# Political Popularity of Misinformation

**Catherine Tao, Aaron Chan, Matthew Sao**

*University of California, San Diego*

March 7, 2021

## 1 Abstract

For our research on Political Popularity of Misinformation, we want to research the influence politicians have on Twitter, a well known social media platform for users to voice their opinions to a wider audience. The information shared on Twitter that we are interested in will be grouped into scientific information or misinformation. Politicians can easily sway public opinion with a simple tweet, therefore we wanted to analyze how much they influence other Twitter users.

We gathered ten politicians who we considered to spread scientific information on Twitter and ten politicians who we considered to spread misinformation on Twitter. We analyze the two groups to show how controversial a tweet appears. We do this by looking at tweet engagement as well as a popularity metrics to see growth over time.

The results of our investigation showed that politicians who spread misinformation have a higher ratio value on average and have less overall likes over their tweets. Our permutation tests shows that our scientific group has been consistently growing and increasing in growth over time. In contrast, our misinformation group has grown significantly, but only in the more recent years. Overall, our results show that a politician can experience the most growth through spreading non-controversial, scientific information.

## 2 Introduction

The rise of the internet and easily accessible and instantaneous information in the recent century has caused a significant change in the way that the public ingests their news. In a large part, this shift to instantaneous public information has allowed this generation to be the most informed that it has ever been, but also the most opinionated and misconstrued.

Social media has become the main source of easily accessible and digestible information for field experts and organizations to publicly spread news, but at the same time it has become a place where individuals can spread their beliefs as fact and influence others' opinions on subjects that readers have yet to be informed about. As the internet is a place open for anyone to share information, the validity of information presented is not always guaranteed to be accurate or benevolent.

For our research on Political Popularity of Misinformation, we want to analyze the growth of politicians on Twitter, a well known social media platform for users to voice their opinions to a wider audience. The information shared on Twitter that we are interested in will be grouped into scientific information or misinformation. We have chosen ten politicians to represent our scientific group and another ten politicians to represent our misinformation group. This specific analysis is interesting because we are able to determine how the content of a politician's tweet affects their growth on Twitter.

Throughout our investigation, we used mathe-

matical methods in order to analyze engagement of the tweets and to compare our two groups. The ratio metric is used to analyze engagement of a politician's tweets. This method takes in account the retweets, likes, and comments of a specific tweet. We estimate following and growth using a politician's likes for each tweet over time. Finally, we used permutation tests in order to compare our two sample groups to draw conclusions.

Data visualizations are shown to illustrate the technical findings into a visual representation where we can view trends and patterns. The graphs shown are a way to compare different groups of politicians.

[1] Many of the politician's tweet IDs were gathered from a third party source which stores all individuals holding office from the Senate and Congress. The starting tweets for each individual varies depending how long they have been actively tweeting on their specified Twitter account.

## 3 Data Collection

Our data consists of a collection of tweets for each individual politician, also known as their timeline. We obtain the tweet IDs that compose our politicians' timeline from George Washington University's TweetSets database. The TweetSets database have datasets consisting of tweets for research and archival purposes, covering a wide range of topics such as climate change, the 2018 Winter Olympics, the two most recent presidential elections as well as tweets made by politicians of the 115th and 116th Congress.

For our analysis, we chose to focus on politicians who served in the 116th United States Congress, which corresponds to two datasets, Congress: Representatives of the 116th Congress and Congress: Senators of the 116th Congress. We specifically chose the 116th Congress as it is the most recently concluded session at the time of writing. The two datasets combined contain 2,756,042 tweet IDs and were collected between January 27, 2019 and May 7, 2020 from Twitter's API using Social Feed Manager. [2] The earliest tweet in this dataset relevant to our project occurred on December 16, 2008 while the last tweet was made on May 5, 2020. It is worth noting that not all of the politicians have tweets spanning all years. This is a result of some politicians having just been recently elected to Congress, such as Alexandria Ocasio-Cortez whose first term was the 116th Congress.

