Final Project

Jessica Nadalin

December 11, 2019

Abstract

I analyze Twitter data from United States senators and representatives in the context of presidential impeachment, extracting Tweet content, Twitter interactions, and Twitter following relationships using the Twitter API. I develop a methodology to identify influential users in the discussion of impeachment, and identify sixteen influential users. I analyze the relationships between these users, along with their Twitter activity over time and in relation to key impeachment events. I introduce a measure of polarized discussion and apply it to this topic, finding relatively high polarization along party lines. The polarized nature of discussion appears to be related to critical impeachment events.

1 Introduction

1.1 Background

This project focuses on Twitter data from United States legislators as they discuss the impeachment of President Donald Trump. The impeachment inquiry began with a whistleblower complaint on September 9th, 2019 concerning Donald Trump's dealings with Ukraine, and developed into a formal impeachment proceeding, culminating in a inquiry report released by the House of Representatives on December 3rd, 2019. A more thorough timeline can be found below, in Section 1.3.

A vast literature exists on the study of Twitter network data, and I base part of my approach on several existing strategies. The structure of networks **F** and **I**, as discussed in Methods, draws from examples in [1],[2]. Strategies for detecting influential users using eigenvector centrality draw from [3],[4], and polarized community detection draws from work in [1],[5].

1.2 Twitter Data

Using the Twitter API, I collected the 1000 most recent tweets from all U.S. senators and representatives on Twitter (532 legislators in all), and related interaction data. A tweet from a given user consists of a string of at most 280 characters, and I consider four different types of interactions for this project: mentions, replies, quotes, and retweets. A mention is a Twitter user including the username of another user in their tweet; a reply is a tweet made by one user in response to a particular tweet from another user; a quote is a user posting the tweet of another user, along with any of their own commentary; a retweet is a repost of another user's tweet, without commentary. I also used the Twitter API to collect the full list of users each legislator follows on Twitter. The data was last updated on December 8th, 2019.

1.3 2019 Impeachment Timeline

In Results, we will be considering activity of legislators over time, particularly in relation to events in the impeachment process. For reference, I've included a table of (some, but not all) relevant events pertaining to the impeachment inquiry below.

Date	Event
09-09-19	House announces investigation into Giuliani's Ukraine efforts and halting of aid Congress notified of whistleblower complaint
09-10-19	Adam Schiff demands whistleblower complaint, Bolton resigns
09-11-19	Ukraine aid released
09-13-19	Schiff subpoenas for whistleblower complaint
09-18-19	The Post reports the complaint involves Trump's communications with foreign leader and some kind of promise
09-24-19	Trump confirms he withheld funding for Ukraine's military aid Nancy Pelosi announces support of formal impeachment inquiry
09-25-19	White House releases transcript of Trump-Zelensky call
09-26-19	White House declassifies whistleblower complaint, which Schiff releases
10-01-19	Mike Pompeo sends House Democrats letter stating State Department employees summoned for depositions will not appear, critiques impeachment inquiry
10-03-19	Trump states he had hoped Zelensky would start an investigation into the Bidens and suggests "China should start an investigation into the Bidens"
10-06-19	News of a second whistleblower surfaces
10-12-19	Post reports Sondland is to testify "no quid pro quo" came from Trump
10-17-19	Mulvaney momentarily confirms quid pro quo in press conference
10-31-19	House votes to formalize impeachment inquiry and open hearings
11-04-19	House releases first closed-door deposition transcripts
11-13-19	First open hearings in impeachment inquiry (Taylor and Kent)
11-15-19	Yovanovitch publicly testifies
11-19-19	Vindman, Williams, Volker, and Morrison testify
11-20-19	Sondland testifies
11-21-19	Hill and Holmes testify, completing public testimony of inquiry
12-03-19	House releases inquiry report documenting impeachment case against Trump

2 Methods

In this project, I address two separate questions: in relation to impeachment, which legislators are the most influential, and how politically polarized is this discussion?

2.1 Finding users with greatest influence

I seek to understand which legislators have the greatest influence on other legislators, in the context of Twitter interactions about impeachment. Which legislators appear to promote discussion of impeachment among their followers? Which followers are spoken about the most, or speak with others the most, in the context of impeachment? How does the behavior of these most influential users change over time, especially in reference to impeachment events? What does the network of interaction and network of follows look like for these top users?

