAlexander	SamFeistCNN	Michelle4NM	TroyWSJ
RepNewhouse	RichardViguerie	SenMarkey	JohnKingCNN	LauraKellyKS	GeraldFSeib
RepBrianBabin	EconTalker	RoyBlunt	MerylGovernski	RalphNortham	anthonydb
RepAdams	DeanClancy	SenatorBurr	GrantMegan	LarryHogan	WSJPolitics
RepTomEmmer	barneykeller	SenShelby	CandyCrowley	AsaHutchinson	DrewFitzGerald
RepLoudermilk	arthurbrooks	CoryGardner	SuzanneMalveaux	GovRaimondo	aviswanatha
RepLawrence	McCormackJohn	kyrstensinema	sinderbrand	JohnCarneyDE	Dchinni
RepKenBuck	BernardGoldberg	InhofePress	tomforemancnn	govkristinoem	nataliewsj
USRepGaryPalmer	thegoldfarb	stabenow	stevebruskNews	GovEvers	DannyDougherty
RepWesterman	TonyFratto	SenatorCollins	CNNPolitics	NYGovCuomo	NaftaliBendavid
ChuyForCongress	TuckerCarlson	senrobportman	jrubin	PhilBryantMS	MsVanessaW
DNI_Ratcliffe	ThatWinkler	SenatorMenendez	CNNTonight	LouisianaGov	BlairGuild
RepLeeZeldin	emilymiller	JerryMoran	CNNOriginals	KristiNoem	WPJohnWagner
SteveScaliseGOP	MJGerson	ChrisVanHollen	NewDay	GinaRaimondo	scotusreporter
RepMoolenaar	Gundisalvus	lisamurkowski	TheLeadCNN	GovernorBullock	GlennKesslerWP
TomOHalleran	philipaklein	SenSchumer	DaniellaMicaela	GovernorVA	rachaelmbade
DonYoungAK	kathrynlopez	jontester	Michaela	GovWhitmer	chelsea_janes
LuetkemeyerB	allahpundit	CoryBooker	AlannaPetroff	ChrisSununu	CarolLeonnig
Tcardenas	AndrewBreitbart	marcorubio	EarlyStart	GregAbbott_TX	byamberphillips
TeamPelosi	kenblackwell	ChrisCoons	RachelSmolkin	Tim_Walz	Colbyltkowitz
RepHastingsFL	mattklewis	GrahamBlog	OutFrontCNN	mikeparson	terri_rupar
GrothmanforWI	EdMorrissey	JeanneShaheen	Eugene_Scott	GovHerbert	Fahrenthold
Norma4Congress	JimPethokoukis	JohnCornyn	kyletblaine	CharlieBakerMA	amybwang
ClyburnSC06	andyroth	SteveDaines	MichLKosinski	BrianKempGA	marycjordan
RepRobWoodall	guypbenson	RogerWicker	FWhitfield	JBPritzker	cathleendecker
RepJimmyGomez	NolteNC	ChuckGrassley	CristinaAlesci	GovernorLittle	newsbysamuels
DebDingell	TPCarney	MarkWarner	CNNOpinion	CAgovernor	CleveWootson
DonNorcross4NJ	MichaelRWarren	JimInhofe	donie	KimReynoldsIA	AnnieLinskey
DonBeyerVA	johnhawkinsrwn	JohnBoozman	AC360	MikeDeWine	ericawerner
Palmer4Alabama	jimgeraghty	SBAJovita	PiersMorganLive	NedLamont	Krissah30
RepKClark	CalebHowe	DHS_Wolf	AlisonKosik	Mark_Gordon_WY	peterwallsten
BostForCongress	Heminator	SecGeneScalia	MaxFosterCNN	JayInslee	sandhyawp
RepByrne	dmataconis	USAmbUN	VictorBlackwell	dougducey	thamburger
Babin4Congress	jpodhoretz	EsperDoD	sunny	GovernorKayIvey	myhlee
JenniferWexton	MikeRiggs	SecWilkie	JuanCarlosCNN	GovAbbott	mikedebonis
jdanbishop	JamesPPinkerton	SecPompeo	joejohnscnn	GovMikeDeWine	agearan
RepGuthrie	MediaLizzy	RussVought45	barbarastarrcnn	MassGovernor	PostPolls
USRepKCastor	BrianFaughnan	SecAzar	BrianToddCNN	henrymcmaster	EmGusk
RepBlaine	mkhammer	SecBrouillette	GaryTuchmanCNN	jaredpolis	postpolitics
DarrenSoto	EWErickson	MickMulvaneyOMB	evanperez	GavinNewsom	pbump
JimHagedornMN	PaulBegala	USTradeRep	SaimaMohsin	karentravers	loganclarkhall