To start our data collection process, we first identified twenty politicians, ten of which we believe to spread misinformation during their time in office and ten which we believe to spread scientific information. In order to classify a politician as someone who spreads misinformation we researched notable current politicians and justified their classification through reports and news articles detailing their statements on topics ranging from the coronavirus to the most recent election. [3] [4] For example, Senator Joni Ernst, who falsely claimed that healthcare providers are inflating the number of coronavirus cases, or Representative Matt Gaetz, who falsely claimed that Antifa members were part of the riots on Capitol Hill. To classify a politician as scientific we identified current politicians who often tweet out scientific information such as Representative Lauren Underwood, a former nurse who regularly tweets and retweets information about the coronavirus.

After identifying our politicians, we gathered the user IDs for their Twitter accounts using an online Python library Tweepy, which we then used to query the two Congressional datasets. We use a politician's user ID as opposed to their username because a politician's username may change over time while their user ID remains constant. The datasets also contain a file of the House and Senate members along with their user IDs which is an alternative way to obtain these IDs. To query the datasets, for each politician, we selected either the Representative or Senator dataset depending on their position and inputted their user ID in the "Contains any user id" box under the "Posted by" section. This process gives us a text file of tweet IDs for each politician which we then rehydrate using Twarc which is a API used for accessing archived Twitter JSON data. The output is a JSON file for each politician that contains tweet objects returned by Twitter's API. The average number of tweets for our scientific politicians is 4,563 while the average number of tweets for our misinformation politician is 5,446.

In order for us to calculate a tweet's ratio, we need to have information about the number of times a tweet has been replied to. Unfortunately, we are not able to access the reply_count attribute on a Tweet object without the Premium or Enterprise tier of Twitter's API. As an alternative, we make cURL calls to the Twitter API's Metrics field, which allows us to access engagement metrics for

Tweet objects. For each politician, we use cURL to request a tweet's retweet, likes and reply counts and save the output into a csv. At the end of our entire data collection process, each politician has a txt containing their Tweet IDs, a JSON file containing their Tweet data, and csv file containing likes, replies, and comments.

# 4 Methods

For this section, we discuss the three different methods we use to analyze and draw conclusions to our results. The three methods include the ratio metric, popularity estimates, and permutation tests.

## 4.1 Ratio Metric

We analyze the community engagement by using a ratio metric. This method incorporates the number of likes, retweets, and comments a specified tweet holds. We define an equation to measure the amount of community engagement with the given numbers from each tweet. A high ratio will generally mean the Tweet has received a negative reaction whereas a low ratio would indicate a positive or neutral reaction. We intend to track the reaction of each tweet a politician tweets over time to see the politician's overall approval.

To analyze reception to a particular tweet, we chose to use the concept of ratios or "getting ratioed" on Twitter. This is the number of replies compared to the number of likes and retweets a tweet receives. Ratios allow for a quantitative way to measure how controversial a tweet is, with higher ratios signaling a more disputed tweet. [5] We formally define our measure of ratio below.

$$\frac{2 * \#\ of\ replies}{\#\ of\ likes + \#\ of\ retweets} \tag{1}$$

We decide to weigh comments negatively because both the like and retweet function of a Tweet are used as ways to indicate approval or agreement. Although comments can also contain positive feedback, a large amount of comments compared to a smaller number of likes and retweets generally indicate that the Tweet was not well received.

We weigh comments more heavily than likes and retweets due to the increased amount of effort it takes to write out a reply to a tweet as opposed to liking or retweeting that same tweet.

The ratios for tweets per politician are averaged for each politician in order to determine their average ratio. Each average ratio does not include days where a politician does not tweet because the ratio would result in an undefined value since the likes, comments, and retweets would be zero. These tweets are removed from the other ratios in order to prevent skewing of a politician's average ratio result.

### 4.1.1 Ratio Metric: Data Visualizations

In Figure 1, we graphed our ten politicians we grouped as scientific. As we can see from the graph, Lisa Murkowski and Mitt Romney have the highest ratios compared to our other eight politicians grouped under scientific politicians. It is interesting to note that these two politicians represent the Republican party, while our other eight represent the Democratic Party.

