# Examining Celebrity and National Figure Twitter Posting with Efforts to Overturn the 2020 U.S. Presidential Election

Cathleen Cusachs, Paula Penghui He, Harini Jaisankar, Nicole Yang, Levi Bevis

College of Communication, Boston University

EM747: Trending Insights

Dr. Chris Chao Su

April 29, 2022

## Introduction

The 2020 election was unlike any other in the history of the United States. As the global COVID-19 pandemic upended almost all aspects of modern life and forced campaigning to become an almost exclusively digital affair, significant public attention moved from campaign rallies and events to candidates' social media platforms. Following the election, logistical challenges in mailing, receiving, and counting millions of mail-in ballots, along with numerous accusations by former President Donald Trump that widespread voter fraud had "stolen" the election from him, the narrative of a stolen election gained traction. The Republican Party largely embraced this narrative, and a partisan divide over the election processes emerged. Accusations of voter fraud were present on social media following the election, with some interactions quickly becoming hostile. On January 6, 2021, a rally was held in Washington, D.C. in support of the "Stop the Steal" narrative. After hearing numerous speakers, including former President Trump, rally attendees marched to Capitol Hill and infiltrated the Capitol building. The Capitol insurrection brought international attention to the message put forward by President Trump's supporters, and it renewed interest in the role partisanship and social media played in the build-up to the event.

Twitter has proven to be the ideal platform to study in terms of partisanship and online activity. One 2021 study conducted by Gaisbauer et al. indicates that Twitter users who retweet or more readily engage with far-right politicians on the platform tend to be more aggressive in responding to those who do not identify with their beliefs, thus increasing the visibility of far-right stances and ideology (Gaisbauer et al., 2021).

Another 2021 experimental study conducted by Berlinski et al. examined how unsubstantiated claims of voter fraud influenced confidence in the United States' electoral process. The researchers found that those who were exposed to unsupported claims of voter fraud, most often Republicans and Trump supporters, expressed lower levels of confidence in the election (Berlinski et al., 2021).

The influence of high-profile national figures' social media accounts has also been the subject of research, and a 2018 study by researchers Stolee and Caton offer new insight on how Trump's use of Twitter speaks directly to a determined base instead of a more broad audience. They find that Trump's dramatic shift away from the historical use of messaging to a wide audience may indicate a substantial shift in presidential campaign outreach, with candidates more likely to speak more directly at a particular audience and base than users at-large (Solee and Caton, 2018).

The goal of this research is to investigate the relationship between partisanship, support for former President Donald Trump, and emotional sentiment displayed in tweets toward the 2020 election and January 6th insurrection. To this end, we ask the following research question, and pose the following hypotheses:

*RQ1: Does support for then-President Donald Trump's claim that the 2020 election was stolen significantly differ based on support for then-President Trump's 2-2-presidential candidacy?*

*H1: Author partisanship is significantly correlated with a tweet disagreeing with 2020 election results.*

*H2: Tweets featuring the hashtag #StoptheSteal will positively correlate with aggressive and/or violent language in comparison to tweets featuring the hashtag #Biden2020.*

*H3: The tone of tweets mentioning Trump will be more negative than the tone of tweets* *mentioning Biden.*

## Method

We sampled tweets from verified accounts only. Our time frame was between November 6th, 2020 and January 6th, 2021. This time was selected because it encapsulates reactions to the 2020 presidential election results, and sentiments leading up to the January 6th Capitol insurrection. We collected a sample of 1,000 tweets in order to complete our manual coding. Each tweet and the content within the tweets were our units of analysis. To conduct an inter-coder reliability test, two coders conducted a content analysis of 100 tweets separate from our sample of 1,000 tweets. These results were then fed through an IRC online calculator to find their agreement percentage and Krippendorff's alpha. The full results are included in our appendix, but it should be noted that all variables aside from variable 11 and variable 13 had a sufficient percentage agreement (above 90%).