A.4 List of Twitter Accounts for Filtered Stream

petekingcpg	Jimmyspolitics	SecretarySonny	IshaSesay	MarthaRaddatz	RealCandaceO
RepJasonSmith	ThisBowers	SecretaryRoss	bencnn	jparkABC	RealSaavedra
AlmaforCongress	joshharkinson	OMBPress	CNNPR	tarapalmeri	KensingtonRoyal
WestermanAR	KeithOlbermann	EPAAWheeler	PoppyHarlowCNN	rickklein	BillKristol
MikieSherrill	KatrinaNation	BetsyDeVosED	ShimonPro	WrightUps	stuartpstevens
MikeforWI	thinkprogress	SecretaryCarson	SunlenSerfaty	GStephanopoulos	HashtagGriswold
JoeNeguse	AndyKroll	SecElaineChao	ericbradner	ABCPolitics	continetti
RepBeatty	markos	WhiteHouse	DrewGriffinCNN	jonkarl	elianayjohnson
MooneyforWV	drvox	SecBernhardt	BrookeBCNN	voxdotcom	PRyan
USRepLong	jonathanalter	Cabinet	sarasidnerCNN	qz	TianaTheFirst
PeterWelch	CitizenCohn	VP	NatSecCNN	verge	DouthatNYT
AnthonyBrownMD4	joshtpm	stevenmnuchin1	natemcdermott	storyfulpro	jwgop
RepRobinKelly	CharlesMBlow	CIA	FareedZakaria	AP	xan_desanctis
RepChrisSmith	DavidShuster	USDOT	Kevinliptakcnn	TheAtlantic	RameshPonnuru
pallonefornj	Milbank	USUN	vausecnn	people	JRubinBlogger
Abby4Iowa	peterdaou	ENERGY	holmescnn	60Minutes	alexis_levinson
RepJuanVargas	Travis_Waldron	SBAgov	CNNSitRoom	ABC11_WTVD	NoahCRothman
chiproytx	AdamSerwer	USTreasury	AnaCabrera	nypost	SethAMandel
RepBrownley	ggreenwald	CommerceGov	HalaGorani	ABC7NY	RichardGrenell
RogerMarshallMD	emptywheel	DeptVetAffairs	NicRobertsonCNN	ABC30	SonnyBunch
RepCardenas	AriBerman	Interior	fpleitgenCNN	latimes	DonaldJTrumpJr
SteveScalise	aterkel	TheJusticeDept	ChrisCuomo	thedailybeast	baseballcrank
repmarkpocan	Atrios	DeptofDefense	klustout	USATODAY	MeghanMcCain
RepDonaldPayne	cbellantoni	USDA	edlavaCNN	Slate	Reince
RepDWStweets	smkeyes	ODNIgov	jimsciutto	CBSNews	SenJohnMcCain
MacThornberry	samstein	HHSGov	jasoncarrollcnn	propublica	GPollowitz
RepScottPeters	kate_sheppard	usedgov	kellywallacetv	HuffPost	RachaelBL
RepCheri	badler	USDOL	niaCNN	TIME	MZHemingway
RepRaulRuizMD	janehamsher	HUDgov	KateBolduan	NBCNews	benshapiro
RepAndyBarr	KirstenPowers	DHSgov	EliseLabott	politico	conncarroll
RepFilemonVela	billscher	EPA	RosaFlores	CNNgo	amandacarpenter
RepBillFoster	jeromearmstrong	StateDept	mikeroweworks	USArmy	reihan
RepRWilliams	ddayen	chicago	NoahGrayCNN	SkyNews	joelpollak
RepLoisFrankel	ryanjreilly	chicagosmayor	frates	YahooNews	SpeakerBoehner
RepPaulCook	JessicaValenti	houmayor	ryanobles	TheEconomist	petersuderman
RepVeasey	AmandaMarcotte	BaltCityHall	susancandiotti	BuzzFeed	IChotiner
SusanWBrooks	MarkHarrisNYC	MayorJenny	donlemon	NPR	TimothyNoah1
RepSeanMaloney	LarryKudlowShow	MayorGallego	lisaling	AJEnglish	AlecMacGillis
RepChrisStewart	DylanByers	LACity	CNNnewsroom	Newsweek	OsitaNwanevu
RepTedYoho	tanyahayre	mayor_margo	cnni	MSNBC	Yascha_Mounk
RepSchneider	PageSix	MayorGinther	BleacherReport	espn	noamscheiber
RepHuffman	mcdermott	MayorMemphis	aliarouzi	washingtonpost	jedshug
RepLaMalfa	mediabeat	PhillyMayor	PeteWilliamsNBC	FoxNews	michelleinbklyn
RepDennyHeck	francesmartel	LondonBreed	tombrokaw	mashable	mjs_DC
RepWalorski	ravisomaiya	CityofDetroit	JennaBushHager	TechCrunch	jonathanchait
TulsiPress	NoelSheppard	IndyMayorJoe	RonMott	nytimes	pareene
RepDougCollins	helena_andrews	MayorAdler	rickwtyler	CNN	JamilSmith
RepMullin	mlcalderone	ColumbusGov	kristendahlgren	BBCWorld	ClaraJeffery
RepDerekKilmer	LauraDonovanUA	CLTMayor	KMoDatelineNBC	cnnbrk	imillhiser
RepAnnieKuster	joepompeo	MayorBowser	tomcostellonbc	FiveThirtyEight	brianbeutler
RepHolding	JenniferPreston	MayorMikeDuggan	annenbcnews	CandaceSmith_	irin
RepTomRice	jonresnickUAS	ElPasoTXGov	trymainelee	CourtneycABC	kdrum
TXRandy14	AmyArgetsinger	CityofPhoenixAZ	jeffrossen	matthewjdowd	JillFilipovic
RepJoeKennedy	craigmelvin	MayorBetsyPrice	BillNeelyNBC	arobach	igorvolsky
RepLowenthal	Mediaite	nycgov	nbcMik	DianeSawyer	cmclymer
RepAnnWagner	HeardontheHill	MayorHancock	jdbalart	RobinRoberts	hilaryr
RepGraceMeng	carr2n	CityofSanJose	SheinelleJones	sswinkgma	page88
RepDanKildee	vulture	CityofJax	GarrettHaake	PierreTABC	owillis
RepMarkTakano	poniewozik	CityOfBoston	jonathan4ny	micahcohen	joanwalsh
RepLipinski	ChrisLicht	CityOfMemphis	joefryer	LinseyDavis	BarackObama
RepDelBene	TimJGraham	COSAGOV	mikeseidel	TomLlamasABC	MichelleObama
RodneyDavis	mattchew03	MayorPugh50	AymanM	DavidMuir	DrBiden
RepDaveJoyce	ErikWemple	marty_walsh	KeirSimmons	michaelstrahan	HillaryClinton
RepThomasMassie	keachhagey	HoustonTX	darabrown	CeciliaVega	AndrewYang
MarkMeadows	WillieGeist	sliccardo	janestreet	danabrams	TheDemocrats
RepAndyHarrisMD	nytmedia	CityofFortWorth	KellyCobiella	NateSilver538	HouseDemocrats
DeGette5280	colbyhall	Kevin_Faulconer	RichardLui	sarahaines	JohnKasich
RepBera	schneiderfactor	CityOfDallas	OARichardEngel	TerryMoran	SenateGOP
RepSwalwell	LaurenSko	Johnson4Dallas	AymanMSNBC	ABCNewsJason	GOP
MikeKellyPA	joseiswriting	austintexasgov	KasieDC	bworley	charliekirk11
RepRichHudson	joeljmiller	CityofSanDiego	DeadlineWH	niagarasquare	shannonrwatts
ABrindisiNY	HowardKurtz	Ron_Nirenberg	NBC_EarlyToday	NBCNewsTHINK	IvankaTrump
Rosa_DeLauro	DCist	lennycurry	NicolleDWallace	TheBeatWithAri	TiffanyATrump
JimPressOffice	FamousDC	CLTgov	SRuhle	NBCLeftField	EricTrump
CarolMillerWV	FishbowlDC	NYCMayor	jjramberg	NBCNewsBETTER	LaraLeaTrump
JamesComer	alexweprin	CityofDenver	ThomasARoberts	11thHour	POTUS
RepDavidRouzer	brianstelter	PhiladelphiaGov	VelshiMSNBC	amjoyshow	StephGrisham45
michaelgwaltz	MichaelHyatt	CityofSeattle	phillipmenaNBC	NBCOUT	KellyannePolls