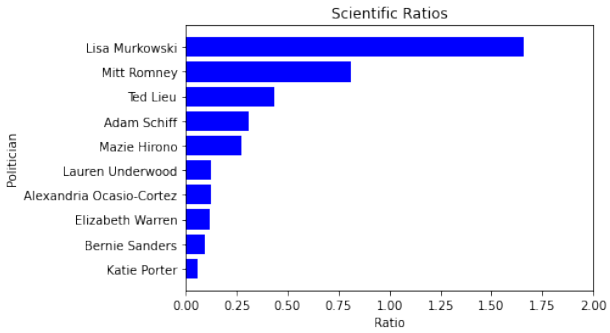


**Figure 1:** *Shows a horizontal bar graph for the ten politicians grouped under scientific. Represents the averaged ratios for each politician.*

Figure 2 graphs the ratios for the politicians in our misinformation group. One interesting finding is that Tulsi Gabbard, who is the only Democrat of the misinformation group has the lowest ratio, meaning that her average tweet engagement is overall positive. The other nine politicians are representatives of the Republican party. The margin of difference for each politician is not overly extensive in comparison to the Scientific Ratio graph. As seen in Figure 2, Lindsey Graham, Matt Gaetz, and Joni Ernst are the three politicians with the highest ratios, indicating that their tweet engagement is relatively negative.

## 4.2 Popularity Metrics

Since Twitter does not provide data on the number of followers a user has at a given time, we find a

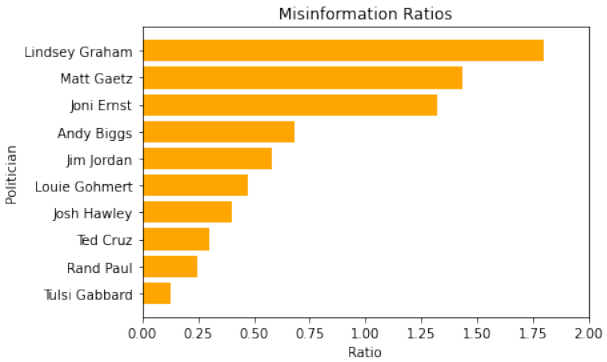


**Figure 2:** *Shows a horizontal bar graph for the ten politicians grouped under misinformation. Represents the averaged ratios for each politician.*

different way to estimate a politician's following and growth. Using the likes that we got from each tweet for each politician, we created metrics as a way to estimate the following that each politician has and gains over time.

These metrics are tracked in two ways: over time and over activity. Over time measures the likes that a politician receives for each month. We decided on the time frame of one month because we believed that any smaller period of time would not be a large enough time period to indicate any significant growth in following. Over activity measures the likes that each politician gets for each tweet. It is important to note that the politicians may have a difference in start date as well as a difference in number of total tweets depending on the frequency at which they Tweet at.

The metrics are further divided into either cumulative or rolling. Cumulative builds on the previous amount of likes. This is useful to measure the politician who generated the most following to find who may be the "most popular" politician. Our rolling metrics take into account a period of time or a number of tweets to aggregate on as a way to see how popular politicians are at a given moment of time while also taking into account some recency. All rolling periods are trailing to account for their recent tweets rather than future ones.

Our rolling metrics can be split again into max and average. Max will mark the max amount of likes of a tweet over the trailing window. Average will take the average amount of likes for the tweets over the window. The window size can be adjusted for both our over time metrics and over activity/tweets metrics.

### 4.2.1 Popularity Metrics: Data Visualizations

For some graphs, the graphs analyze the likes of tweets collected over time while others analyze the likes collected over number of tweets. The graphs showing tweets over time have a window size of 4 months for their tweets. Depending on the graph, the x-axis is tweets over time or total number of tweets while the y-axis shows the average, max, or cumulative number of likes. Each graph allows us to visualize trends and patterns between each group. We are able to make interesting findings between our two groups as well as individuals.

