I Think What I Hear, The Influence of Social Media on Political Ideologies

Zakery A. Fyke

Texas Tech University  
2500 BroadwayLubbock, Texas

zakery.fyke@ttu.edu

Richard Matovu

Texas Tech University  
2500 BroadwayLubbock, Texas

richard.matovu@ttu.edu

Fang Jin

Texas Tech University  
2500 BroadwayLubbock, Texas

fang.jin@tttu.edu

**ABSTRACT**

Social Echo Chambers create boundaries between people, radicalizing ideologies and making collaboration and compromise more difficult. In this paper, we analyze the data available on Twitter to analyze and illustrate the party divide reinforced by these Echo Chambers.

**CCS Concepts**

• **Mathematics of computing~Exploratory data analysis**   • *Networks~Social media networks*   • Human-centered computing~Social media

**Keywords**

Social Media; Data Analysis; Echo Chambers; Politics

# INTRODUCTION

The current political atmosphere across the world has changed. The results of the 2016 Presidential Election and the 2016 Brexit vote in the United Kingdom surprised many, defying prediction models and expectations of voters across both countries. Politics in the United States has become heavily divided between the two main parties, the Republicans and the Democrats, and some experts believe that the shock from the election results can be attributed to the formation of Echo Chambers in the public sphere, often blaming the widespread use of social media and other electronic communication methods for their existence [1][2][3].

Echo Chambers are the metaphorical description given to the situation in which beliefs, information, and ideas are reinforced or cemented inside a defined system. A common example is an individual exclusively associating themselves with those who agree with or share their ideals and their point of view. The Echo Chamber reinforces one’s own present world view because of the lack of exchange of dialogue with those who hold a different point of view. The advent of Social Media has allowed Echo Chambers to be formed more easily than ever, and with over 60% of Americans getting their news from Social Media according to a study by Pew Research Center [4] this problem could be getting worse. Analysis of the data found on Twitter may allow us to locate these Echo Chambers present based on the political party divide.

Social Echo Chambers have become a hot topic, especially in public media, as many believe that the formation of Echo Chambers is leading to a radicalization of political views on both sides of the party divide, making it increasingly difficult for both sides to find a common ground, and leading to social unrest. This inability to find common ground between the two parties makes it increasingly difficult for the two to find compromises that both sides will accept, leading to events like the 2013 Shutdown of the Federal Government when both sides were unable to find a compromise on the Federal Debt Ceiling.

This tendency of similar and likeminded individuals to form ties with one another in a form of Confirmation Bias is also known as homophily. The phenomenon which causes this is often cited as cognitive dissonance and selective exposure theories, well researched and explained areas of human psychology. According to these theories, people experience positive feeling when they are presented with information which confirms their already held opinions on the subject. When faced with opinions which conflict with their own, humans are more likely to experience stress and a pressure to conform. This leads to individuals being more likely to seek out others who agree with them and to find information and discussions which reinforce their original and already held view, causing individuals to join together into smaller homogeneous groups out of the overall public sphere, affiliating with other individuals with similar beliefs, educations, and world views. It follows, then, that the Echo Chamber Effect is caused by this tendency of individuals desiring to create homogeneous groups, intentionally or otherwise, by exclusively affiliating with individuals whom share their own political view.

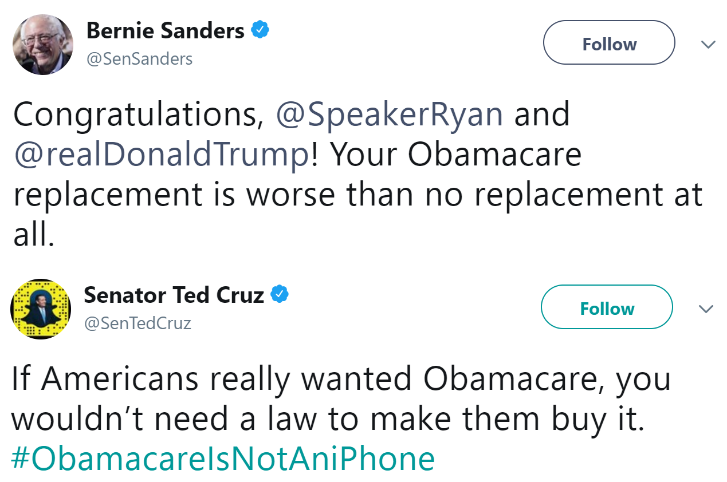


Figure 1. US Senators discussing the Affordable Care Act, also known as Obamacare, to their followers on Twitter.

We can find and visualize the formation of these Echo Chambers by utilizing the information made available on social media, namely by analyzing tweets and followers found on Twitter. Twitter is one of the largest social media platforms in the world, and with the politically charged atmosphere currently present in the United States, a large amount of data has been made easily available, allowing us to visualize the formation of these Echo Chambers on Twitter by analyzing the tweets and followers by leading members of both parties.