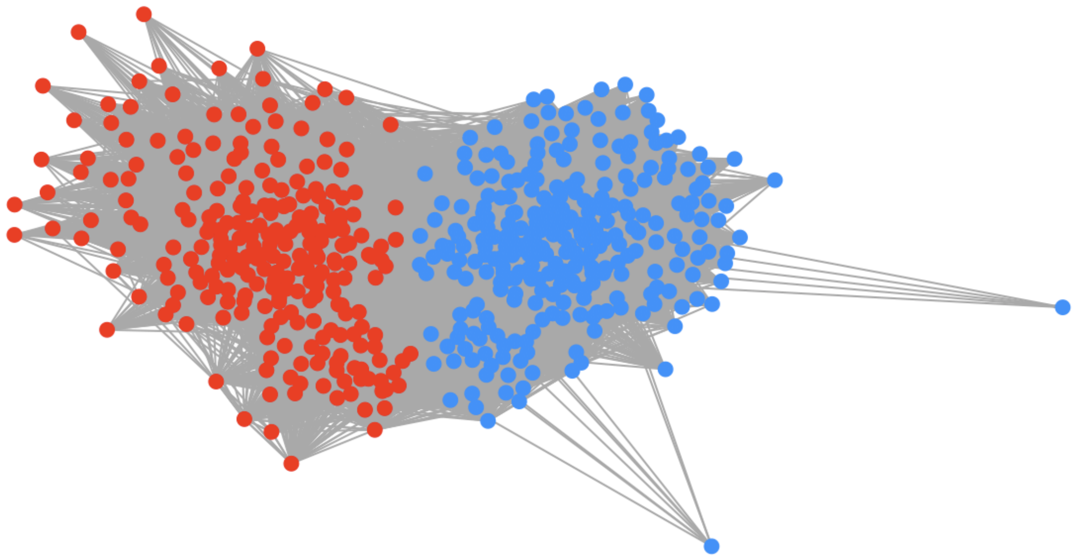


Figure 1: The full follow network **F**, considered over all time.

To address this problem, I consider two networks: the first is a directed network of Twitter follows among legislators (referred to as network **F**), i.e. nodes are legislators and an edge from node A to node B indicates legislator A follows legislator B (Figure 1). I seek to understand for which users it is true that when they tweet about impeachment, their followers tweet about impeachment within the week. To do this, I consider the following induced subgraph of **F**: on a given day, nodes are restricted to legislators who have tweeted about impeachment in the past week. Edges are restricted to follows from legislators who have tweeted about impeachment on this given day. In this way, a node with high in degree is a legislator who has tweeted about impeachment in the past week, and who has a high number of followers that have since tweeted about impeachment. I consider high in degree for these sub-networks to correspond to influence, as there's a relationship between a legislator tweeting about impeachment and their followers subsequently tweeting about impeachment. An example network can be seen in Figure 2, where we've restricted nodes of **F** to users who've tweeted about impeachment in the week 9-8-19 through 9-14-19, and we've restricted edges to follows from users who tweeted about impeachment on 9-14-19. We can see one Republican and two Democratic users with relatively high in-degree, and hence high influence among their followers.

The second network I consider is the directed interaction network **I** between legislators, where again a node is a legislator, but edges from A to B are now an interaction (i.e. a mention, reply, quote, or retweet) from legislator A to legislator B (Figure 3). If multiple users are mentioned in an interaction (e.g. A mentions both B and C), we consider these separate edges (an edge from A to B and another from A to C). We consider an induced subgraph of **I** for a given day, where edges are restricted to interaction tweets about impeachment from the past week, and nodes are restricted to those with degree > 0. By breaking the full timeline into week-long intervals, we can see how interactions, and who the most influential users are, can change over time. To gauge influence, I consider eigenvector centrality of the induced subgraph. Those with high eigenvector centrality will be those who have interacted the most with others on the topic of impeachment, either on the giving or receiving end, and those who interact with users who have high interaction. An example subgraph can be seen in Figure 4, where we consider the week from 10-29-19 to 11-04-19. The