For the first data visualization for this metric, we show the average number of likes per month for two politicians from our scientific group. We chose the politician with the highest and lowest average ratio values in order to view any key differences from the scientific group.

In Figure 3, we see that Representative Katie Porter has many more average number of likes per month in comparison to Senator Lisa Murkowski. This shows a major range difference between these two politicians under our scientific group. Katie Porter's number of average likes per month exceeds Lisa Murkowski by a huge margin, indicating her larger following.

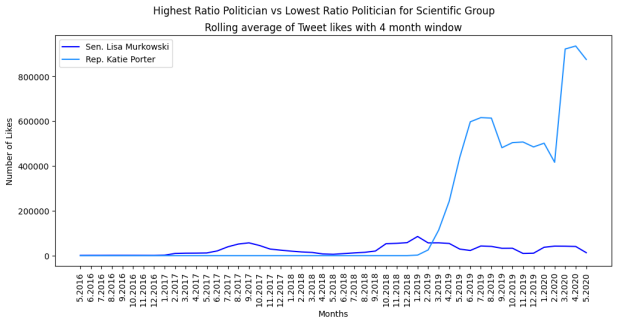


**Figure 3:** *This graph shows the average number of likes per month for the politicians Lisa Murkowski and Katie Porter. Both of these politicians are within our scientific group.*

In Figure 4, we analyze two representatives under our misinformation group which include Lindsey Graham and Tulsi Gabbard. For this graph, it is clear that Gabbard has less average likes per month in comparison to Graham.

For Figure 5, we compare two very popular politicians from our two groups. Alexandria Ocasio-Cortez represents our scientific group and

# Political Popularity of Misinformation

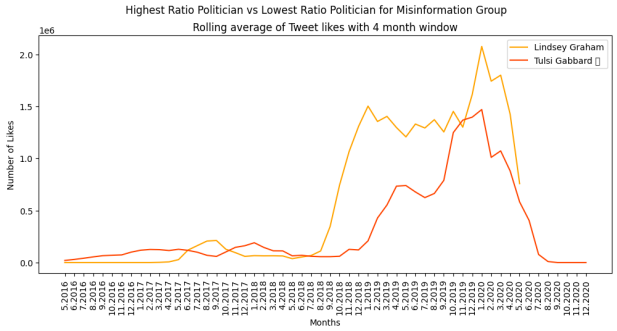


**Figure 4:** *This graph shows the average number of likes per month for the politicians Lindsey Graham and Tulsi Gabbard. Both of these politicians are within our misinformation group.*

Ted Cruz represents our misinformation group. We see that Ocasio-Cortez has a much larger number of maximum likes over the four month window in comparison to Ted Cruz.

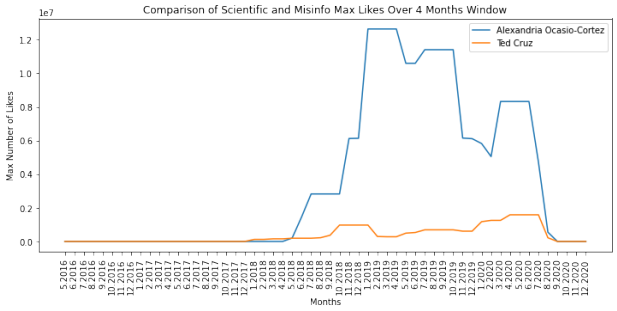


**Figure 5:** *This graph compares the rolling maximum number of likes for Alexandria Ocasio-Cortez and Ted Cruz.*

For Figure 6, we compare the highest ratio politicians from our scientific and misinformation groups. We found that Senator Lisa Murkowski had the highest ratio from our scientific group and Lindsey Graham had the highest ratio from our misinformation group. In this data visualization we can clearly see that Lindsey Graham had a much higher ratio trend for number of average rolling tweet likes over a four month window. Murkowski's trend line does not appear to grow as dramatically.