We collected the tweets and followers from leading members of both major parties in the United States, then analyzed them to gather our conclusions. Our contributions are as follows:

* We illustrate the divide between the two parties by analyzing the shared followers of our representatives.
* We implicitly discover the current state of affairs in the US Political Sphere by analyzing the tweets of representatives from both parties.
* We compare how both parties discuss popular issues. As illustrated in Figure 1, we see two top US Senators, Bernie Sanders and Ted Cruz, discussing the Affordable Care Act in very different ways.

# RELATED WORK

The effects of conversations happening in cyberspace have a very real and measurable impact in the real world. Hampton et al [5] found in 2016 that “Twitter users who felt their audience on Twitter agreed with their opinion were more willing to speak out on that issue in the workplace”. There have been numerous methodologies used to locate and visualize the Echo Chambers that become present online and on social media, including analyzing blog posts and the comments on those posts and analyzing the messages and connections found on Twitter and Facebook.

The topic of whether ideological polarization is exhibited in online exchanges is still an open debate among researchers. Conover et al [6] found that political ideology could be predicted with a high level of accuracy by analyzing tweets of the users. By contrast, Bakshy et al [7] found that there was very little online ideological segregation in absolute terms, with open exchanges and exposure to ideological differences being common for nonpolitical issues. Variations like these are common between studies, and one reason for them might be that some studies use a self-selected sample of partisan individuals, while other studies have not.

Similarly, which side of the political spectrum would be more likely to engage in selective exposure to information -- information which confirms the opinions which they already hold -- is still an open topic with a variety of answers from different researchers. While not the direct focus of our studies, it should be noted in the peripheral that different studies have come to different conclusions on the subject. The ones performed by Bakshy et al [7] found that liberals are much more likely than conservatives to engage in cross-ideological dissemination of political and nonpolitical information, while Colleoni et al [8] found that Democrats exhibit higher levels of political homophily, indicative of an Echo Chamber, but that Republicans that follow official Republican accounts exhibit higher levels still.

## Blog Analysis

In the past, previous work focused the presence of Echo Chambers forming around online blogs. Gilbert et al [9] found in their study that blogs frequently created what they defined to be an Echo Chamber, with certain genres of blogs being more likely to meet these conditions. However, their metric for deciding if a blog could be categorized as an Echo Chamber was based on the ratio of comments left on posts which agreed with the writer of the blog. As they mention, there is no agreed upon metric which designates something as an Echo Chamber, and we instead focus on the communication between separate groups, rather than between a single individual, the author of the document, and a group of individuals.

Adamic and Glance [10] analyzed the differences in behavior of liberal and conservative blogs and how frequently one referred to another. They chose 1000 political blogs, and analyzed the top 20 conservative and top 20 liberal blogs, ranking them based on the number of citations each blog received from October and November of 2004, the period of the George W. Bush vs John Kerry Presidential election. From these 40 blogs, they collected roughly 23,000 posts from 8/29/04 - 11/15/04, with a slight bias in the count towards the left leaning blogs. During analysis, they found that conservative blogs link to one another more frequently than liberal blogs, but that particular liberal blogs are more likely to be linked to by other left leaning blogs. Their analysis also showed that only 15 percent of all links were bipartisan. This analysis of how frequently blogs link to one another is an interesting metric, but because the context of why the links were shared was not analyzed, it’s difficult to say whether it’s a good metric. In this study, we instead focused on followers, as, by following our representatives, the followers have shown a marked interest in seeing what that person has to say.

## Twitter Analysis

More recent work has begun to focus on the influence of Social Media on the public sphere and on the formation of Echo Chambers. Twitter has been the focus of many of these studies, due to the ease of access for information on users and their corresponding tweets. By default, all tweets and followers are freely available to anyone. The content of tweets often proves to be easier to analyze as well; because they are limited to 140 characters, tweets must be short and to the point, allowing for easier topic modeling.

Barberá et al [11] analyzed a collection of 3.8 million Twitter users in the United States. They pointed out that methods which focus on analyzing the content of Twitter messages to draw conclusions was not scalable, bringing up questions about the validity of any conclusions made by these techniques. Instead, they focused on Latent Space Modeling, focusing on the connections between different Twitter users and their retweeting, or sharing of tweets, to draw their conclusions. They also chose 12 significant political and nonpolitical events from the period of 2012 to 2014, including the United States Federal Budget, Marriage Equality, the Winter Olympics, and the Academy Awards, and used these to collect roughly 150 million tweets which mentioned any of these keywords.

Using this collection of tweets, they found their 3.8 million active Twitter followers to analyze. By using known political pages, such as The Tea Party, Stephen Colbert, and others, they created a Latent Space Model of all these Twitter user. Barberá et al [11] concluded that while ideological contours do not necessarily bound the Social Media sphere, especially when it comes to nonpolitical issues like the Winter Olympics, when it comes to politicized issues, individuals are significantly more likely to pass on information to other users that they have received from sources which are like their existing network.

Williams et al [12] utilized a similar form of Network Analysis, collecting nearly 600,000 distinct tweets from nearly 180,000 distinct users regarding Climate Change by utilizing hashtags such as *"#globalwarming", "#climatechange",* and *"#climaterealists"* to name a few to perform an analysis of social media debates regarding Climate Change. They created three separate social networks: "Follower" networks, "Retweet" networks, and "Mention" networks, where each was created from users following one another, retweeting each other's tweets, or mentioning one another in a tweet respectively. The most active users analyzed were classified by researchers as either activists, skeptics, neutral, or unknown. If researchers could not reach a unanimous decision, the user was marked as ambiguous.