### *Sentiment analysis*

The method that we have used in our analysis is the Hu and Liu's opinion lexicon as used in class in order to determine the following elements: whether the message is positive or negative, and the magnitude of the sentiment.

The primary reason behind adopting this method of analysis is to determine the sentiment in these tweets to help us answer our research question, which primarily is to understand the sentiment expressed based on the political party or leaning of the tweet.

### *Unsupervised Machine Learning*

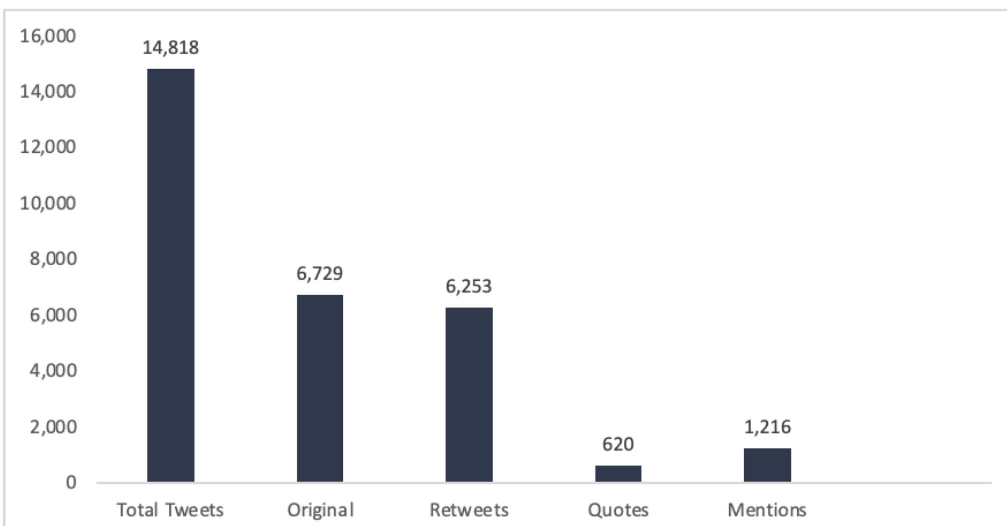
We have adopted topic modeling using latent dirichlet allocation (LDA) to group individuals in our data. In our research, we use this form of analysis to understand the most commonly found topic and words during the election period of 2020 to get an insight into the widely discussed topics and words pertaining to the election results.

## Results

Before conducting an analysis based on the posed research question and hypotheses, a few descriptive statistical tests were run in order to determine general information about the data. First, researchers wanted to determine when the volume of tweets, and when the most tweets were authored during the given timeframe.