Figure 7 analyzes the top politicians with the highest number of rolling average total likes from our scientific and misinformation group. We found that Alexandria Ocasio-Cortez had the highest number of rolling likes from our scientific group while Jim Jordan had the most from our misinformation group. Within this graph, we see that Ocasio-Cortez has an overall larger number of

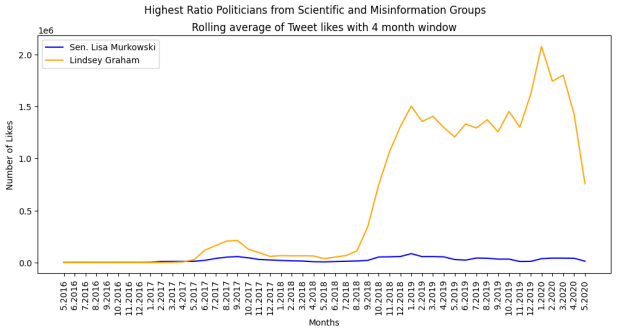


**Figure 6:** *The graph shows a comparison of rolling average likes over a 4 month window for Lisa Murkowski and Lindsey Graham.*

rolling likes compared to Jordan. This graph is significant because we can see the difference in total likes that politicians from our scientific group have our misinformation group.

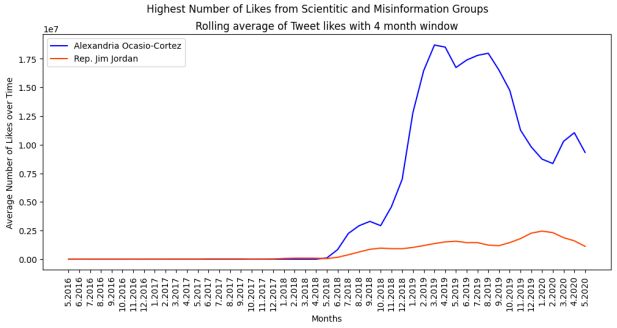


**Figure 7:** *This graph compares the highest number of likes from our scientific and misinformation groups. We compare Alexandria Ocasio-Cortez and Jim Jordan.*

For our Figure 8, we wanted to compare the top politicians from each of our groups with the highest number of tweets over a window size of 200. After analyzing, we found that Alexandria Ocasio-Cortez and Jim Jordan were once again our top two politicians to compare for most likes on their tweets. In this graph, we see that even with a window size of 200, Ocasio-Cortez has a higher trend line on the graph, indicating that she consistently exceeds the average number of likes over politician Jim Jordan.

For our final graph, we wanted to take the median number of likes per month for both groups to compare both groups as a whole. As we can see, the scientific group exceeds the number of median likes compared to our misinformation group. This supports our argument of politicians who spread scientific information on Twitter have more likes overall compared to those who spread misinfor-

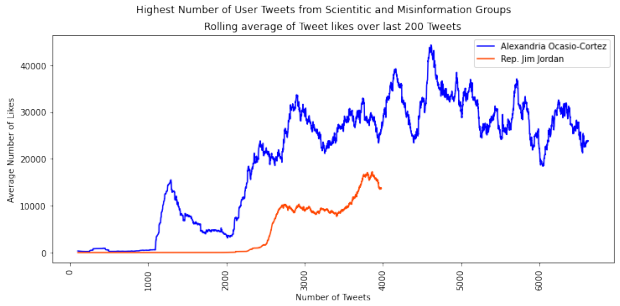


**Figure 8:** *This graph compares the highest number of likes for politicians Alexandria Ocasio-Cortez with Jim Jordan over a 200 tweet window size.*

mation on Twitter.