Williams et al found in their study that there was a high degree of polarization in attitudes regarding climate change discussion on Twitter; users who were active in online discussions tended to have strong attitudes and users on Twitter were likely to self-segregate into likeminded communities regarding the issue. This self-segregation may be indicative of divides along overall political ideologies. However, because this study only focused on the subject of Climate Change, they could only draw conclusions on that single topic. In this study, we choose to analyze the entire state of affairs discussed by our ten representatives, instead of focusing on a singular topic.

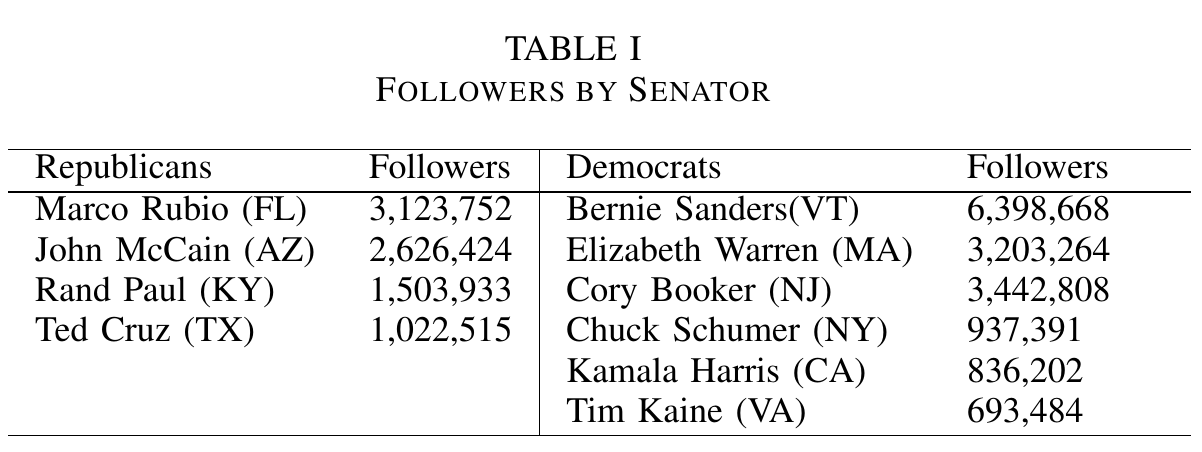
Colleoni et al [8] utilized data collected by Kwak et al [13] in 2009 which consists of all nodes and ties on Twitter in 2009, where nodes are individual users and ties represent the relationships between them. This dataset consists of over 40 million nodes and 1.47 billion ties, though, because it is almost a decade out of date, this number has certainly grown considerably since then. They also utilized a collection of 467 million tweets from a 7-month period from June 1st to December 31st, 2009 collected by Yang and Leskovec [14], utilizing training data containing 60,000 political and 170,000 nonpolitical titles from news blogs. A training set containing 1,683 Democrat users and 8,868 Republican users was used, the researchers assuming that users who are following exclusively Republicans or Democrats are themselves Republicans or Democrats.

Colleoni et al used this collection of data to predict the political orientation of users from the content they shared, with an accuracy of roughly 0.96 for Republicans and 0.79 for Democrats, identifying 72,302 Republicans, and 782,371 Democrats. Using this, they measured the political homophily by defining it as the number of outbound ties (that is, directed to users with similar political orientation plus directed to users with different political orientation). While they concluded that Democrats in general had higher levels of homophily than Republicans, except when considering Republicans which follow official Republican sources, they also had a heavily unbalanced dataset overall, with far more Democrats than Republicans identified, despite having far more Republicans than Democrats in their training data. This heavy imbalance in their dataset leads to questions about the accuracy of their conclusions; for our study, while we do have more Democrats than Republicans, we attempted to keep the numbers more closely tied.

# EXPERIMENT SETUP

## Dataset

Data used for this study was collected from the ten United States Senators which have the largest following on Twitter, measured by the number of followers according to [15], as their party affiliation is already established, as are their Twitter handles. From these accounts, we collected a subset of 500,000 followers per Senator, for a total of 3,338,398 unique users shared across all accounts, with Republican Senators having 1,678,040 unique followers, and Democrat Senators having 2,167,455 unique followers. Individual follower accounts for each senator are given in Table 1.



For each of our ten senators, we harvested up to 3,200 tweets via use of the Twitter API for Python, starting with their most recent tweets, for a total of 12,662 tweets from our four Republican Senators and 18,623 tweets from our Democrat Senators, for a total of 31,225 tweets in all.

Republican tweets contained a total of 120,896 words and a vocabulary of 17,100 unique words, while Democrat tweets contained 190,746 words and a vocabulary of 20,512 unique words.

