



An Analysis of the 2020 Presidential Candidates' Twitter Sentiment



Introduction

In 2020, the hottest topic besides Covid-19 was the US presidential election between President Donald J. Trump and former vice president Joe Biden. This election was full of history making firsts; It was the first election held during a worldwide pandemic, the first election to have 3 states whose margin of victory was under 1%, and the first incumbent president not to concede. Because of the uniqueness of this election, and the highly contrasting personalities between the two candidates, our team of data science students decided to analyze the sentiment of each candidate's tweets to find out how much their sentiment helped, or hindered, their election campaigns and how the general twitter population reacted to them.

In the analysis, we looked at the candidates by analyzing their sentiment via their social media presence and how they were received. Specifically, we investigated their tweet's sentiment and whether said sentiment correlated with the number of likes and retweets they receive. Since Trump and Biden are very different people with very different fanbases and followers, we wanted to see just how different the reactions to the attitudes of their tweets were within the twitter population.

We got the motivation for this topic from an article published by Cambridge University Press titled: "Differences in negativity bias underlie variations in political ideology". This article discusses how negative thoughts gain more attention and popularity than positive thoughts do. Negative thoughts also stay within our memory for a longer period. This also links to negative bias in politics, which triggered an idea to apply sentiment analysis in politics. The 2020 presidential election was the perfect area to focus this analysis on negative political bias.

We chose Twitter as the source of our analysis because it is the best platform to give us direct negative and positive thoughts that come directly from the candidates themselves. Through Twitter we also have some insight into the popularity and attention they receive by the number of impressions their tweets have.

In this analysis we will determine if Biden's and Trump's tweets evoke higher or lower impressions depending on the tweet's sentiment. With two very different personalities, the presidential candidates assumedly evoke different responses from people. Our hypothesis going into the project is that Trump's negative sentiment tweets get more likes and retweets, whereas Biden's positive sentiment tweets get more likes and retweets based on their media personalities and campaign advertisements.

Twitter overview

Before we begin the analysis, we will go over what Twitter is for those who do not use the platform so we can all be on the same page when talking about our methodology and findings.

Twitter, founded in 2006, is a social media platform where users interact using messages known as “tweets”. A tweet can be a sentence(s) that has a maximum length of 280 characters, a picture, or a short video. A tweet can contain a hashtag, denoted by the “#” symbol, to link it to a current event or idea, or can contain a tag, denoted by the “@” symbol, used to mention another twitter profile. Users can like a tweet, reply to a tweet with a comment, or do something called “retweet”. Retweeting is when a user tweets out a message that is just a link to a previous tweet made by another user. Lastly, users can follow other users to get notifications of new tweets from the users they are following.



Figure 1. Stock Twitter Profile. Obtained from <https://aea365.org/blog/dan-mcdonnell-on-setting-up-your-new-twitter-profile-page/>



Figure 2. Tweet example. Obtained from <https://blog.hubspot.com/blog/tabid/6307/bid/27675/how-to-retweet-the-right-way-in-4-easy-steps.aspx>

Candidate Twitter Background

Donald Trump has a much stronger presence on Social Media than Joe Biden. On Twitter alone, Trump has 88.9M followers and has made over 58K tweets. That is almost ten times more tweets and 80 million more followers than Biden has. On Facebook and Instagram, Trump also has a much larger following and presence than Biden. Trump also tweets at a much higher rate than Biden. He averages out at 20 tweets per day, while Biden usually has 6. As election day got closer, Trump was tweeting up to 800 times a month, which eclipsed Biden’s tweet rate.



Figure 3. Trump's and Biden's Twitter statistics. Obtained from <https://www.financialmirror.com/2020/10/29/trumps-social-media-army-7-times-bigger-than-bidens/>

Gathering the Data

The first step in our project was to gather as much data as we could from Trump and Biden so we could perform our sentiment analysis. The data used are the public tweets from their Twitter profiles. In order to gather said tweets, we needed to obtain permission from Twitter, which was given in the form of an API key used to extract all of their tweets from the source from September until election day. The tweets pulled from the Twitter API contain the following features:

- date - date and time the tweet was published
- tweet - a short message (less than 280 characters), a picture link, or a video link
- tweet_length - the number of characters a tweet has
- id - unique Identifier for the tweet
- likes - number of “likes” a tweet received by other users
- retweets - number of times the tweet was “retweeted” by other users

	date	tweet	tweet_length	id	likes	retweets
0	2020-11-12 19:20:00	I extend my deep condolences to the loved ones...	139	1326968002839908353	351008	26642
1	2020-11-12 18:30:00	Because of the Affordable Care Act:\n\n- Peopl...	140	1326955421320450049	82659	8509
2	2020-11-12 01:19:45	Ron Klain's deep, varied experience and capaci...	140	1326696150364286976	202370	16814
3	2020-11-11 21:30:00	This Veterans Day, I feel the full weight of t...	139	1326638331204038656	282464	19264
4	2020-11-11 16:08:11	Today, we honor the service of those who have ...	140	1326557341697839106	357665	32327

Figure 4. Raw data in Pandas DataFrame.

Data Modeling

We will start the data cleaning by importing the raw data, stored as a pkl file, into a Jupyter notebook environment. In this environment, we have the following libraries installed:

- tweepy – library used to store raw data from Twitter through their API with a key
- pandas – library used for data modeling
- numpy – statistical package used to process and analyze data
- matplotlib – visualization package used to explore data and make rudimentary charts
- pickle – library used to read raw pkl data into python
- textblob – Sentiment analysis library
- re – Regex library used to edit and clean text data

From this file, a Pandas DataFrame will be made, as shown in Figure 4. A little bit of feature engineering will be done to make the data better for the sentiment analysis. The first thing we will do is take out all the tweets where presidents retweet themselves, as we believe this to be duplicate data. Next, we will use Regex scripts to take out any hyperlinks in textual tweets and any tweets whose only content in a link to a video, photo, external website, or retweet. The reason for this is that the sentiment analyzer does not know how to interpret hyperlinks, so it's counted as noise.

At this point, the data is ready to be analyzed with Textblob's sentiment analyzer. Each tweet is ran through a Textblob script based on dictionary rules and NLP to give it a polarity rating between -1 and 1 and a subjectivity rating between 0 and 1. The polarity rating tells us how negative (-1) or positive (1) a tweet is. Meanwhile, the subjectivity rating tells us the likelihood of whether the tweet is an opinion or a fact, with 0 being an opinion and 1 being a fact.



Figure 5. Example of positive tweet.

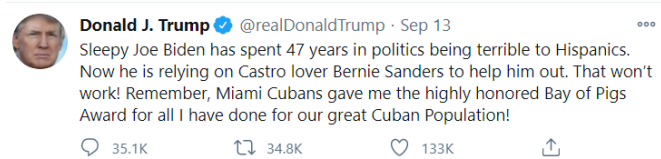


Figure 6. Example of negative tweet.

At this point in the problem, we ran into a bit of a problem, in particular with President Trump's tweets. As mentioned before, there was a huge discrepancy when it came to the number of tweets Trump made per day. This was messing with our sentiment analysis because a large amount of those tweets were only one or two words, retweets of someone else, retweets quoting himself, or tweets including a video or picture link as the majority of the text. Because of the low information and vagueness regarding these particular tweets, the sentiment analyzer didn't know how to interpret them, and ignored them or gave them a wrong score, making them skew the results. These tweets were usually very generic and unpopular compared to his more popular, lengthy tweets. To remedy this, we filtered his tweets to only include the top 70 percentile most liked tweets and his least subjective tweets, since we found that these two values correlated heavily with the useless tweets in the data. This process would get rid of the tweets that were useless to the sentiment analyzer, but still keep the most important data for the analysis. We ran the same cleaning function on Biden's tweets for parity, and surprisingly, they both ended up with around the same amount of tweets afterwards despite Trump having a huge lead in number beforehand.



Figure 7. Example of useless tweet treated as noise on the left, and useful tweet for the sentiment analyzer on the right.

Historical Average Sentiment

At this point our dataset is cleaned and each tweet has a sentiment associated with it. We've also filtered out all the useless data treated as noise from the dataset. The first sets of graphs we made were an average sentiment by data for Biden and Trump. We basically took the tweets made on each day and used them to evaluate how negative, positive, or neutral the candidate was on that given day. The x axis contains the days, and the y axis has the average sentiment of the tweets on that day. We'll start off with the graph containing Trump's historical tweet sentiment between the beginning of September until election day.