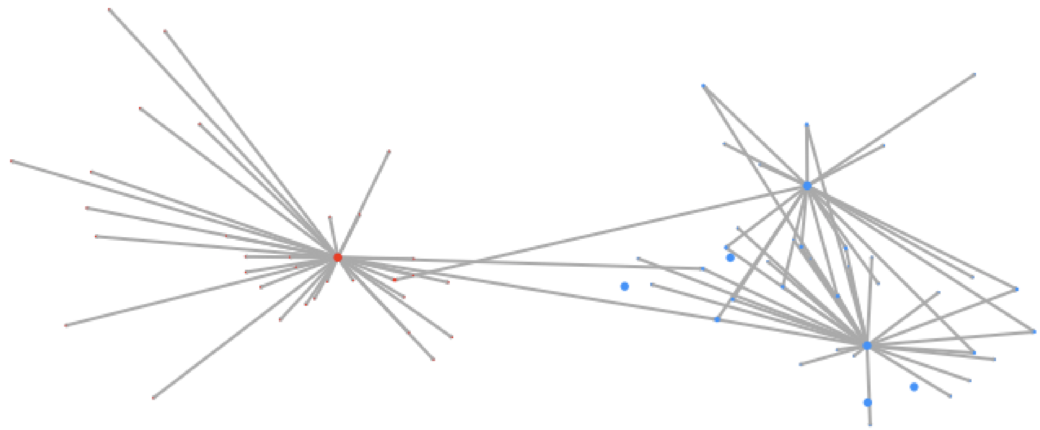


Figure 2: The network **F** with nodes restriced to users tweeting about impeachment between 9-8-19 and 9-14-19, and edges restricted to follows from users tweeting about impeachment on 9-14-19. Node color corresponds to party and node size corresponds to in degree.