Vote4Cleaver	tommyxtopher	MayorOfLA	TheRevAl	RobertsMSNBC	FLOTUS
RepDannyDavis	SteveKrak	washingtondc	AlexWitt	shiftmsnbc	PressSec
Gwen4Congress	tvnewser	sfgov	FrancesRivera	NBCBLK	SecondLady
IlhanMN	jackshafer	KYSecState	yasminv	waytooearly	QAnon_Report
RepCartwright	rachelsklar	oklahoma_sos	davidgura	NBCAsianAmerica	laurenboebert
RoKhanna	davidfolkenflik	NHSecretary	mega2e	MSNBCPhoto	DeAnna4Congress
McCaulforTexas	jeffjarvis	sodaksos	YLindaQiu	allinwithchris	theangiestanton
DaveJoyceOH14	NikkiSchwab	FLSecofState	nytstevek	TODAYPets	montaga
katieporteroc	Mediabistro	NEvnen	Ari_NYT	NBCLatino	mtgreenee
RossSpano	Megan	ltgovmeyer	npfandos	PoliticsNation	PeteHegseth
Troy_Balderson	TomDeFrank	IDSecOfState	aliciaparlap	TamronMSNBC	Blklivesmatter
SusieLeeNV	pkcapitol	ak_elections	Nate_Cohn	Morning_Joe	PPFA
CoopForCongress	robyoon	SecretaryWay	Tmgneff	TODAYshowHealth	AMarch4OurLives
garretgraves	SeanGHiggins	illinoissbe	EricLiptonNYT	NBCInvestigates	Heritage
standwithbarry	SarahBakerNBC	SecretaryOfMass	sangerkatz	thereidout	NRA
LacyClayMO1	AP_Politics	ceedepuertorico	adamnagourney	TheLastWord	theMRC
RepDLamborn	burgessev	SecretaryHobbs	GardinerHarris	SNFonNBC	BillClinton
DaveForIowa	MEPFuller	wvsosoffice	tackettdc	MSNBCYourBiz	SarahPalinUSA
JoyceBeatty	marymdouglas	Stapleton_MT	jakesNYT	MSNBCPR	StephenBannon
RepMarcyKaptur	jackhealyNYT	NMSecOfState	MikaylaBouchard	NBCNewsPR	GenFlynn
ChrisPappasNH	FoxReports	elections808	mattfleg	JDBmsnbc	MichaelCohen212
Grace4NY	justinsink	MESecOfState	stavernise	MaddowBlog	redsteeze
bradyfortexas	okayhenderson	WI_Elections	MarkMazzettiNYT	mitchellreports	Scaramucci
RepDean	JennyBlancoCNN	NVElect	SangerNYT	TODAY_Parents	gtconway3d
RepBonamici	JillAbramson	VermontSOS	Neil_Irwin	msnbcLockup	TTuberville
RepHalRogers	katiesmithallen	RISecState	BumillerNYT	TODAYFood	CaptMarkKelly
VoteYvette	anniekarni	EdBuchananWyo	jimtankersley	tvkatesnow	Hickenlooper
jamie_raskin	politicoroger	NVSOS	MarkLandler	MorningJoeguest	ossoff
wildforcongress	dmrcaucus	NCSBE	SteveEder	WPTV	ReverendWarnock
JackieSpeier	markzbarabak	MississippiSOS	jonathanmahler	NBCOlympics	GreenfieldIowa
RepBradWenstrup	NPRmelissablock	VotingInMass	SherylNYT	DatelineNBC	SaraGideon
BettyMcCollum04	PostRoz	md_sbe	dougmillsnyt	MeetThePress	JohnJamesMI
PascrellforNJ	mikiebarb	PAStateDept	BCAppelbaum	TheRoadWarriors	CalforNC
daveloebsack	AlexParkerDC	SecretaryLawson	nytgraphics	NBCNewsGraphics	mjhegar
WelchForVT	reidcherlin	VoteTexas	Jo_Becker	NBCNewsScience	GinaOrtizJones
BradSherman4SFV	NYTnickc	IowaSOS	jbrodernyt	NBCNewsHealth	TonyGonzales4TX
Eshoo4Congress	BobCusack	Louisiana_sos	dagnysalas	TODAYEnt	Carolyn4GA7
McHenryCampaign	JohnJHarwood	SOTSMerrill	CoralMDavenport	NBCSports	RichforGA
Marcy_Kaptur	nickconfessore	ElectionsUtah	AllMattNYT	NBCPolitics	Yvette4congress
RepRoybalAllard	EamonJavers	MichSoS	fstockman	NBCNewsWorld	stephaniebice
hakeem_jeffries	jimrutenberg	oregonelections	tripgabriel	NBCNightlyNews	BurgessOwens
reschenthaler	kkondik	KansasSOS	thomaskaplan	NBCNewsNow	candacefor24
CartwrightPA	StuPolitics	Vote4DC	emilybazelon	TODAYshow	Bethvanduyne
RepJeffries	dickstevenson	NYSDOS	mmcintire	BreakingNews	NancyMace
TimRyan	JamesPindell	MNSecofState	Stcrow	kwelkernbc	marwilliamson
RepCicilline	Emma_Dumain	MissouriSOS	NNievesCBS	KellyO	millermeeks
joekennedy	GingerGibson	sectrehargett	EricaLG	shaqbrewster	RitaHartIA
Donald_McEachin	singernews	GaSecofState	adamgoldmanNYT	AlexNBCNews	VoteJackie4NY
RepRichardNeal	emilyaheil	ARSecofState	grynbaum	mmurraypolitics	GarbarinoforNY
JoshHarder	JudyKurtz	OregonSOS	esullivannyt	aseitzwald	SriPKulkarni
Walker4NC	juliehdavis	scvotes	nixonron	albamonica	SheriffTNehls
RashidaTlaib	ggbirnbaum	vaELECT	AnaSwanson	VaughnHillyard	iamjohnoliver
RepSusanDavis	gdebenedetti	alasecofstate	helenecooper	LACaldwellDC	TheDailyShow
GKButterfield	Neda_Semnani	TXsecofstate	nytmike	chucktodd	Trevornoah
JoshGottheimer	MarcACaputo	COSecofState	UpshotNYT	KatyTurNBC	StephenAtHome
XochforCongress	ByGlenJohnson	CASOSvote	nytvideo	HallieJackson	hasanminhaj
BLeeForCongress	dcbigjohn	FrankLaRose	carolynryan	alivitali	dbongino
AngieCraigMN	davelevinthal	secstatewa	mattapuzzo	CarrieNBCNews	crookedmedia
Dr_RaulRuiz	PoliticoScott	realDonaldTrump	amyfiscus	elysepg	realBobWoodward
JerryNadler	poconnorWSJ	JoeBiden	arappeport	kailanikm	Comey
JudgeJohnCarter	BresPolitico	GovAndyBeshear	GioRussonello	AnthonyNBCNews	MotherJones
RepRickLarsen	michaelphirsh	gretchenwhitmer	igrullonpaz	BenjySarlin	NewYorker
MarkAmodeiNV2	JamieNovogrod	GovBillLee	caseysjournal	jacobsoboroff	tweetdrudge
JimmyGomezCA	DanEggenWPost	GovPritzker	stefsaul	frankthorp	BreitbartNews
StewartforUtah	MarcNBCBoston	GovJanetMills	AsteadWesley	kasie	theblaze
HaleyLive	CookPolitical	GovernorGordon	MaggieAstor	VivianBKim	realDailyWire
RepHorsford	MrWalterShapiro	GovSisolak	ByMattStevens	Miguelnbc	Alyssa_Milano
RepMcNerney	TimAlberta	GovNedLamont	kateconger	ChloeNouvelle	coldxman
rickallen	DavidMarkDC	GovMLG	NellieBowles	BetsyNBCNews	piersmorgan
RepJohnYarmuth	susanferrechio	GovKemp	Bill_Brink2	RJonesNews	Mike_Pence
RepTerriSewell	tomdiemer	GovTimWalz	dionnesearcey	ChrisJansing	WestWingReport
MikeTurnerOH	JulieSobel	GovDunleavy	katieglueck	NBCKatie	TeamTrump
ericswalwell	WaPoSean	GovLauraKelly	patrickhealynyt	ErikaJournal	MrsVanessaTrump
MacTXPress	jamieshuffle	GovParsonMO	rachel_dry	DartDClark	DiamondandSilk
MarioDBCamp	gbennettpost	GovMurphy	CarynAWilson	KerryNBC	GOPChairwoman
WeberforTexas	DLeonhardt	kayiveyforgov	jonathanellis	JuliaEAinsley	TheOnion
TeamMeadows	JonEasley	JanetMillsforME	sarahlyall	wloreilly	ProjectLincoln
rehemaellis	RepsForBiden				

Appendix B

Methods

The methods that we use for analysis require an understanding of how to handle our data and an understanding of the information we need from the data. Handling the data calls for a method to interface with Google Drive (where we store the data) and a method to work with pickle files.