Daily Average Sentiment of Trump's tweets



Figure 8. Graph of Average Sentiment by day for Trump

This graph shows the average sentiment per day for Donald Trump, starting at the end of August until the end of October. As you can see, Trump was a mixed bag in the beginning of September but was mostly negative. He then started to quiet down by the end of September and was very bipolar in his tweets going into October, being very positive one day and very negative another day. By the end of October, as election day got closer, he started being overwhelmingly more positive in his tweets. Now we'll examine Biden's.

Daily Average Sentiment of Biden's tweets

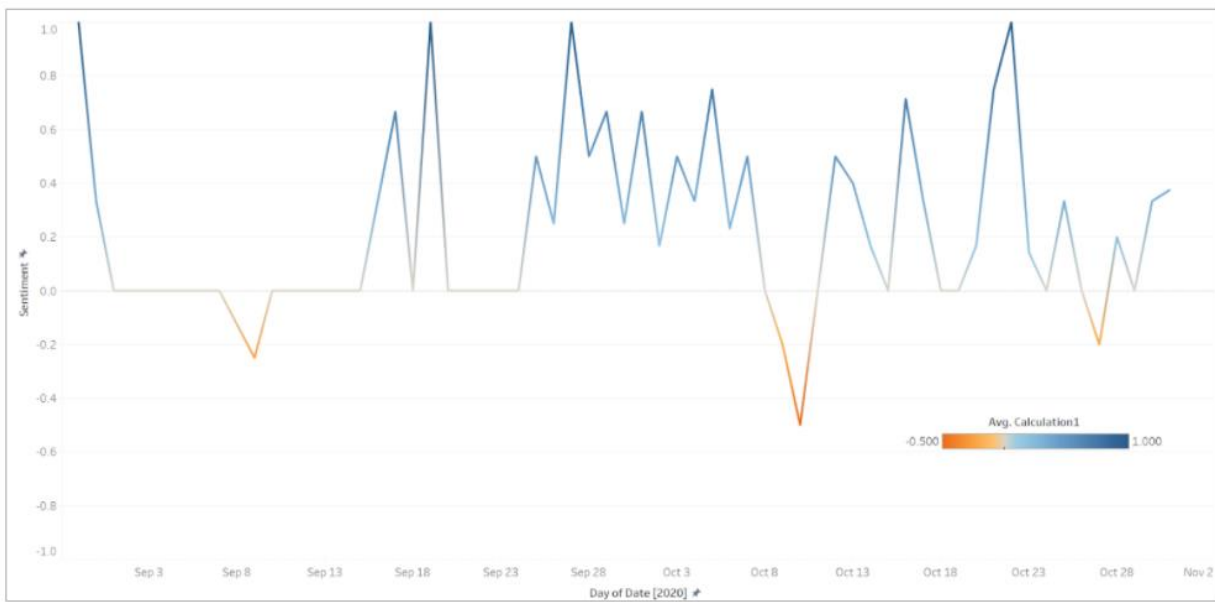


Figure 9. Graph of Average Sentiment by day for Biden.

Moving on to Biden, he was pretty neutral in the beginning of September and got more positive as the month continued. By late September and early October, he was overwhelmingly positive, with a few splashes of negativity on some days. By late October to early November, he was mostly positive or neutral in his sentiment.

Now, the thing that mostly surprised us from these graphs is just how positive Trump was in his tweets. Our team did not really follow politics too closely before this analysis, so we didn't know much about Trump's personality other than his scandals, and we assumed most of his twitter would be littered with negativity. Because of this twist we did not see coming, we decided to investigate the tweet data more closely.

This data told us a few things we already knew about the candidate's twitter habits. Trump is very negative and almost bipolar in the way he posts, but also very positive, and Biden is more positive or neutral, with generic non offensive tweets making up the most of his tweets. We were surprised that Trump's tweets were almost equally as positive as he was negative. We decided to investigate further and found that the reason he has so many tweets that are positive is because he constantly tweets about himself and what a good job he is doing, which the sentiment analyzer marks as positive in nature. Meanwhile, Biden's positive tweets are more general and about other people rather than himself. While tweets bragging about yourself and your perceived accomplishments are indeed categorized as positive in nature, it's still interesting that it was these types of tweets that made the bulk of Trump's positivity.