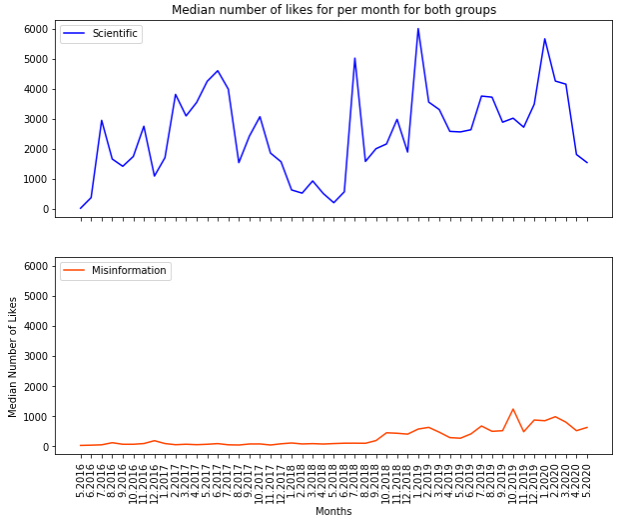


**Figure 9:** *The graph compares the median number of likes per month for both our scientific and misinformation groups.*

## 4.3 Permutation Test

In order to actually see if the popularity of the groups are changing over time, we run multiple permutation tests. A permutation test takes in two samples and determines the chance that these samples come from the same population. By running this test on our likes for our two groups or politicians, we can see how similar in popularity our groups are or if they are different. The distribution that we run our test on is the normalized likes per year for both groups. Each tweet's likes in the current year that we are looking at is subtracted and divided by the mean likes of the previous year. This way we are able to measure the growth rather than raw numbers.

Our null hypothesis and alternative hypothesis are as follows:

*Null Hypothesis*: The growth of likes for our misinformation group is the same as the growth if likes for our scientific group over each year.

*Alternative Hypothesis*: The growth of likes for our misinformation group will be different from our scientific group over each year.

For this process, we normalize the growth of likes for each year by calculating the percentage growth from the previous year. We run three main permutation tests to determine comparison of growth. To compare the scientific and misinformation groups, we run a permutation test over each year comparing the distribution of normalized likes for each group. For example we compare Scientific 2015 vs Misinformation 2015 or Scientific 2016 vs Misinformation 2016. This will allow us to see how the growth of the two groups compare with each other. After applying a Bonferroni correction due to running multiple hypothesis tests which gives an alpha of 0.05 / 8, we still find that all years show significance except for 2017. This indicates that the only year in which our scientific group and misinformation group's growth matched was during 2017.

```
Scientific vs Misinformation permutation test for year 2012: the p-value is  9.999000099990002e-05.
Scientific vs Misinformation permutation test for year 2013: the p-value is  9.999000099990002e-05.
Scientific vs Misinformation permutation test for year 2014: the p-value is  0.0026997300269973002.
Scientific vs Misinformation permutation test for year 2015: the p-value is  9.999000099990002e-05.
Scientific vs Misinformation permutation test for year 2016: the p-value is  9.999000099990002e-05.
Scientific vs Misinformation permutation test for year 2017: the p-value is  0.47325267473252675.
Scientific vs Misinformation permutation test for year 2018: the p-value is  9.999000099990002e-05.
Scientific vs Misinformation permutation test for year 2019: the p-value is  9.999000099990002e-05.
Scientific vs Misinformation permutation test for year 2020: the p-value is  9.999000099990002e-05.
```

**Figure 10:** *This graph shows our results for the scientific versus misinformation groups for our permutation test.*

We then run two more permutation tests for the two groups themselves. This test is on consecutive years and is meant to show us if growth for each of the groups is increasing or stagnating. For example Scientific 2015 vs Scientific 2016 and Misinformation 2015 vs Misinformation 2016.

For these tests, we find that our scientific group shows stagnated growth for the years 2013 to 2014. This indicates that the scientific group consistently is growing more and more compared to its previous years.

The misinformation group shows stagnated years for 2013 to 2014, 2014 to 2015, and 2017 to 2018. Showing that they are not increasing in following as often.

```
Scientific 2011 vs Scientific 2012 permutation test: the p-value is 9.999000099990002e-05.
Scientific 2012 vs Scientific 2013 permutation test: the p-value is 9.999000099990002e-05.
Scientific 2013 vs Scientific 2014 permutation test: the p-value is 0.9396060393960604.
Scientific 2014 vs Scientific 2015 permutation test: the p-value is 9.999000099990002e-05.
Scientific 2015 vs Scientific 2016 permutation test: the p-value is 9.999000099990002e-05.
Scientific 2016 vs Scientific 2017 permutation test: the p-value is 9.999000099990002e-05.
Scientific 2017 vs Scientific 2018 permutation test: the p-value is 9.999000099990002e-05.
Scientific 2018 vs Scientific 2019 permutation test: the p-value is 9.999000099990002e-05.
Scientific 2019 vs Scientific 2020 permutation test: the p-value is 9.999000099990002e-05.
```