## Latent Dirichlet Allocation Analysis

Latent Dirichlet Allocation (LDA) is a statistical model based on mathematical distributions which was laid out in the paper by Blei et al [16]. It uses a hierarchical Bayesian approach in which Dirichlet priors are placed on the underlying multinomial distributions [17]. A graphical model representation of LDA is given in Figure 2.

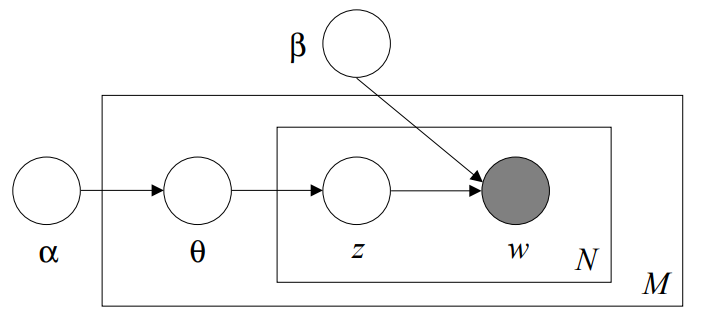


Figure 2. Graphical model representation of Latent Dirichlet Allocation in Plate Notation, where the boxes are “plates” representing replicates [16]. Here, the inner plate represents the repeated choice of topics and words within a document, and the outer plate.

In this paper, we utilize LDA on the tweets we have harvested as a form of topic modeling. To achieve this, we combined the tweets of all our Republican senators into a single dataset, and did the same for our Democrat senators. Each dataset was then placed into a dictionary which was then transformed into a Document Term Matrix (DTM). The vocabulary of this DTM was extracted, then the LDA model was fitted with the DTM with a set number of topics and a set total of 2000 passes over the set. After, we retrieved the topics from the LDA model.

We assumed that our Republican and Democrat senators would have a different number of topics they discuss on Twitter. Therefore, we needed to find the ideal number of topics for each. To do this, we ran our LDA model with an increasing number of topics from 1 to 100. We then plotted this number against the Log Likelihood corresponding with the topic count. The results are shown in Figure 3, which shows that the ideal number of topics for Democrats was 7, and the ideal for Republicans was 13.