B.1 Data Handling

We store our data on a shared Google Drive where each data file corresponds to one day's worth of data. In order to access one day's worth of data, we need to download the pickle file, then read the file. In our analysis, we work with a 98-day range from August 1st to November 7th so manually downloading (and deleting) every file whenever we need it would be tedious. Thus, we write a script to interface with the Google Drive API to easily download and read a data file. This script is viewable in subsequent Figures B.1.1, but the following provides a high-level overview.

In order to access the gDrive API, we first obtain API credentials through the Google Developers Console. Then, we build and return the Drive API service object via `authenticate_drive_api`. We can now download files from the shared drive using the returned drive object. To do so, we first use the drive object to get a list of file objects that match a certain query string using the `get_file_objs_list`

method. For example, if we want to download the files from August, we would first obtain a list of file objects by calling `get_file_objs_list(DRIVE, 'name contains "combined_tweets-2020-08"')`. We then loop through the list of file objects to download each file of data. The download method is `download_pkl_file` which first locally downloads the requested pickle file from the shared drive, then reads and stores the contents of the pickle file, removes the local copy, and finally returns the contents of the file.

The contents of the pickled file is a list of thinned tweet objects (see Section 3.1). We extract the necessary information from the downloaded thinned tweet data as we loop through and download each file object. Working with the data between downloads allows us to keep only one day's worth of data stored in our local memory. This saves space, but can cause time issues if our extraction methods are flawed since we may need to redownload every file.

B.1.1 gDrive API Script

```
"""
Edited on Mon September 7 11:50:33 2020

@author: sklucioni

Google Drive API setup.
"""
from __future__ import print_function
import pickle
import os.path, io
import general_utilities
from googleapiclient.discovery import build
from google_auth_oauthlib.flow import InstalledAppFlow
from google.auth.transport.requests import Request
from googleapiclient.http import MediaIoBaseDownload

# Google Drive API Docs: https://developers.google.com/resources/api-libraries
# /documentation/drive/v3/python/latest/index%2Ehtml

# If modifying these scopes, delete the file token.pickle.
SCOPES = ['https://www.googleapis.com/auth/drive.readonly']

def authenticate_drive_api():
    """
    Builds the drive api service and stores it in DRIVE.
    :return the drive api service.
    """
    creds = None
    # The file token.pickle stores the user's access and refresh tokens, and is
    # created automatically when the authorization flow completes for the first
    # time.
    if os.path.exists('token.pickle'):
        with open('token.pickle', 'rb') as token:
            creds = pickle.load(token)
    # If there are no (valid) credentials available, let the user log in.
    if not creds or not creds.valid:
        if creds and creds.expired and creds.refresh_token:
            creds.refresh(Request())
        else:
            flow = InstalledAppFlow.from_client_secrets_file(
                'credentials.json', SCOPES)
            creds = flow.run_local_server(port=0)
        # Save the credentials for the next run
        with open('token.pickle', 'wb') as token:
            pickle.dump(creds, token)

    DRIVE = build('drive', 'v3', credentials=creds)
    return DRIVE
```

```python
def get_file_objs_list(DRIVE, query):
    '''
    Use the drive api service to get a list of files according to the `query`
    string.
    :param DRIVE: the drive api service.
    :param query: string, A query for filtering the file results. See
    https://developers.google.com/drive/api/v3/search-files for supported
    syntax.

    :return A list of file objects.
    '''
    results = DRIVE.files().list(q=query).execute()
    items = results.get('files', [])
    return items


def download_file(DRIVE, file_obj, print_status=True):
    '''
    Use the drive api service to download a file from your personal gdrive. Does
    NOT remove the locally downloaded file. Prints a status report of the
    print_statusdownload (if `print_status == True`).
    :param DRIVE: the drive api service.
    :param file_obj: a file object, must have keys `id` and `name`.
    :param print_status: bool, Whether or not to print the file download status.

    :return None.
    '''
    if 'id' not in file_obj or 'name' not in file_obj:
        print('`file_obj` needs `id` and `name`.')
        return None

    request = DRIVE.files().get_media(fileId=file_obj['id'])
    fh = io.FileIO(file_obj['name'], mode='w')
    downloader = MediaIoBaseDownload(fh, request)
    done = False
    while done is False:
        status, done = downloader.next_chunk()
        if print_status: print("Download %d%%." % int(status.progress() * 100))

    return None


def download_pkl_file(DRIVE, file_obj, print_status=True):
    '''
    Use the drive api service to download and return the data from a pkl file
    from your personal gdrive. Removes the locally downloaded file. Prints a
    status report of the download (if `print_status == True`).
    :param DRIVE: the drive api service.
    :param file_obj: a file object, must have keys `id` and `name`.
    :param print_status: bool, Whether or not to print the file download status.

    :return pkl_data: the pkl data, most likely a list of `thinned_tweet_obj`s.
    '''
    download_file(DRIVE, file_obj, print_status)
    pkl_data = general_utilities.read_pkl(file_obj['name'])
    os.remove(file_obj['name'])
    return pkl_data


'''
Base code to download all combined_tweets-2020 files, grab pkl data, remove files

DRIVE = authenticate_drive_api()
file_objs_lst = get_file_objs_list('name contains "combined_tweets-2020"')
for file_obj in file_objs_lst:
    thinned_obj_tweets = download_pkl_file(DRIVE, file_obj, print_status=True):
    # Do something with the `thinned_obj_tweets` or store them to use later
'''
```