**Figure 11:** *This graph shows our results for the scientific group for our permutation test.*

```
Misinformation 2011 vs Misinformation 2012 permutation test: the p-value is 9.999000099990002e-05.
Misinformation 2012 vs Misinformation 2013 permutation test: the p-value is 0.0022997700229977.
Misinformation 2013 vs Misinformation 2014 permutation test: the p-value is 0.0086991300869913.
Misinformation 2014 vs Misinformation 2015 permutation test: the p-value is 0.20467953204679531.
Misinformation 2015 vs Misinformation 2016 permutation test: the p-value is 9.999000099990002e-05.
Misinformation 2016 vs Misinformation 2017 permutation test: the p-value is 9.999000099990002e-05.
Misinformation 2017 vs Misinformation 2018 permutation test: the p-value is 0.8886111388861114.
Misinformation 2018 vs Misinformation 2019 permutation test: the p-value is 9.999000099990002e-05.
Misinformation 2019 vs Misinformation 2020 permutation test: the p-value is 9.999000099990002e-05.
```

**Figure 12:** *This graph shows our results for the misinformation group for our permutation test.*

### 4.3.1 Permutation Test: Data Visualizations

For our permutation test data, we wanted to visualize to see the growth over time as well as total likes for each group. This allows us to see the changes for each group from 2012 to 2020.

In graph 13, we compare our two groups by analyzing the total number of likes per year. The graphs indicate that our scientific group has significantly more total likes per year than our misinformation group. This has been a common trend we see throughout our research.

Another important spike we see in our graphs is in 2019. We can infer that there is a significant spike during this year because this is when the COVID-19 pandemic started to be more spoken about online. This pandemic was a major viral disease which spread around the world, causing many to become extremely ill. Our graph shows that 2020 is much lower in total likes than 2019, however it is important to note that our 2020 data does not contain the last 7 months of 2020 due to the time of data collection.

For our final permutation graph in Figure 14, we compare our two groups to see the growth ratio of likes per year. This graph shows that in 2016, there was a high growth ratio for both of our groups. During this year, the presidential election for Donald J. Trump and Hillary Clinton took place in the United States. There was much controversy during this time regarding Clinton's email scandal and collusion regarding Russia and the election in favor of Trump.