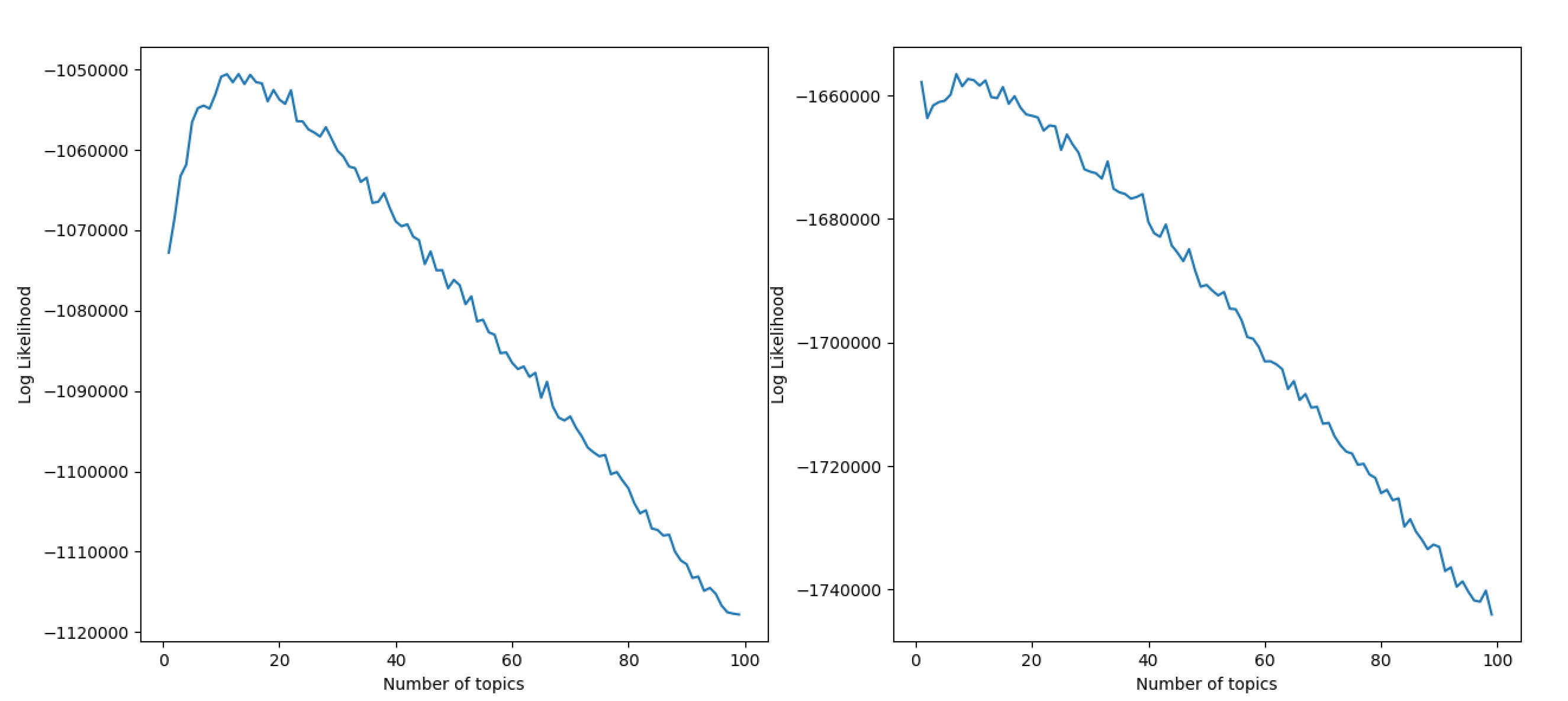


Figure 3. Log Likelihood for increasing topic numbers. Republicans are shown on the left, Democrats on the right.

# EXPERIMENTATION RESULTS

## Follower Analysis

First, we analyzed the list of Twitter followers of the ten most influential US Senators on Twitter, measured by their number of followers. We then compared these sets of followers to find if followers of one political party are less likely to follow the members of the other. This was indeed the case and, in addition, we found that, in general, Democrats were less likely than Republicans to follow senators of the other party, but that Republicans who followed all top members of their party were less likely than Democrats to follow senators from the other party.

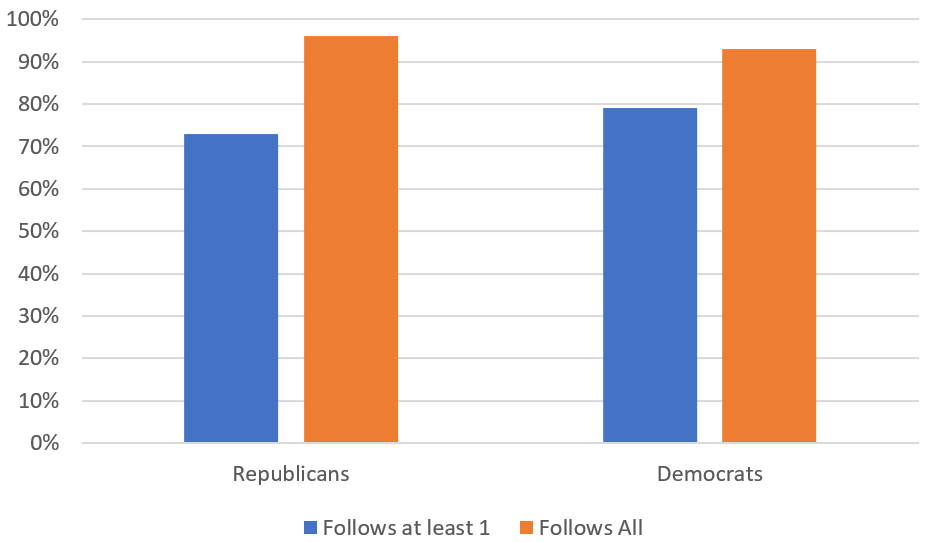


Figure 4. Distribution of followers who exclusively follow one party.

Out of 1,678,040 followers who followed at least one of our four Republicans, we found that 1,220,940, almost 73 percent, do not follow any of our six Democrats. Similarly, we found that out of our 2,167,455 followers who followed at least one of our Democrats, 1,710,358, almost 79 percent, do not follow any of our four Republicans. We then narrowed this further, by examining those who are following all the top members of a party. Out of 1,678,040 followers who followed at least one of our Republicans, we found that 11,393 followed all four of our Republican Senators. Of this, we found that 10,928, almost 96 percent, did not follow any of our Democrat Senators, meaning they exclusively followed the four Republicans. Similarly, of our 2,167,455 followers who followed at least one of our Democrats, we found that 6,498 followed all six of our Democrat Senators, and that 6,033, almost 93 percent, of these did not follow any of our Republican senators as shown in Figure 4. The percentage of Twitter users who follow at least one Republican Senator and no Democrat Senators is lower than the percentage of users who follow at least a single Democrat Senator and no Republican Senators. This potentially indicates that followers of Republican Senators are slightly more likely to also follow a Democrat Senator than followers of Democrat Senators are to follow a Republican Senator. However, users who follow all four of our Republican Senators are also shown to be slightly less likely than users who follow all top Democrat Senators to follow members of the other party.

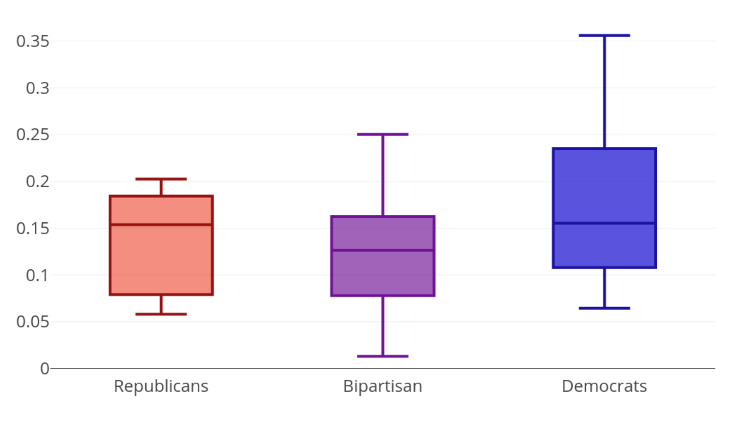
**Figure 5. Percentage of shared followers within and between parties.**

Figure 5 shows the percentage of shared followers between the ten senators, both within the same party and between the two parties.