**Figure 1**

*Tweets amounts and types*

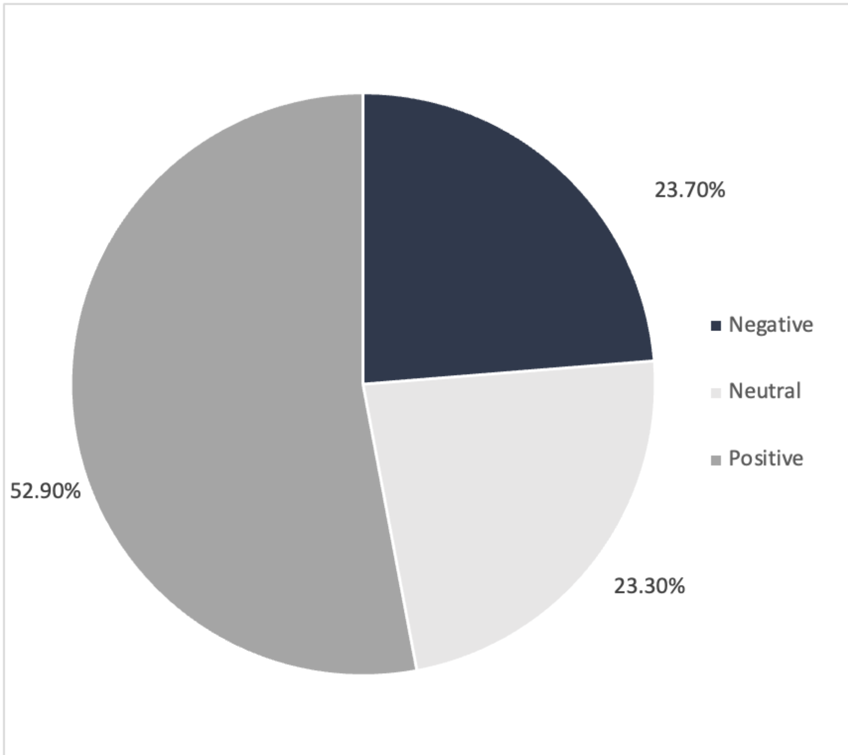



*Note: The unit of measure is a tweet. The time distribution is 11/6/2020 to 1/6/2021.*

In total, 14,818 tweets were pulled between November 9th 2020, and January 6th, 2021. In the first thirty days between November 9th and December 8th, there were 9,685 tweets. In the next thirty days between December 9th and January 6th, 5,133 tweets were pulled. There were nearly two times as many tweets in the data immediately after the election, and in the weeks that followed, than there were leading up to the capitol insurrection on January 6th, 2021.

**Figure 2**

*Sentiment analysis of all tweets*

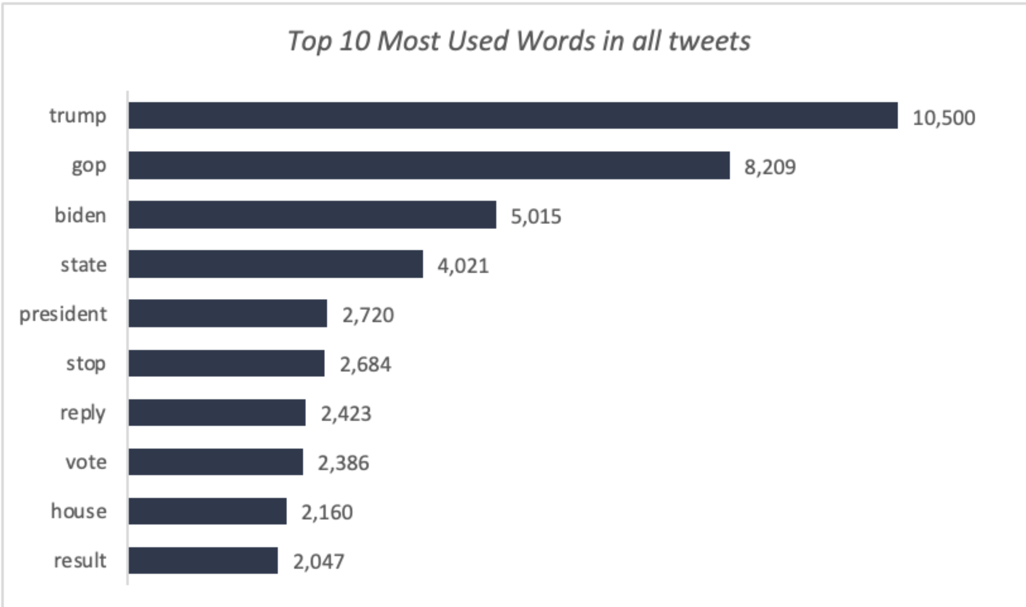



Note: *The scale for negative is -1, for neutral is 0, for positive is 1.*

Next, researchers conducted a sentiment analysis of all 14,818 tweets using the Opinion Lexicon developed by Minging Hu and Bing Liu (Hu and Liu, 2004), and a scale of -1 for negative, 0 for neutral, 1 for positive. The analysis showed that 52.9% of the tweets were neutral in tone, 23.7% were negative, and 23.3% were positive. In order to conduct hypothesis testing, further sentiment analysis was done on tweets that specifically mentioned the term "biden," and the term "trump." The analysis found that the "biden" tweets had a mean value of .38. "Trump" tweets had a higher mean value of .43. This indicates that tweets mentioning "biden" were actually more negative in tone than tweets mentioning "trump."