B.2 BERT Training

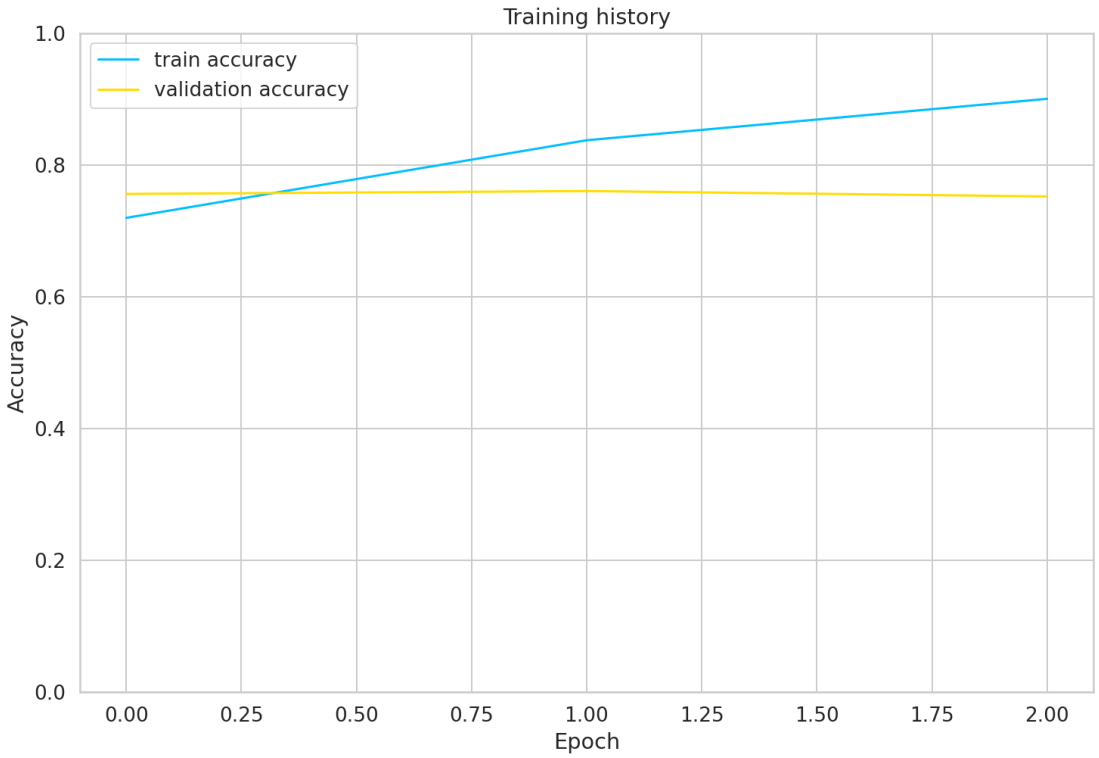
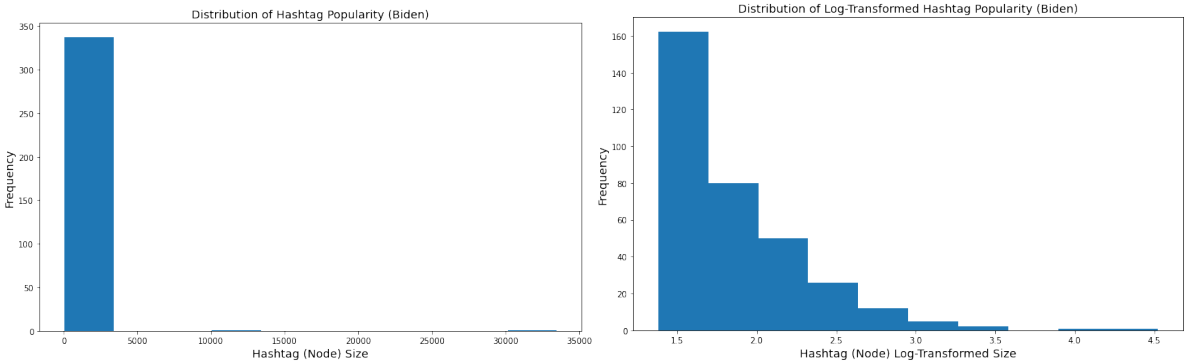


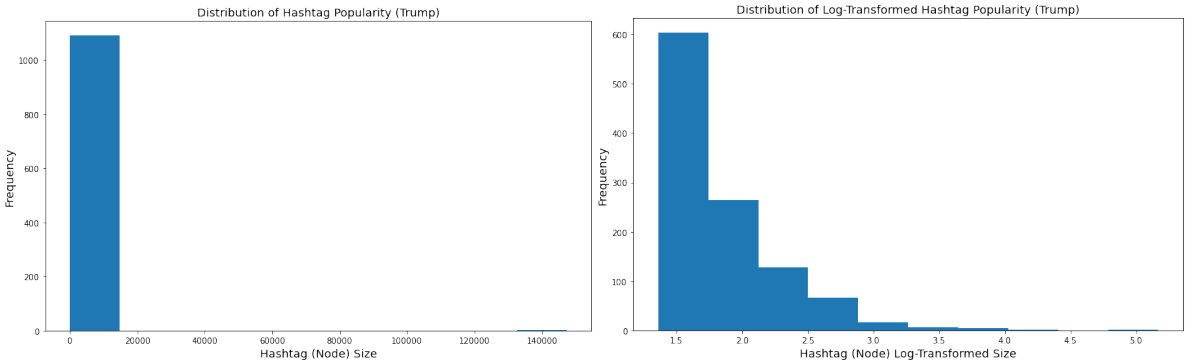
Figure B.1 BERT Training History

B.3 Ego Network Node Size Distribution



(a) Distribution of Hashtag Sizes **(b)** Distribution of (Log) Hashtag Sizes

Figure B.2 Distributions of Hashtag Sizes Centered at #Biden



(a) Distribution of Hashtag Sizes **(b)** Distribution of (Log) Hashtag Sizes

Figure B.3 Distributions of Hashtag Sizes Centered at #Trump

Appendix C

Results

C.1 List of Ego-Network Hashtags

trump	potus45	ivankatrump	msmistheenemyofthepeople	melania	dialitforward
trump2020	love	fuerzadiego	speakupvotebiden	macron	latinosportrump
bidenharris2020	taliban	bidencorruption	lanz	barr	equities
biden	silentnomore	thepeoplespresident	forthepeople	nbc	wewillvote
maga	harrisbiden2020	mueller	graham	detroit	crack
vote	comey	romney	conservatives	trumpisnotwell	2ndamendment
restartleader	endsars	buildthewall	oann	aoc	enoughisenough
kag	gopcomplicittraitors	merkel	energy	hohoho	bideniscorrupt
covid19	albany	kamalawonthedebate	amycomeybarrett	trumptrain2020	antifaterrorists
miga	cartoonrepost	turkey	makingherstory	russiagate	redkingdom
election2020	obamawasbetterateverything	democratsarecorrupt	vets4biden	votetrump	ny
maga2020	georgie	پیمان_نوین	nastywoman	clinton	london
trump2020landslide	末満健一	onev1	trumpviruscatastrophe	godblessamerica	vpdebates
restart_opposition	kingdomwithpahlavi	fraud	ludemedia	pandemic	50cent
keepamericagreat	trumpsamerica	corruptjoebiden	tech	whereshunter	wieszwięcej
kag2020	liar	google	wednesdaymotivation	mega	trumpriots
usa	truthmatters	virus	rice	kag2020landslidevictory	gophypocrisy
coronavirus	cia	dnc2020	immigration	israel	eleccionesusa2020
joebiden	zemmour	bigtech	powerbelongstothepeople	thesaracartershow	mypresident
elections2020	themighty200	voterepublican	planyourvote	everyvotecounts	godblesstrump
ccot	trump2020kag	venezuela	lincolnproject	trumplandslidevictory2020	bidencrimefamilly
gop	thenewfederalstateofchina	trumptownhall	republicansagainsttrump	hawaii	gretathunberg
foxandfriends	meaculpa	darktolight	laptopgate	tiktok	fracking
walkaway	unemployment	poll	5Oct	15days	weloveyoutrump
biden2020	bidens	patriot	ditchmitch	tucker	supporters
covid	bidenforpresident	ترامپ	subpoenasong	msnbc	ridenwithbiden2020
armyfortrump	notmypresident	countryoverparty	nazis	yourvotematters	newyorkcity
votehimout	factsmatter	emoluments	medicareforall	deepstate	makeamericagreatagainagain
4moreyears	judgejeanine	lies	durham	kamalaharrisvp	votedemsout
voteredtosaveamerica	randomtrump	iamavoter	ruthbaderginsburg	kaga2020	teaparty
voteearly	trumppressconference	ncpol	azerbaijan	makeamericagreatagain	hopehicks
electionday	happybirthdaykamala	filltheseat	taxreturns	eleccioneseeuu	iranians_hate_pmoi_rezapahlavi
trumppence2020	trumpliesamericansdie	wevegotacountrytosave	amazon	electionday2020	november3rd
donaldtrump	socialsecurity	bargesfortrump	caughtinukraine	fridaythoughts	demcastca
democrats	lesleystahl	pa	votingspells	uselections2020	limbaugh
cnn	trumpisacoward	teamjoe	gtcblue	savethechildren	trumpismypresident
whitehouse	hidenbiden	wwg1wgaworldwide	60minutesinterview	votebluetoendthenightmare	hungry
wwg1wga	presidentialelection2020	bidenharis2020	crimesagainsthumanity	superspreader	crazynancy
kamalaharris	giphy	salvini	trumpvirusdeathtoll211k	rnc	بایدن
bidenharris	gif	paytoplay	deutschland	votefortrump	aka1908