We find that within the Democratic Party, Senators Bernie Sanders and Elizabeth Warren share the highest proportion of their followers, with over 35 percent shared between them, while Sanders and Tim Kaine share the lowest within their party, at only 6 percent. For the Republicans, Senators Marco Rubio and Rand Paul share 20 percent of their followers, while Senators Ted Cruz and John McCain share the least, at only 5.8 percent, similar to Bernie Sanders and Tim Kaine within their party. Overall, the Republicans have a median of 15.3 percent, a number which is comparable to the Democrats’ median of 15.5 percent, seeming to suggest that the followers of the two parties are equally likely to follow other members of the same party.

However, when analyzing the shared followers on a Bipartisan basis, we see a lower proportion shared between each senator. Senators Cory Booker and Marco Rubio share the highest percentage of followers at 25 percent, while Bernie Sanders and Ted Cruz share the fewest at only 1.3 percent, with a median value of 12.6 percent for all bipartisan ties. This difference between the interparty follower sharing seems to indicate that Twitter users are more likely to follow other members of a party which they already follow, rather than to follow members of both parties, illustrating the party divide which we expected.

## Current Affair Analysis

Once this was completed, we then analyzed the Tweets sent out by these ten United States Senators to find the topics which each discussed most frequently, finding that, as expected, both parties discussed different topics on Twitter.

After finding the ideal number of topics for our LDA analysis as discussed previously, we modeled the topics for both of our party senators. Finding the frequency for the phrases which were contained in each of these topics, we were able to create a Word Cloud visualization of those phrases which appeared more frequently. This Word Cloud is shown in Figure 6. Size of words is indicative of frequency, with the most frequent terms being the largest. Words which are from the same topic identified by our LDA model also share the same color. The probabilities for this diagram were normalized across all topics, so that the phrases with the highest probabilities within their own topic would appear with similar size to those in other topics. Lemmatization was also performed, so that words such as “discuss”, “discussed”, and “discussing” would be counted as a single term. From this visualization, we see a wide range of phrases used in tweets by both parties, with the Republicans frequently mentioning phrases like “Obamacare”, “Repeal”, and “Hearing”, along with many others. Meanwhile, Democrats more frequently use terms such as “Bill”, “Health”, and “Student”, among others. We do see some obvious common phrases as well, such as, “Tax” and “American” mentioned frequently by both parties.