Figure 10. Tweets often categorized as positive for Trump.

Most Liked and Retweeted tweets by sentiment

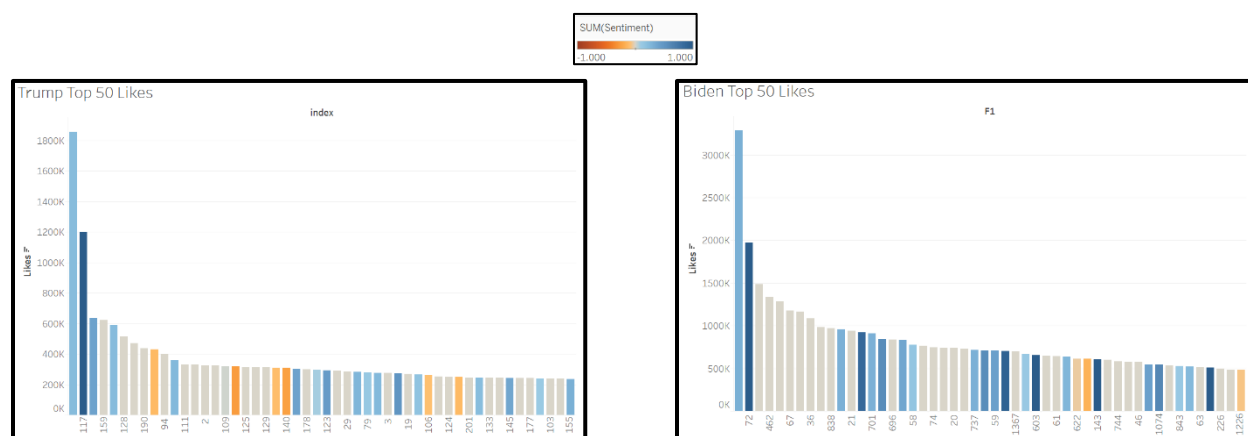


Figure 11. Trump and Biden's most liked tweets by sentiment.

Of Trump's 50 most liked tweets, 54% are positive, 26% are neutral, and 20% are negative. Trump's most popular tweet received 1.1M likes. One of the most popular tweets is his tweet about testing positive for Covid-19: "Tonight @FLOTUS and I tested positive for covid-19. We will begin our quarantine and Recovery process immediately. We will get through this together".

Compared to Biden's 50 most liked tweets, 58% are positive, 32% are neutral, and 10% are negative. His current most popular tweet is his acceptance of winning the presidency and received 3.3M likes: "America, I'm honored that you have chosen me to lead our great country. The work ahead of us will be hard, but I promise you this: I will be a President for all Americans — whether you voted for me or not. I will keep the faith that you have placed in me." Biden has a slightly higher percentage of positive tweets in his top 50 most liked, and half the amount of negative tweets that Trump has.

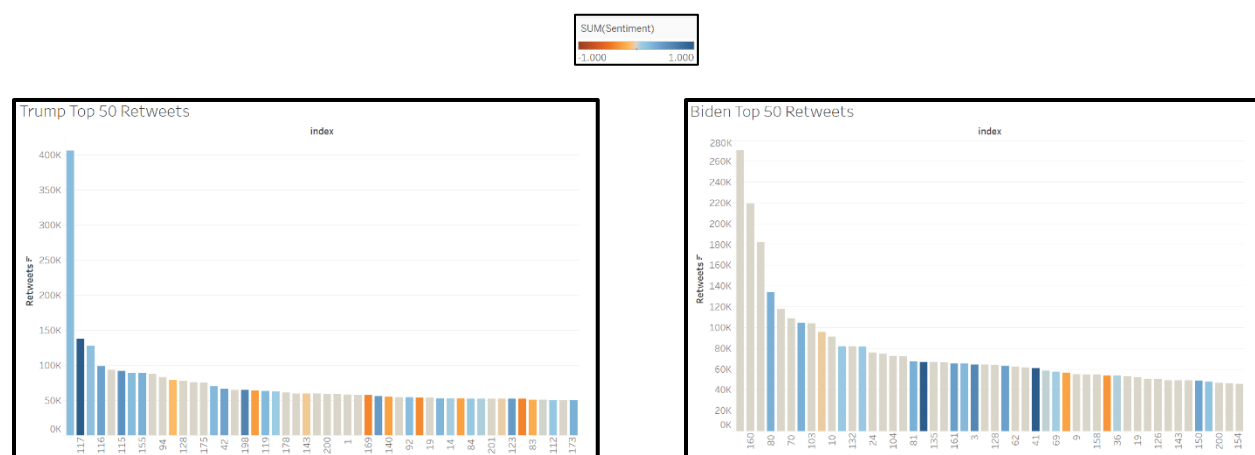


Figure 12. Trump's and Biden's top 50 most retweeted tweets by sentiment.