potus	live	brexit	trumpterroism	abortion	landslidevictory
foxnews	flotus	covid—19	democratsfortrump	votered2020	bluesenate
votered	trumpcrimesyndicate	chinajoe	1runrundiceq	wednesdaywisdom	jobsnotmobs
trump2020tosaveamerica	wechat	red	bobulinsky	msm	liberalhypocrisy
votebluetosaveamerica	russian	rbypoll	unfit	hannity	border
fourmoreyears	murdoch	bobulinski	trumpchaos	blacksfortrump	whereishunter
voteblue	victory	ivotedearly	pencefly	thesilentmajority	votebymail
electionresults2020	voteoutthegop	nra	votegopout	walkawayfromdemocrats	trumpworld
america	save12hkyouths	digitalsoilder	donotelectkamala	floridaforbiden	fliptexasblue
tcot	chrischristie	economy	jillbiden	votebidenharris	chuckschumer
trump2020landslidevictory	ری_استارت_تنها_راه_نجات	markmeadows	childabuse	teamtrump	theawakening
bidenharristosaveamerica	factcheck	maga2020landslide	political	blackvoicesfortrump	bernie
cnnsotu	trumpgravedancer	トランプ大統領勝利	trumpliedamericansdied	ukraine	teampetty4kamala
obama	getoutthevote	ingrahamangle	smartnews	obamagate	viral
pennsylvania	newjersey	antrim	unfittobepresident	bestpresidentever	ai
resist	blackwomenfortrump	vicepresidentialdebate	trumpisamerica	sleepyjoe	medicare
morningjoe	uswahl	modiji	louisiana	stimulus	hcq
bluewave2020	joebidenkamalaharris2020	holdtheline	followtrick	antifa	thestory
trumprally	memes	trumphascorona	hamiltontownhall	fbi	nuitamericaine
redwave	wisconsinriots	byetiffany	electionsamericaines	newyork	joebidencrimefamily
bidenharrislandslide2020	whitesupremacy	mussolini	chinabankaccount	nancypelosi	pridenews1
americafirst	trumprallyflorida	nyc	noalnom	california	teampa
electionnight	westandwithpahlavi	فوری	latino	ccpvireus	tuesdayvibes
realdonaldtrump	trumpparade	kamalaharrisforvp	stocks	conservative	juntossomoslaverdad
trumpmeltdown	pride2020	springsteen	handsome	trumplandslide2020	physician
harris	floride	netanyahu	wuhanvirus	hillary	bobwoodward
trumphascovid	bikersfortrump	bidencalm	trumpisaliar	hillaryclinton	joebidentownhall
breaking	unitedwestand	trumpknew	salongate	walterreedhospital	neverbidenneverharris
get2getherexperience	humanrights	abdninsecimi	world	uselectionresults	op1
hitrefresh	washington	dadjokes	gotmfv	democracy	mccabe
humantraffickingawareness	scalise	msnbc2020	russiacollusion	votetrumpout	obamacare
china	votejoebiden2020	melaniatrump	restart_oppsition	alllivesmatter	lesliestahl
election	trumpcountdown	jewsfortrump	trumptraitor	iwillvote	gotrump2020
trumptrain	silentmajorityrising	stevekbannon	doctors	follow	huge
draintheswamp	erictrump	fakenewsmedia	platinumplan	veterans	repubs
florida	liberal	nbcblackout	herdimmunity	byedon2020	white
new	spygate	bannon	neveragain	breakingnews	firetrump
michigan	warroompandemic	onlytheyoung	everyvoice	defenddemocracy	votejoe
voteblue2020	jfkjr	flynn	trumpdepression	trumpisaloser	trumpadministration
eeuu	census2020	hunterbidenemails	culp2020	freedom	traitor
republicans	forgodandcountry	abd	thread	redwavechallenge	جاویدشاه
dumptrump	trumppence	trusttrump	vpharris	voteinperson	government
vote2020	トランプ	gsa	pakistan	usa2020	votehimoutandlockhimup
hunter	decision2020	death	gopsuperspreaders	r2p	trumpwarroom
trumpcovid	meme	mcconell	duluth	lovemypresident	merz
news	rnc convention2020	redpilled	nba	byedon	stoptrumpsterror
georgia	dexamethasone	stopthecount	chinajoebiden	facebook	feminism
latinosfortrump	politicalcartoon	rudygiuliani	schiff	bidenwonthedebate	rif_raf
republican	newsmax	republicansforbiden	dontgambleontrump	votethemout	dontgetcomplacent
uselection2020	hunterbidenslaptop	dopeydon	leadership	saveourchildren	southcarolina
mediabuzz	bidentaxcuts	regeneron	pedophile	santathug	azerbajian
hunterbiden	cbs	removetrump	votedonaldjtrump	schoolchoice	pacifier
bidencrimefamily	bebest	racism	salvinicheerleader	proudboys	karma
stopthesteal	bidenwon	wwg1gwa	alaska	turntexasblue	wakeup
voterfraud	vppick	lindseygraham	thejagmeetsingh	savannahguthrie	trunalimunumaprzure
fakenews	wtp2020	godblessourmilitary	trumpearnedmyvote	cnnpolitics	bluetsunami
bidenharris2020landslide	unitedstates	trumplies	debatenight	bidenriots	health
trumpvirus	patriotsunite	syria	tedcruz	25thamendmentnow	trumpfascist
walterreed	peace	patriotismwins	trumps4eva	tonyfauci	trumpprotest
corona	donaldjtrump	lockdown	satire	ue	bidenlaptop
bluewave	magats	stockmarket	2020debate	williambarr	bbpe
voteredtosaveamerica2020	geogia	floridafortrump	mondaymorning	thebestisyettocome	criminal
pence	vaccine	trumpisaracist	bush	debate2020	gotrump
us	christian	earlyvoting	casabianca	uselection	アニメイト
arizona	trumpwins	wattersworld	trumptantrum	covidiots	withbidenwecan
bidencrimefamilyexposed	fbrparty	moscowmitch	transgender	usaelection2020	wtpbiden
politics	remember911	americaneedspennsylvania	kushner	dnc	trumptosaveamerica
sundaythoughts	kenosha	traitortrump	democratshateamerica	amymcgrath	deblasio
elecciones2020	trumpfailedamerica	berniesanders	time	bidensunitingus	npo
elections	sleepyjoebiden	epstein	ridenwithbiden	2a	stimuluspackage
bidenharris2020tosaveamerica	mrrogers	nunsfortrump	violence	donald	potustrump
foxandfriend	trumpcoupplot	weareth erally	nojoe	rt	yourefired
trumpcovid19	canada	fresh	johnlewis	americaortrump	lilwayne
welovetrump	stopthecheating	digitalsoldier	cult	promisesmadepromiseskept	why
trumpvsbiden	voterfruad	restart	trumptrainwreck	godblesstheusa	whistleblowermovement
vpdebate	ymca	landslide	fuckingidiot	2020elections	youarechumps
debates2020	bidenfamilycorruption	benghaziaintgoingaway	strzok	thenewfedralstatofchina	ヤングエース
leadright	黒世界	people	teampettyheadlines	democratic	liberals