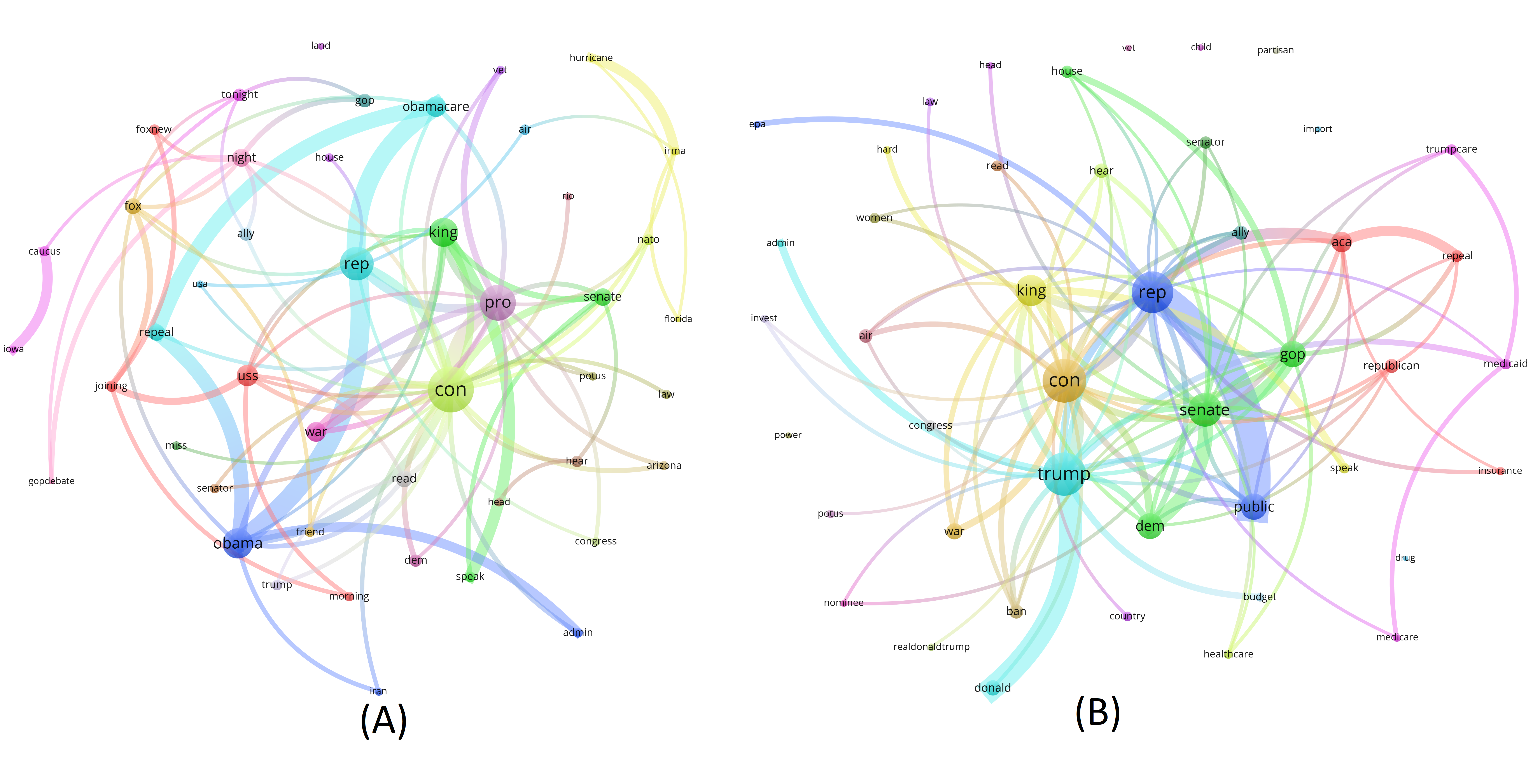


Figure 7. Map of links between words in tweets. (A) shows the links for Republicans, (B) shows the links for Democrats.

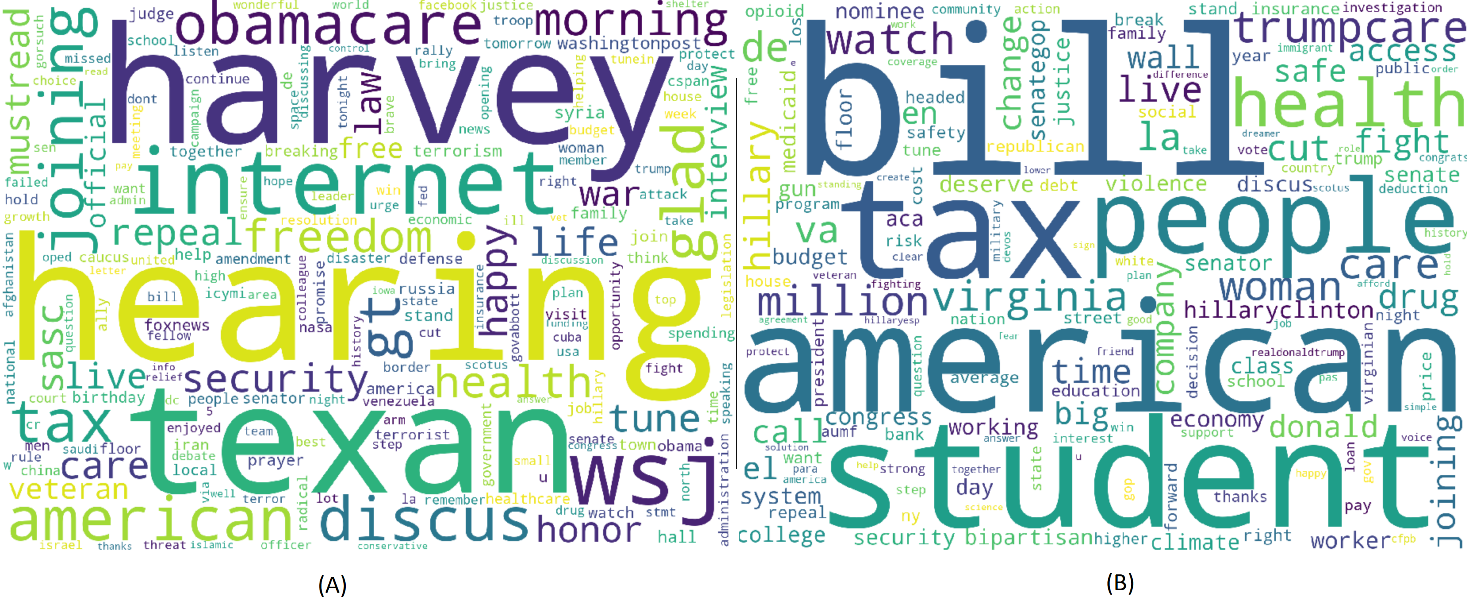


Figure 6. Word cloud of Senator Tweet Topics. (A) shows the most discussed words in the topics by the Republican Senators, (B) shows the most discussed words in the topics by the Democrat Senators.

For the Republicans, we also see “Harvey” appearing prominently, along with “Texan”, which, with two of our Republican Senators from states impacted by Hurricane Harvey in 2017, is expected.