**Figure 3**

*Word frequencies of all tweets*




Note: *The Axis data labels represent the number of times each word appears in the total number of tweets.*

Researchers also conducted a frequency test to determine the most used words in the dataset. The word "trump" was used the most frequently, appearing 10,500 times in the whole dataset. "Gop" was the next most frequent word, used 8,209 times. The rest of the words were used significantly less frequently, with "election" appearing 5,829 times, and "biden" appearing 5,015 times.

**Figure 4**

*Word Cloud of "Trump"*

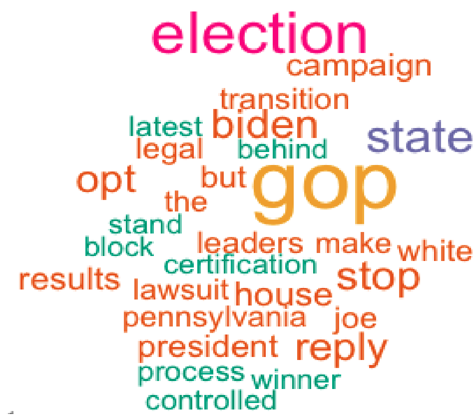



**Figure 5**

*Word Cloud of "Biden"*


Another descriptive analysis that researchers conducted was word associations between the common terms "biden" and "trump." Three of the most common associations with the term "trump" were "election," "gop," and "state." For the term "biden," the most common associated terms were "gop," "trump," and "election."

**Table 1**

***Topic Model Results***

| Topic | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| state | elect | hous | gop | biden | republica | trump | biden | american |
| result | trump | make | trump | presid | say | democrat | vote | one |
| campaign | offici | process | lose | administr | senat | covid | elector | democraci |
| lawsuit | gop | joe | elect | tell | presid | need | win | year |
| pennsylvania | fraud | leader | court | nation | georgia | support | million | peopl |
| block | voter | transit | reject | today | gop | donald | certifi | time |
| control | claim | legal | effort | day | parti | will | michigan | country |
| late | ballot | white | overturn | new | call | offic | colleg | power |
| certif | continu | move | lawmak | offici | member | gop | lead | mani |
| stop | evid | stand | challeng | capitol | congress | check | victori | everi |

*Note: data was stemmed prior to being run through LDA*

In order to begin the full data analysis, the research team decided to conduct unsupervised machine learning, and use a Latent Dirichlet Allocation (LDA) to create a topic model. The LDA returned ten topics consisting of ten terms each. We determined that nine of those topics were usable. The tenth appeared to miscellaneous stopwords that were not included in the initial data cleaning list.

Three of the topics appear to be discussions surrounding the electoral votes in different states: Michigan, Pennsylvania, and Georgia. Both during and after the election, these three states were points of contention in the ongoing discourse about election fraud in mail-in voting. Several of the topics included the term "trump" or "biden," but none of the topics included both terms. In contrast with the results from testing hypothesis three, the topics with the term "trump" also include negative terms such as "fraud" or "lose."

*Research Question 1*

**Table 2**

Descriptive statistics

| Author | N | Mean | SD | SE Mean |
|---|---|---|---|---|
| Republican | 49 | 1.3265 | .47380 | .06769 |
| Democratic | 252 | 1.9722 | .16466 | .01037 |

**Table 3**

Independent sample t-test

| | t | df | Sig. (2-tailed) | Mean difference | 95% CI lower | 95% CI upper |
|---|---|---|---|---|---|---|
| Equal variances assumed | 17.055 | 299 | <.001 | .64569 | .57119 | .72020 |
| Equal variances not assumed | 9.429 | 50.276 | <.001 | .64569 | .50817 | .78321 |

*Note: Mean value is significantly different if Sig. (2-tailed)<0.001.*

To test the research question, researchers conducted an independent sample T-test. Scale 1 was if the tweet disputed the 2020 election results, and scale 2 was if it upheld the election results. Due to an inability to conduct ideological scaling in R, we instead utilized the manual coding sample consisting of 1,000 tweets. The mean difference between the two scales was 0.64 ($p < .001$) with a 95% confidence interval spanning from 0.51 to 0.78..