restartmiga	gopdeathcult	europe	amerika	khive	illegitimatepresident
texas	teampetty4dems	saynotojoe	zaddy	magadragtheinterstate	isis
amyconeybarrett	dc	congress	vance	votelikeyourlifedependsonit	tds
twitter	americanpride	silentmajority	iartg	vog	math
maga2020landslidevictory	votethemallout2020	nfl	trumpmustgo	buildbackbetter	kaga
votebluetoendthisnightmare	バイデン	johnson	nothingcanstopwhatiscoming	ivanka	standupforamerica
covid_19	trumpgenocide	twgrp	hunterslaptop	saturdaymorning	laptop
votebluedownballot	progressives	italy	bidenisaracist	wallstreet	voteforbidenharris
presidenttrump	unite	philadelphia	trumplieseverytimehespeaks	bidencoalition	bidenemaildeclassification
sundaymorning	lockhimup	pennsylvaniafortrump	letswinthis	elecciones	downballotblue
womenfortrump	foxnew	healthcare	puertorico	etatsunis	sealteam6
patriots	voteinperson2020	tn	gettysburg	make_iran_great_again	journalism
lawandorder	johnkerry	tuckercarlson	notsmart	votebluedowntheballot	taxcuts
icecube	dhs	cdc	genocide	trumpcollapse	jordansview
bidentownhall	eua	minnesota	ballotharvesting	qarmyjapanflynn	hypocrite
blacklivesmatter	rosegarden	senatebluewaveday	babieslivesmatter	givebacktrumpvotes	truthsandwich
votebiden	bengarrison	wednesdaythoughts	kirstjennielsen	wethepeople	bbc
usaelections2020	noh8	pafortrump	bigdata	like	oil
rememberinnovember	melaniaisahooker	trump2024	kentucky	fox	parade
thefive	mjhegarfortx	25thamendment	dumptrumpdaynov3	comment	comedy
60minutes	borderwall	oms	thegreatawakeningworldwide	bidencrimefamiiy	turncalifornia red
flipthesenate	texasforbiden	iowarally	vietradiaz	digitalsoldiers	militia
trumpislosing	operationmaga	impeachment	hospital	trumplandslide	movingmountains
president	afghanistan	traitorinchief	page	votetrump2020tosaveamerica	โควิด19
godwins	boatparade	liberalismisamentaldisorder	novel	presidentialdebate2020	pridefamily
burisma	terrorist	socialism	win	election2020results	terrorism
counteveryvote	gtv	electionfraud	michelleobama	truth	nagornokarabakh
trump2020nowmorethanever	treason	debates	nixon	american	uspresidentialelections2020
dumptrump2020	kansas	voteearlyday	wwg1wgaworlwide	bluelivesmatter	dolar
nebraska	deplorables	americasgreatestmistake	jeffreyepstein	kamala	trumphasmyvote
saturdaythoughts	startups	godblesspresidenttrump	yates	senate	nyt
omaha	trumpvirusdeathtoll275k	theresistance	free	teamdl	amlo
blexit	assange	iowa	joebiwan	billbarr	letsdothis
scotus	kayleighmcenany	tonybobulinski	justice	covid__19	artsakh
mitchmcconnell	nofear	creepyjoe	democats	townhall	étatsunis
votehimout2020	teampetty4joe	thursdaythoughts	amjoy	qanons	mailinvoting
fauci	barrett	cartoon	rapistinchief	amas	californiansfortrump
trumpisalaughingstock	veteransfortrump	crookedjoe	entrepreneur	joebiden2020	judgejanine
presidentt	evangelicals	dementiajoe	nickpacilio	fuck2020	qcrushdigs
ccp	racistinchief	returnoftheusa	karabakhisazerbaijan	mcconnell	bidentherapist
democrat	bidenforfl	hollywood	gay	trumptaxreturns	رضاشاه_روحت_شاد
putin	woodward	uswbatlight	ifbap	bidenharristoendthisnightmare	hope
northcarolina	doitbold	trumprallybutler	colluded	racist	redtsunami2020
fightback	donthecon	giuliani	dontbelievearmenia	wearadamnmask	democraticconvention
qanon	penceknew	kenoshariot	donaldtrump2020	covidcaughttrump	trumppence2020landslidevictory
ohio	chinavirus	votehealth	maskup	prayfortrump	trumpcrimefamilyforprison
debate	stopthefraud	politicaltwitter	vota	covidiot	philly
trumpisanationaldisgrace	section230	amishsupporttrump	itiswhatitis	corruption	science
trumpcrimefamily	funny	bidenlies	pensilvania	counteverylegalvote	tuckercarlsontonight
bidencrimefamiily	atlanta	trumpsupporters	propaganda	bestpresidentever45	facts
winning	business	gnews	afd	stevebannon	rncconvention
blm	democraticparty	neverbiden	redpill	cuba	wtpblue
pelosi	orangeshitgibbon	trumpiscompromised	fact	dems	fascist
vpdebate2020	evicted	youaintblack	presseclub	mondaythoughts	dragtheinterstate
2020election	trumpsuperspreader	borisjohnson	mundo	kerry	brennan
bidencares	eleccioneseeuu2020	trumplied200kdied	westwingspecial	gotv	goodtrouble
votebidenharris2020	electioninterference	veteransagainsttrump	sidehustle	demcast	h1n1
backtheblue	maduro	trump20	rogerstone	pennsylvanie	الانتخابات_الامريكيه
crookedjoebiden	últimahora	bolsonaro2022	faith	wearamask	nevertrump
russia	denvershooting	letsridefor45	house	chrisfarrell	thursdaymotivation
cbsnews	letsgetserious	riggedelection	christopherwray	saveamerica	followthemoney
redwave2020	mask	beijingbiden	votebluenomatterwho	fbr	mondaymood
americans	joe	elezioniusa2020	ingodwetrust	thegreatawakening	brasile
americaneedsjoe	keepthefaith	mikepence	terrorists	uselections	nwo
democratsaredestroyingamerica	louisdejoy	halloween	suspended	uspolitics	wow
blametrump	barackobama	fukouttahere	specialreport	votered2020removeeverydemocrat	quidprojoe
supremecourt	keitholbermann	socialist	seanconley	invincibleherotrump	kkk
rushlimbaugh	filltheseatnow	joebidenukrainescandal	méxico	strongertogether	46reasonswhy
votethemallout	patriotsfight	voting	voteoutcorruptgop	laptopfromhell	hispanicsfortrump
iran	susancollins	govote	mexico	uk	whitmer
votebidenharristosaveamerica	debatetonight	promiseskept	walkawayfromdemocratsforever	senaterepublicans	newyorktimes
flipthesenateblue	votes	votereadу	gold	acb	pennsylvaniaforbiden
cnnspolitics	thehill	doj	trumpsforever	q	pray
trump2020landside	whistleblower	greatawakening	mondayvibes	fucktrump	scotusnominee
humantrafficking	toobigtorig	trumpispathetic	dems4usa	uswahl2020	netflix
byebyetrump	pencewonthedebate	battlegroundstates	steelbelt	voteredlikeyourlifedependsonit	paulawhite
demvoice1	endthenightmare	lockherup	trump2020tosaveourcountry	prolife	trumpprojection
rnc2020	ga	washingtonpost	counterpunch	nobelpeaceprize	smile