In Figure 7, the map of connections between words used by both parties is shown. Larger nodes represent words more frequently used, and thicker edges denote a stronger connection between the two words. As we see, Republicans frequently use the terms “Obama”, “Admin”, “Obamacare”, and “Repeal” together, whereas the Democrats refer to the healthcare act by its more formal name of the ACA, or Affordable Care Act, and when discussing it with the phrase “Repeal”, it’s frequently along with “Republican”. Also, when discussing “Admin”, the Democrats are more likely to link to the current president, Donald Trump, than the former President Obama. This shows that, while both sides may be discussing some of the same terms, they are discussing them in very differently from one another, as we would expect to see.

# Conclusion

It seems that with each passing year, American politics and politics around the globe are becoming increasingly divided along party lines. Each year it’s becoming harder for the two sides to find compromises that will make both parties happy, leading to more and more political strife as time goes by. The formation of Echo Chambers, caused by self-segregation of individuals into smaller homogenous groups, have largely contributed to this by making it more difficult to consider the validity of the other side’s point of view.

In this study, we utilized the information available on Twitter to analyze the tweets and collections of followers of the ten most followed United States Senators. By comparing the groups of followers for each senator, we showed that the followers of the two parties are divided not only by ideology, but also by communication on Social Media. Our findings seemed to indicate that Twitter users who follow at least one of our Democrats are less likely to follow any Republicans and vice versa. We also found that the number of shared followers between the Senators is similar within their own parties, but that the bipartisan follower sharing is markedly lower, seeming to suggest a barrier in discussion between the two sides.

We also showed that the leaders of the Republicans and Democrats are discussing a range of topics which often do not coincide with one another, and when they do, they are often discussed in widely different ways.

# References

[1] NPR. Tech creates our political echo chambers.

it might also be a solution. [Online]. Available:

https://www.npr.org/sections/alltechconsidered/2017/04/12/522760479/tech-creates-our-political-echo-chambers-it-might-also-be-a-solution

[2] CNN. Most republicans and democrats have few or no friends in the opposing party. [Online]. Available: http://www.cnn.com/2017/10/05/politics/friends-political-party/index.html

[3] N. Y. Times. The dangers of echo chambers on campus.

[Online]. Available: https://www.nytimes.com/2016/12/10/opinion/sunday/the-dangers-of-echo-chambers-on-campus.html

[4] J. Gottfried and E. Shearer, “News use across social media platforms 2016,” 2016.

[5] K. Hampton, I. Shin, and W. Lu, “Social media and political discussion: when online presence silences offline conversation,” pp. 1–18, 08 2016.

[6] M. D. Conover, B. Gonçalves, J. Ratkiewicz, A. Flammini, and F. Menczer, “Predicting the political alignment of twitter users,” in Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on. IEEE, 2011, pp. 192–199.

[7] E. Bakshy, S. Messing, and L. A. Adamic, “Exposure to ideologically diverse news and opinion on facebook,” Science, vol. 348, no. 6239, pp. 1130–1132, 2015. [Online]. Available: http://science.sciencemag.org/content/348/6239/1130

[8] E. Colleoni, A. Rozza, and A. Arvidsson, “Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data,” Journal of Communication, vol. 64, no. 2, pp. 317–332, 2014.

[9] E. Gilbert, T. Bergstrom, and K. Karahalios, “Blogs are echo chambers: Blogs are echo chambers,” in System Sciences, 2009. HICSS’09. 42nd Hawaii International Conference on. IEEE, 2009, pp. 1–10.

[10] L. A. Adamic and N. Glance, “The political blogosphere and the 2004 us election: divided they blog,” in Proceedings of the 3rd international workshop on Link discovery. ACM, 2005, pp. 36–43.

[11] P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau, “Tweeting from left to right: Is online political communication more than an echo chamber?” Psychological science, vol. 26, no. 10, pp. 1531–1542, 2015.

[12] H. Williams, J. R. McMurray, T. Kurz, and F. Hugo Lambert, “Network analysis reveals open forums and echo chambers in social media discussions of climate change,” vol. 32, 05 2015.

[13] H. Kwak, C. Lee, H. Park, and S. Moon, “What is twitter, a social network or a news media?” in Proceedings of the 19th international conference on World wide web. ACM, 2010, pp. 591–600.

[14] J. Yang and J. Leskovec, “Patterns of temporal variation in online media,” in Proceedings of the fourth ACM international conference on Web search and data mining. ACM, 2011, pp. 177–186.

[15] B. Analytics. United states senators on twitter. [Online]. Available: https://www.birdsonganalytics.com/politics/senators/

[16] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” Journal of machine Learning research, vol. 3, no. Jan, pp. 993–1022, 2003.

[17] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

Columns on Last Page Should Be Made As Close As Possible to Equal Length

Authors’ background

|  |  |  |  |
| --- | --- | --- | --- |
| Your Name | Title\* | Research Field | Personal website |
| Zakery Fyke | Master Student |  |  |
| Richard Matovu | Phd Candidate |  |  |
| Dr. Fang Jin | Assistant Professor |  | http://myweb.ttu.edu/fjin/ |

\*This form helps us to understand your paper better, the form itself will not be published.

\*Title can be chosen from: master student, Phd candidate, assistant professor, lecture, senior lecture, associate professor, full professor