*Hypothesis 1*

**Table 4**

Correlation tests (when used the “unsure” data in manual coding as well)

| | | the partisanship of the author | claims against the results of the 2020 election |
|---|---|---|---|
| the partisanship of the | Pearson | 1 | -.182** |
| author | Correlation | | <.001 |
| | Sig. (2-tailed) | | 1001 |
| | N | 1001 | |
| claims against the results of the 2020 election | Pearson | -.182** | 1 |
| | Correlation | <.001 | |
| | Sig. (2-tailed) | | |
| | N | 1001 | 1001 |

*Note: **.Correlation is significant at the 0.01 level(2-tailed)*

**Table 5**

Correlation tests( one just for just “democrats” and “republicans”)

| | | the partisanship of the author | claims against the results of the 2020 election |
|---|---|---|---|
| the partisanship of the author | Pearson Correlation | 1 | -.702** |
| | Sig. (2-tailed) | | <.001 |
| | N | 418 | 301 |
| claims against the results of the 2020 election | Pearson Correlation | -.702** | 1 |
| | Sig. (2-tailed) | <.001 | |
| | N | 301 | 418 |

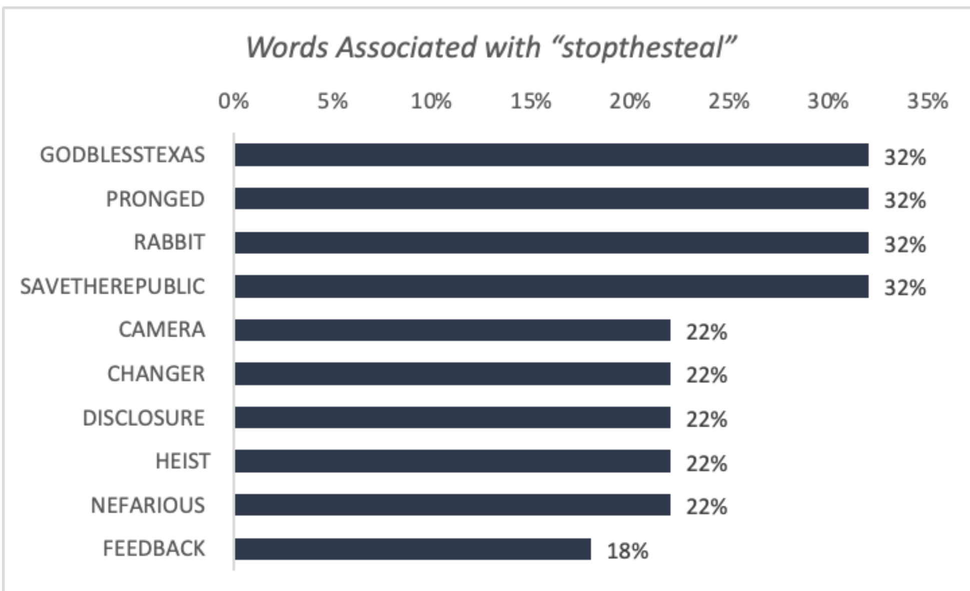
*Note: **Correlation is significant at the 0.01 level(2-tailed)*

There was mixed support for the first hypothesis. To test this hypothesis, we ran a correlation test between claims against the results of the 2020 election, and the partisanship of the author. This analysis required coding for partisanship, so we used our manually coded sample for this section as well. In the original codebook, coders were instructed to mark “republican,” “democrat,” “independent,” or “unsure.” A significant portion of the data had “unsure” author partisanship. When the correlation was run using the whole dataset, we obtained a Pearson’s correlation coefficient of |.18| (p <.001). This indicated that there was low correlation between partisanship of the author, and claims against the election results. However, when the correlation test was run using only tweets coded as “republican” or “democrat,” we obtained a Pearson’s correlation coefficient of |.70| (p<.001). This indicated a very strong correlation between the author's partisanship, and whether or not they agreed with the 2020 election results.