Looking at Trump's 50 most retweeted tweets, 20% are negative, 42% are positive, and 38% are neutral. The top 10 liked tweets have no negative tweets at all. When compared to Biden's top 50 retweeted tweets, 6% are negative, 32% are positive, and 62% are neutral. Trump's most retweeted tweet has a positive sentiment with over 400k retweets. While the 2nd and 3rd highest have less than half the amount at 170k and 140k retweets, respectively. The number of Biden's retweets range between 60K to 270K retweets. Whereas Trump's has a much wider range of 70k to 400k.

To summarize the sets of graphs above, both Trump's and Biden's 50 most liked tweets were overwhelmingly neutral or positive, at a ratio of 80%+. Trump had 12% negativity compared to Biden's 6%, but overall, both candidates were almost equal in their most liked tweets. For the top 50 most retweeted tweets, Biden was overwhelmingly neutral and positive, at a 90%+ ratio of positive and neutral tweets. Trump was equally neutral and positive, around 40% each, but was noticeably higher than Biden in the negative department (20%).

What this shows us is that people tend to like Trump's and Bidens positive and neutral tweets at an equal rate. But when it comes to retweets, Trump's negativity is retweeted at a much higher rate than Biden. This may be because people who aren't Trump supporters retweet

his most scandalous tweets to hold it against him but refuse to like the tweet. Below is an image of some of the most retweeted tweets from Biden and Trump. Reading the retweets and the comments showed a large number of users criticizing Trump while mostly praising Biden, adding to the theory that Trump's negativity is retweeted at a much higher rate.



Figure 13. Trump's and Biden's most retweeted.

Conclusion

By the end of the project our research question was answered, but not in the way we thought. Trump and Biden do elicit different responses from their sentiments, since Trump gets a lot more response from his negative tweets from people than Biden does. However, Trump also gets a lot of responses from his positive tweets, like Biden. We had to look further to see what positivity meant to Trump to see why this was and found that he's mostly focused on himself and his fanbase when being positive while Biden focuses more on everyone when being positive. We also found that Trump's tweets are much more opinionated than Biden's, which talk more about facts.

Trump being an outrageous character has certainly made him more popular and got him more impressions than Biden on twitter, as people respond more to that type of behavior. He gets an equal amount of likes on average as Biden by being equally nasty towards others and positive about himself, but his retweets sentiment show that his negative tweets are talked about a lot by people who do not like him rather than just his fanbase. This would be fine if no publicity were bad publicity, but as we saw by the election results, that was not the case this time and he may have turned too many people off from his platform by being too negative and having that negativity be spread through the twitter user base against him. This should serve as a reminder to future politicians looking to emulate Trump's unusual personality, that there is a limit to how nasty you can be for attention before it starts to backfire.

Appendix: Challenges and Limitations

When performing our project, we ran into three big challenges that we had to compromise with or make assumptions on. First, we found that many tweets had inaccurate sentiment scores. For example, Biden's most liked tweet about winning the election was scored as a neutral tweet, but we would call that a positive sentiment tweet. This shows that the sentiment analyzer is not entirely accurate only using dictionary rules and NLP. Because of the limited time we had for this project, we had to compromise and use a sentiment analyzer focused on those aspects rather than make one by training a machine learning model to come up with sentiment. Twitter uses a lot of slang and newer diction and figures of speech, which may have been lost in Texblob's analyzer.



Figure 14. Biden's tweet confused the Sentiment Analyzer.

The second issue we ran into was that we only had access to tweets that occurred in the last 3 months, so it is possible that the last 3 months tweets do not accurately represent their tweet behavior in general. We made the assumption that, given the last 3 months current events, tweets would have more extreme sentiments than usual in this time period and would make up for the lack of historical data.