votetrump2020	lepen	patriotsunited	recount2020	polls	cuomoliedseniorsdied
thisisnotover	trumplandslide2q2q	trumpcovidhoax	jfk	fridayfeeling	themoreyouknow
ivoted	israeli	latinos	getwellsoonmrpresident	usps	noescape
wakeupamerica	generation	uspoli	jackassjoe	presidentelectjoe	russianhoax
media	fake	fail	karabakh	votebluetosaveamerica2020	nazi
nevada	pence2024	liarinchief	trump2016	bolsonaro	nj
firetheliar	aca	trending	michiganfortrump	rally	houellebecq
wisconsin	whitehousevirus	whatilearnedtoday	nigeria	womenarevoting	keepamericagreat2020
mailinvotingfraud	democratstheenemywithin	eu	dictatortrump	resistance	bfmtv
chinabitchbiden	trumpjr2024	money	jaimeharrisonforsenate	cspan	trumprussia
gojoe	4Nov	endourgenocide	ralphreed	trumpliedpeopledied	afazenda12
losangeles	gojoe2020	queen	bidenwillcrushcovid	modi	sale
censorship	newmexico	trusttheplan	hayscounty	votetrumpout2020	notmeus
tuesdaythoughts	bidenleaks	voteredinperson	staystrong	military	ilhanomar
ccpvirus	failure	obamabidengate	ganjacity	voteresponsibly	lockdown2
presidentialelection	trump2020lanslide	trumphasnoplan	octobersurprise	electionresults	nato
black	patriotsawakened	direct	who	trumpnowmorethanever2020	bassetti
war	トランプ大統領	god	trumpisacompletefailure	trumpsthebiggestliarever	fly
miami	pa01	walkawaycampaign	peacefulprotest	gopbetrayedamerica	believeinamerica
ridinwithbiden	ری_استارت	portland	importante	1u	présidentielle
joebidensneighborhood	copolitics	californiafortrump	iranprotests	chriswallace	wwg1wgaww
pelosimustgo	voted	votefordemocracy	antifas	nhpolitics	bolivia
nobel	ハンターバイデン	trumpnowmorethanever	donlemon	nobelprize	courtpacking
climatechange	penceharrisdebate	india	trumpknewanddidnothing	joewillleadus	christians
saturdaymotivation	trumpisacriminal	florida4trump	humor	abc	borat
bluetsunami2020	ufo	trumpwins2020election	bombshell	landslide2020	secretservice
loser	weloveyou	mondaymotivation	az	flygate	blue
bitcoin	burningupdemo	hydroxychloroquine	hongkong	police	h1b
trumpisunwell	trumpcampaign	hoax	traitors	france	hitler
dejoy	biafraexit	ballots	climateaction	donaldtrumpjr	conte
trumpisatraitor	abraham	crazyuncletrump	germany	campaign	impeached
redtsunami	resisters	retweet	support	magaa	japan
trump2020tosavetheworld	un	trumpdance	trumpdoesntcare	imspeaking	mitchmustgo
video	golf	demconvention	trumpforprison2020	theview	reagan
thepostelection	electoralcollege	presidentelect	rep	prayforpresidenttrump	womenforbiden
earlyvote	pensacola	eleccion2020	grabhimbytheballot	walnut	disney
poutine	lgbt	fredohaslostit	blackpink	calltheelection	enthusiasm
cyrus_i	wehaveherback	rinos	trumpstroke	coup	lol
nypost	wearethenewsnow	nytimes	boris	voters	sondages
huawei	impeachtrump	vicepresidentialdebate2020	walmart	daylightsavings	gretianstag
uswahlen2020	farage	merica	pleasesavethesechildren	countallthevotes	vamos
presidentialdebate	postalservice	catholic	obamaspeech	republicanparty	imvotingforjoe
qanon2020	walterreedmedicalcenter	kidvicious	abcnews	bidenemails	therightview
confirmacb	hamilton	redwavecoming	joniernst	crime	england
remdesivir	deutschebank	demsaredestroyingamerica	historically	uselectionresults2020	conley
taxes	mamala	5g	cina	maddow	dzt
votersuppression	editorialcartoon	turquie	ethiopia	whereisthemoneycatherine	mog
trumpisnotamerica	benghazi	phoenix	consequencesfortrump	defundcbc	meganmcglover
flipcalifomiared	cruz	45kagxpress	corruptgop	chicago	casablanca
trumpstrong	youtube	urgente	virginia	jaredkushner	migrant
trumpforfreedom	viganò	libya	bretbaier	gifs	mondialistes
putinspuppet	wahl	voterfraudisreal	elezioniusa	bidenpresident	rosegardenmassacre
sundayfutures	raf	auspol	censors	trumpboatparade	michaelcohen
jesus	vetsagainsttrump	drfauci	swamp	vp	nosjournaal
fuckaroundandfindout	ignorethepolls	metoo	soros	masks	economic
trumpchinabankaccount	egypt	armenia	isil	theworldmustknow	milestaylor
wwg1wwga	kavanaugh	mattgaetz	census	bidencrimesyndicate	evil
bodylanguage	seattle	cdnpoli	briefwahl	briefwahlunterlagen	tennessee
snlpremiere	wonderfulmichigan	presidentbiden	maine	constitution	nbcnews
bideniswinning	amyconeybarrettscotus	votetrumptosaveamerica	qanon2018	women	forex
endthechaos	trumpliespeopledie	voteouteveryrepublican	coronaviruspandemic	latinosfortrump2020	polling
islam	mvpharris	americadecides2020	cowards	icymi	gdp
géorgie	vigano	gopcorruptionovercountry	complicitgop	whitesupremacists	dax
rbg	trumpterrorism	nationalsecurityrisk	patriotes	uae	cult45
liberalcorruption	dictator	hunterbidenlaptop	mailfraud	voteforher	gaysfortrump
brazil	hypocrisy	sirte	berlusconi	firstlady	deepstateexposed
buckeyebattalion	wipolitics	inspire	antitrump	sudan	mcsally
trumptaxes	caresact	امپراطورپرشیاسینمحمدحسینی	berlin	crookedbiden	wray
flipthehouse	wiright	agbarr	fascism	日本の準備は整いました	fillthatseat
jobs	johnkelly	fraude	country	chinese	bidengate
neverforget	trumpvsbidenfight	nasdaq	usarmy	bible	mailinballots
bbtvi	ultimahora	flu	executiveorder	killercuomo	ia01
exclusive	usmc	chumpsfortrump	paris	ppc	donnabrazile
freakout	wahlbetrug	revoke	australia	scamdemic2020	incompetent
erdogan	covita	keeptexasred	crazy	irise	stayinline
latinosforbiden	meetthepress	ittakesallofus	thegreatawkening	ترامپ	kanyewest
drainthedeepstateswamp	twittercensorship	estadosunidos	lgbtq	corrupt	flynnwasframed
pompeo					