*Hypothesis 2*

**Figure 6**

*Word associations with "stopthesteal"*




*Note: The axis data labels represent the percentage frequency of the degree of association for each word, ranked from highest to lowest.*

The second hypothesis required a sentiment analysis of tweets using the hashtag "#StoptheSteal" and the hashtag "#Biden2020." We intended to conduct a sentiment analysis of each of the dataframes, and then perform a means comparison to identify which data frame had a more negative overall sentiment. Unfortunately, the full dataset only returned five tweets using "#Biden2020," and ten tweets using "#StoptheSteal." This was unfortunately not enough data to do a proper analysis.

Despite the fact that a full analysis was not possible, we performed a word association test on the term "stopthesteal" to see what other terms authors used in the same tweet. The phrases "godblesstexas" and "savetherepublic" both occurred in 32% of the tweets. "Heist" and "nefarious" were also prominent associations, both occurring 22% of the time. Another word association test was run to better understand "rabbit," which occurred 32% of the time. This revealed strong associations with "hole" and "chase."

*Hypothesis 3*

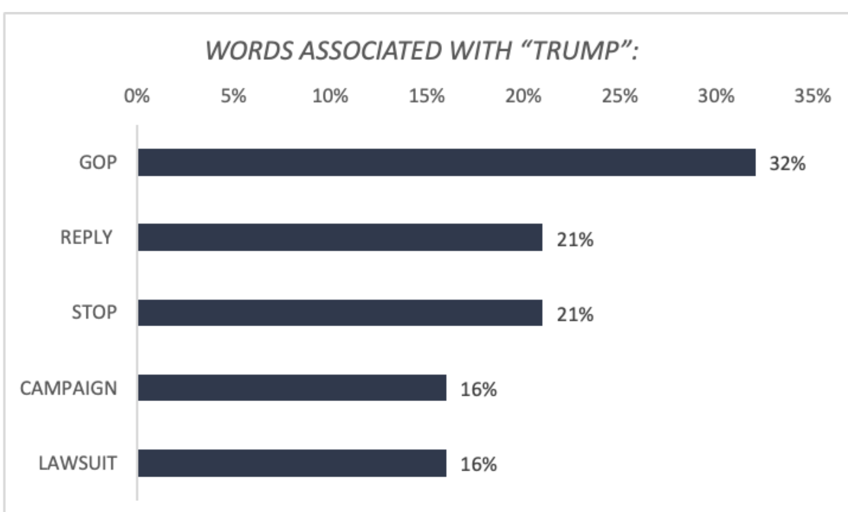
**Table 6**

Descriptive statistics

| Tweets mentioning | N | Mean | SD | SE Mean |
|---|---|---|---|---|
| Biden | 6394 | .38 | .777 | .010 |
| Trump | 9793 | .43 | .798 | .008 |