The last limitation is that we didn't have realistic access to everyone's twitter data, just the two presidents, so we couldn't do an in depth analysis on the general twitter population's sentiment to the candidates or the sentiment of the tweets retweeted from the candidates. Twitter has a limit on how much data you can scrap off their website ever since the Cambridge Analytica Facebook incident, so we decided to focus on the two candidates rather than a large population, since twitter would not have allowed us to scrape data of so many of their users.

Appendix: Twitter user statistics

Lastly, since we were doing an analysis on twitter data, we thought it would be a good idea to figure out how useful this was to other similar minded people. So, we started researching how much twitter represented the general population.

We found that U.S. Twitter users were not exactly a fair representation of the U.S. citizens. Twitter users tend to be younger, but more highly educated and have higher incomes than

average U.S. adults. Twitter users are more likely to identify publicly as Democrats as opposed to Republicans. The biggest takeaway is that, as far as twitter activity, most users rarely tweet, but the most prolific 10% users create 80% of the tweets. Of the 10% who tweet most often, most are women, and tend to focus more on politics, but this probably because of the hectic previous 4 years.

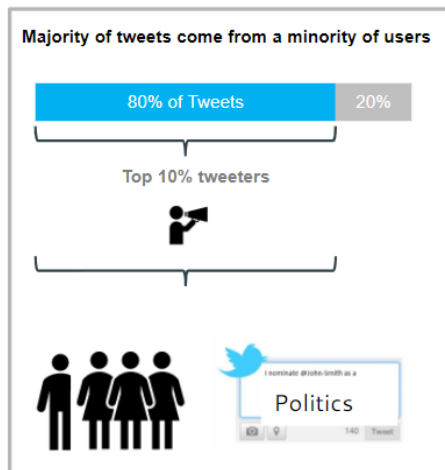


Figure 16. Twitter statistics. taken from <https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/>

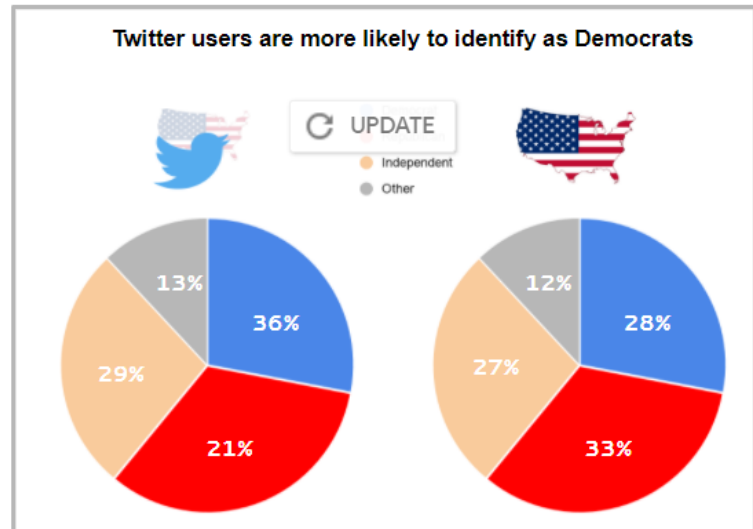


Figure 15. Twitter statistics. taken from <https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/>

So even though Trump and Biden get so many likes and retweets and twitter can give us an idea of a fraction of the US population, it alone is not enough to get the full picture of how the president's sentiment causes people to react when talking about the general US population or how to completely predict an election's results.

Authors

This project was done by a team of data science students from the University of the Pacific's Data Science MS program. Our team consisted of **Theerawat Jindapoo**, **Jereddy Manoj**, **Chirapa Muadchan**, **Rebekah Stottlemeyer**, and **Jaime Vargas Diaz**.

Works Cited

<https://aea365.org/blog/dan-mcdonnell-on-setting-up-your-new-twitter-profile-page/>

<https://blog.hubspot.com/blog/tabid/6307/bid/27675/how-to-retweet-the-right-way-in-4-easy-steps.aspx>

<https://www.verywellmind.com/negative-bias-4589618#citation-3>

<https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/>

<https://www.state.gov/elections-101-the-role-of-social-media-in-us-elections>

<https://twitter.com/realDonaldTrump>

<https://twitter.com/JoeBiden>