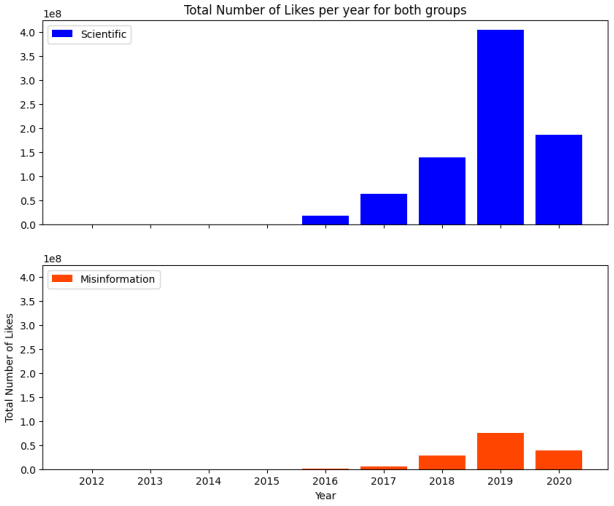


**Figure 13:** *This graph compares the total number of likes per year for our scientific and misinformation groups.*

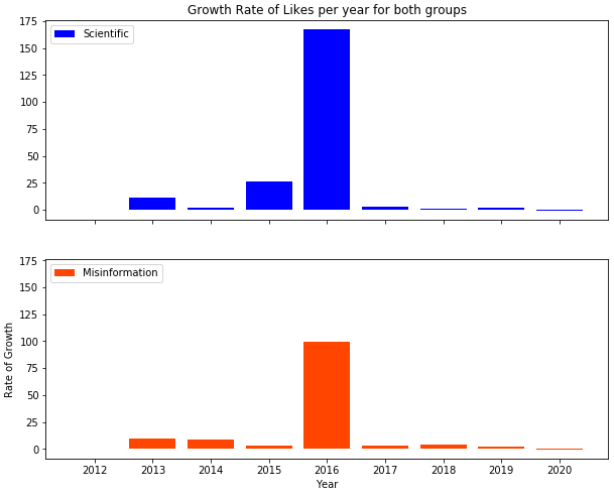


**Figure 14:** *This graph shows the comparison of growth ratios of likes per yer for our scientific and misinformation groups.*

## 5 Results

As a result of our investigation, we found that politicians who spread misinformation often have a higher ratio value and less overall likes per tweet. This higher ratio value means that these politicians are more likely to spread controversial information on Twitter. This also shows that people who are viewing their tweets on Twitter are engaging in the politicians' tweets by commenting more compared to liking or retweeting.

In contrast, we see that politicians who spread scientific information on Twitter have lower ratios and significantly more likes on their tweets. This is interesting to note because it shows a clear

distinction and result between our two groups.

When comparing the two groups, we see that our scientific group has been steadily increasing in growth over the years while our misinformation group has only been growing significantly in the past recent years.

The overall result of our research shows that a politician has the most growth through spreading non-controversial, scientific information because this yields a steady growth over time in comparison to spreading controversial information.

## 6 Conclusion

Twitter is one of the largest social media platforms and as more politicians move to Twitter as a means of sharing their political thoughts and opinions, we see that their popularity and reputation are strongly amplified. The digital world can massively transform the growth of a politician depending on the types of tweets they share.

Our ratio and popularity metrics show us that a politician's controversial tweets can heavily impact their audience engagement. Scientific, non-controversial tweets mainly spread by likes while misinformation or controversial tweets spread by having more retweets or comments addressing the tweet.

The permutation test shows us that the growth for politicians who share scientific information has been more steady since they started tweeting, whereas politicians sharing misinformation has only recently started to see a rise in growth.

These distinct patterns show how a politician can grow over time and the amount of influence they have on their Twitter followers and audience online.

The next envisioned steps of our analysis include collecting a larger sample size of politicians in order to compare each politician to a larger sample size. In addition to this, we would include some former and current presidents such as Donald Trump and Joe Biden. We may also expand on more social media platforms because this would allow us to expand our data and see what other types of content politicians are posting. Additional platforms could include Facebook and Reddit.

## References

[**1**] Justin Littman. (2018). TweetSets. Zenodo. https://doi.org/10.5281/zenodo.1289426

[**2**] Wrubel, Laura; Kerchner, Daniel, 2020, "116th U.S. Congress Tweet Ids", https://doi.org/10.7910/DVN/MBOJNS, Harvard Dataverse, V1.

[**3**] Seddiq, O., Relman, E. (2020, September 02). Republican Sen. Joni Ernst promoted a far-right misinformation theory that falsely claims coronavirus cases are inflated by healthcare providers. Retrieved January 25, 2021, from https://www.businessinsider.com/gop-senator-pushes-qanon-misinformation-theory-on-coronavirus-case-count-2020-9

[**4**] Zadrozny, B., Collins, B. (2021, January 07). Trump loyalists push evidence-free claims that antifa activists fueled mob. Retrieved January 25, 2021, from https://www.nbcnews.com/tech/internet/trump-loyalists-push-evidence-free-claims-antifa-activists-fueled-mob-n1253176

[**5**] Words we're WATCHING: What is 'The Ratio' AND 'RATIOED'. (n.d.). Retrieved January 25, 2021, from https://www.merriam-webster.com/words-at-play/words-were-watching-ratio-ratioed-ratioing