**Figure 7**

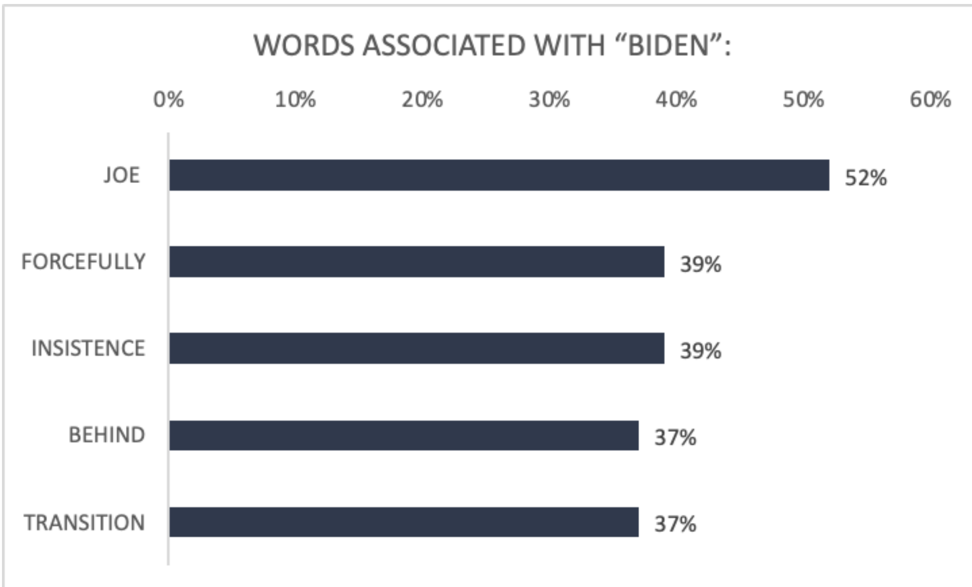
*Word associations with "trump"*



Note: *The axis data labels represent the percentage frequency of the degree of association for each word, ranked from highest to lowest.*

**Figure 8**

*Word associations with "biden"*




Note: *The axis data labels represent the percentage frequency of the degree of association for each word, ranked from highest to lowest.*

To analyze hypothesis three, we first created new data frames, one only including tweets mentioning "biden," and the other including only tweets mentioning "trump." These data frames were not exclusive, so there is overlap in tweets that mentioned both terms. A sentiment analysis was performed on both sets of tweets using a scale -1, 0, 1. We then compared the means of the results. Tweets mentioning "biden" had a mean of .38 ($p < .001$). Tweets mentioning "trump" had a mean .43 ($p < 00.1$). Tweets mentioning "trump" tended to be more positive than tweets mentioning "biden." For this reason, hypothesis three is unsupported.

## Discussion

In general, our research provided a limited look at the public conversation between the 2020 presidential election and the Capitol insurrection. A large part of our sample was retweets, which might reflect social standards of Twitter's verified users or might be a unique occurrence.

The data sample was also unequally balanced across the time range. In the weeks immediately after the election, we had almost twice as many tweets as in the weeks leading up to the insurrection. Further research might show if this pattern was consistent across unverified users. Extending the time range beyond January 6, 2021 might have also changed our results.

In our topic model, tweets were organized around topics that seemed to reflect the election results in specific states, including Michigan, Pennsylvania, and Georgia. After Election Day on November 6, 2020, the presidential results were called into question, including in these three states. Discourse on election fraud continued over the next few months, leading up to the scheduled results certification in Congress on January 6. Lawsuits were filed across the country, and there were multiple key points of contention. Therefore, the results of the topic model were consistent with expectations. Post-hoc research could be done on expanding these clusters and further investigating them.

Our RQ1 was on the relationship between support for claims of election fraud and support for Trump's 2020 candidacy. Technical issues with ideological scaling in R required manual coding for this analysis. The results showed that tweets by Republican authors on average questioned the election results more than tweets by Democratic authors. Post-hoc work correcting the technical issues would be ideal to achieve a better understanding of the entire sample.

Our H1 had mixed support, likely due to the technical issues with ideological scaling. The manual coded random sample was used in this analysis, as well, and coders were provided with a category for “unsure” partisanship or “other.” When tweets with this code were included in the analysis, the correlation was weak. However, when removed, the correlation was significantly strong. This suggests that partisanship is closely aligned with whether a tweet author believed the election was invalid. As mentioned previously, post-hoc work correcting the technical issues might provide a more comprehensive answer for this relationship.

Unfortunately, our H2 could not be supported due to a lack of available data. There were not enough tweets containing hashtags in our set to provide a worthwhile analysis. From a broader perspective, the lack of hashtags featured in our tweets suggests verified accounts might utilize the feature differently than unverified accounts. Future studies should extend beyond verified accounts to explore the relationship between hashtags and political ideologies.

The result most surprising to us was H3. Our hypothesized relationship between sentiment and “trump” and “biden” keywords turned out to be incorrect. Tweets mentioning “trump” tended to be more positive than tweets mentioning “biden.” Post-hoc research taking a more detailed look at the content of both datasets might provide clarity. Our initial thought is that tweets mentioning “biden” contained more accusations of fraud and election rigging.

These results further the discussion surrounding political partisanship and activity on social media. Social media has become a popular platform for political figures to connect with constituents, from representatives to presidential candidates. This analysis helps to uncover the key phrases and conversations used by political figureheads, and how they use either positive or negative sentiment to express their platforms. Our study only determined the Twitter activity of a specific subset of verified accounts. Future research could examine how the average Twitter user differs from verified accounts in terms of hashtag uses, overall sentiment, or common words and phrases.

*Limitations*

This study has a few limitations that provide opportunities for additional research. First, our analysis found that a few tweets were retweeted hundreds of times. The presence of dozens of duplicate tweets may skew the results of our sentiment analysis and not reflect the true public sentiment in a broader sample of tweets in the same time frame. Likewise, sentiment analysis tools may have misinterpreted any sarcastic, ironic, or humorous messages that were part of the retweets, which could also provide inaccurate sentiments for some retweet messages. Additionally, without a full scaling for the partisanship of the authors, it was challenging to determine whether the negative sentiment in tweets that featured “biden” was directed at President Biden. More context is necessary to fully understand the sentiment analysis.

Second, our research only examined verified accounts of celebrities and elected officials, which offers a limited view into the true online landscape surrounding the election. Due to time constraints, a list of select celebrity and national elected official Twitter accounts had to be developed, and a narrowed scope had to be adopted to ascertain a manageable population of tweets. It is likely that this subset of the population has different habits on social media than the average person.

Lastly, our analysis only examined tweets from, or mentioning, a selected group of celebrities and U.S. federal elected officials. The inclusion of a relatively small portion of Twitter’s population offers only a glimpse into the available digital content at this time. Prominent state and local officials, politically active celebrities, and other influencers with large followings who were not included in this study could have offered different views that may be closer to the broader public sentiment and perception of the election. Future research in each of these areas may reveal additional insights that further our understanding of this topic.

## Conclusion

In conclusion, our research and analysis indicate the important role partisanship and social media play in the expression of one's beliefs and in taking action online. The 2020 presidential election and the January 6th Capitol insurrection were significant historical events that had an enormous impact on political discourse, and they deserve dedicated study on the conversations leading up to, and following, these events to gather public perceptions. These two events may have a notable impact on future elections and/or policies, and social media channels could provide clues in advance that indicate a significant event may develop. Future research can provide key insights on how these events transpired and how they impacted the broader public.

## References

Berlinski, N., M. Doyle, A. Guess, G. Levy, B. Lyons, J. Montgomery, B. Nyhan, and J. Reifler. (2021). The effects of unsubstantiated claims of voter fraud on confidence in elections. *Journal of Experimental Political Science*, 1-16. https://doi.org/10.1017/XPS.2021.18.

Gaisbauer, F., A. Pournaki, S. Banisch, and E. Olbrich. (2021). Ideological differences in engagement in public debate on Twitter. *PLoS ONE, 16*(3). https://doi.org/10.1371/journal.pone.0249241

Minqing Hu and Bing Liu. "Mining and summarizing customer reviews." *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-2004, full paper)*, Seattle, Washington, USA, Aug 22-25, 2004.

Stolee, G., and S. Caton. (2018). Twitter, Trump, and the base: A shift to a new form of presidential talk? *Signs and Society, 6*(1). 147-165. https://doi.org/10.1086/694755