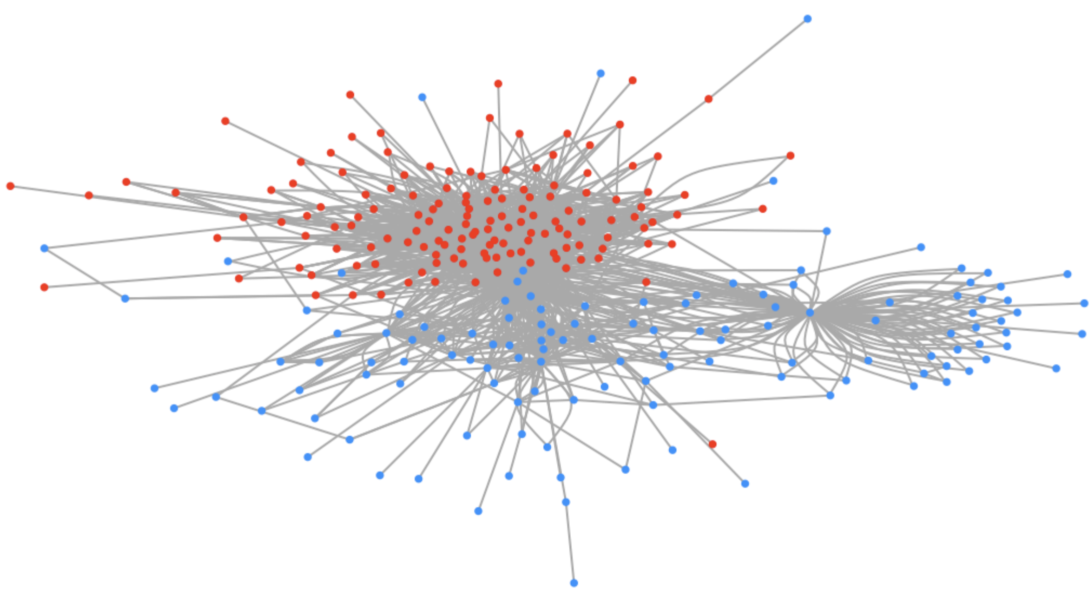


Figure 3: The full network **I**, considered over all time

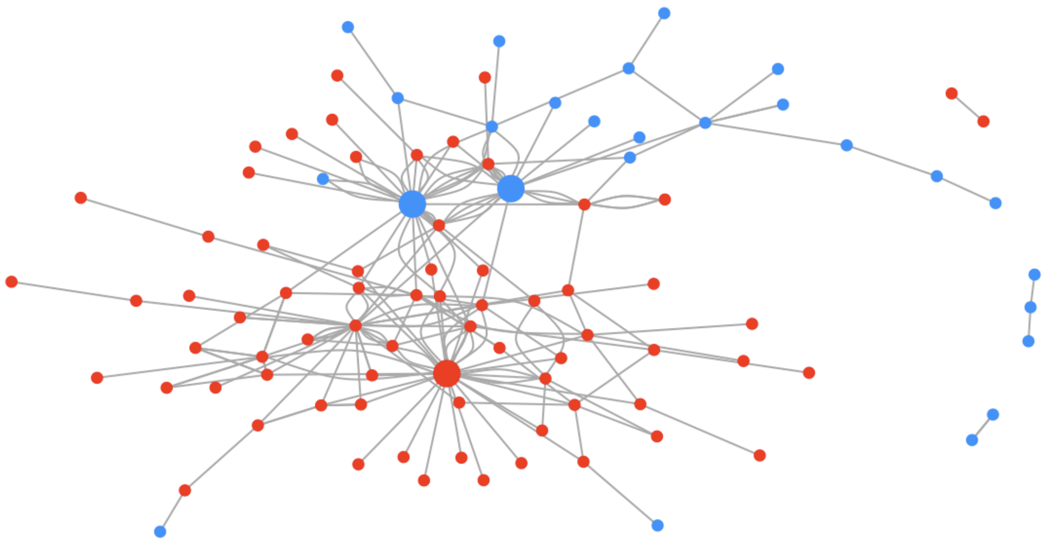


Figure 4: The network **I** with edges restricted to interactions in the week 10-29-19 through 11-04-19, and vertices restricted to those with non-zero degree. Color indicates party, and the nodes with top 3 eigenvector centralities are slightly larger.

three enlarged nodes have the highest eigenvector centralities; they appear to have high interaction overall and with other high interaction nodes.

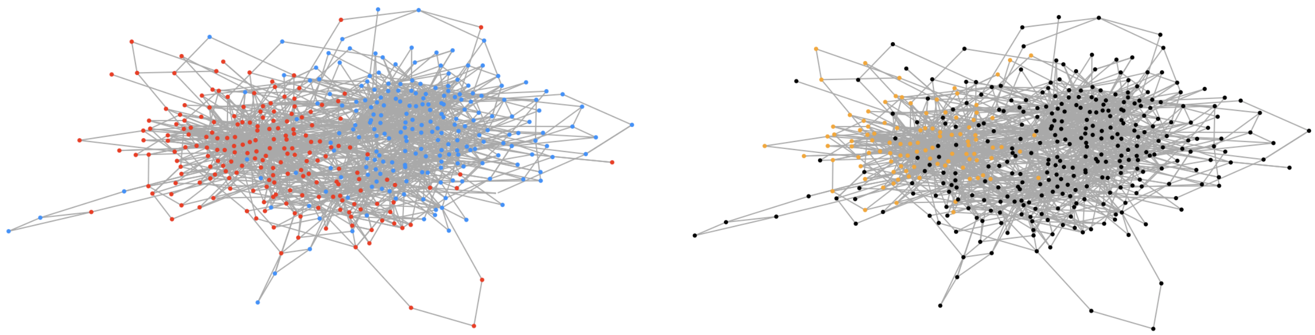
To get a final list of the top influencers, for each day we isolate the three nodes with highest in-degree from the subgraph of network **F**, and the three nodes with the highest eigenvector centrality from the subgraph of network **I**. This produces two lists, one for **F** and one for **I**, where each list entry corresponds to a given date, and is equal to the top three most influential nodes from the subgraphs of **F** and **I** on that day. We consider the 10 users most commonly appearing in the Top **F** list and the Top **I** list, and combine these to arrive at 16 unique users (Table 3.1).

2.2 Assessing degree of polarization

I hope to understand to what extent discussion about impeachment is polarized along party lines, i.e. users communicate primarily within-party on the topic and seldom across the aisle. To do so, I consider the density of **I**, alongside the density of subgraphs $\mathbf{I}_D$ and $\mathbf{I}_R$ which are restricted to nodes from the Democratic and Republican parties, respectively. If density is consistently higher within a particular party than overall, it indicates more inner-party discussion than between-party discussion, and hence higher party-line polarization.

I also assess polarization via community detection. For a given network, I use the fast greedy algorithm `cluster_fast_greedy` to cluster the network into two groups (using `cutat`), and examine how well these two groups align with party lines. If group 1 has 1_R Republican members and 1_D Democratic members, and similarly group 2 has 2_R and 2_D Republican and Democratic members respectively, I define the polarization measure P:

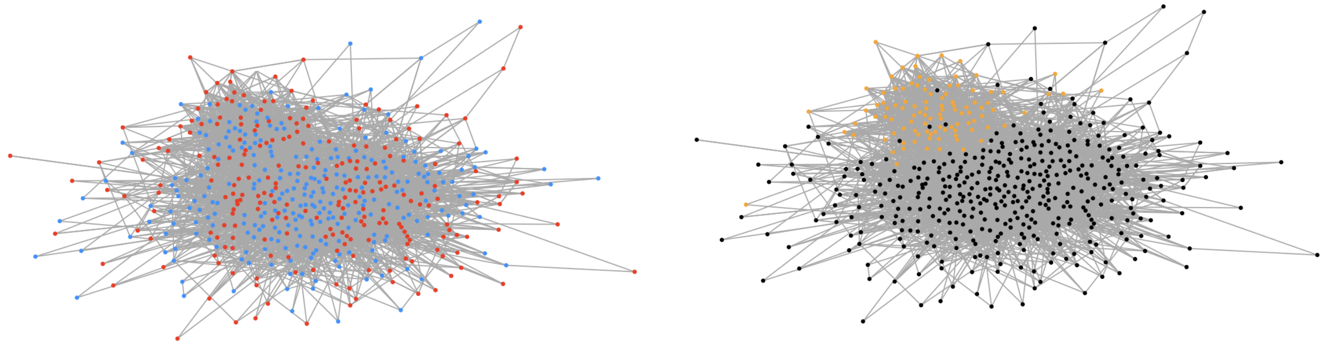
$$P = \frac{1}{2}\left(\frac{\max\{1_R, 1_D\}}{1_R + 1_D} + \frac{\max\{2_R, 2_D\}}{2_R + 2_D}\right).$$



(a) Node color indicates party.

(b) Node color indicates group from fastgreedy clustering algorithm.

Figure 5: Interaction network created using keyword "tax"



(a) Node color indicates party.

(b) Node color indicates group from fastgreedy clustering algorithm.

Figure 6: Interaction network created using "bipartisan"

The first and second summation terms correspond to how homogeneous clusters 1 and 2 are along party lines, somewhere between .5 and 1. We take the average homogeneity value of the two groups to obtain our result P. A network perfectly clustered along party lines would have a value of 1, and a network poorly clustered along party lines (i.e. each cluster has equal amounts of Republican and Democratic users) would have a value close to .5.

For example, Figure 5(a) shows an interaction network constructed similarly to $\mathbf{I}$, but by restricting to tweets containing the keyword "tax", rather than "impeach". Nodes are colored by party (Republicans in red, Democrats in blue), and Figure 5(b) shows the same network where nodes are colored by group, as output from the clustering algorithm (group 1 in black, group 2 in orange). Here, the clustering algorithm aligns fairly well with party lines, and we get a P output value of 0.81. We can also consider the value of P if we construct the interaction network by restricting to tweets containing the keyword "bipartisan", producing the network shown in Figure 6(a), and clustering shown in Figure 6(b). The clustering aligns poorly along party lines, and we obtain a P value of 0.55. This tracks with our intuition, as we would expect highly divisive issues like tax policy to have little inter-party communication and high intra-party communication, but would expect the opposite behavior for tweets discussing bipartisanship.

3 Results

3.1 Users with greatest influence

As described in Section 2.1, we obtain a list of the 16 most influential users, shown in Table 3.1. We note that 7/16 users are current or former members of a Judiciary or Intelligence committee, which are key actors in the impeachment process. Further, 6 out of the remaining 9 members are key party figures (whip, leader, or conference chair). One of the final 3 members is Mark Meadows, one of Donald Trump's closest allies in congress [6]. The reamining two members are Republican House Members Jeff Duncan and Kevin Brady, who do not appear to have any direct linkage to the impeachment proceedings. All but three of these top users are Republicans, and all but two are members of the House (which, notably, has been in charge of the impeachment proceedings for the duration of this study). These results support the measures of user influenced used, as we've selected legislators either critical to the impeachment proceedings or holding positions of high power in congress.

Name	Top **I**	Top **F**	Role
Mark Meadows	1	1	Republican House Member
Kevin McCarthy	1	1	Republican House Leader
Nancy Pelosi	1	1	Democratic House Leader
Steve Scalise	1	1	Republican House Whip
Adam Schiff	1	0	Democratic House Intelligence Committee Chariman
Doug Collins	1	0	Republican House Judiciary Committee Member
Jim Sensenbrenner	1	0	Republican former House Judiciary Chairman
Steve Chabot	1	0	Republican House Judiciary Committee Member
Andy Biggs	1	0	Republican House Judiciary Committee Member
Liz Cheney	1	0	House Republican Conference Chair
Jim Jordan	0	1	Republican House Intelligence Committee Member
Kevin Brady	0	1	Republican House Member
Steny Hoyer	0	1	Democratic House Majority Leader
John Cornyn	0	1	Republican Senate Majority Whip
Marsha Blackburn	0	1	Republican Senate Judiciary Committee Member
Jeff Duncan	0	1	Republican House Member

The follow network **F**, restricted to these 16 users, is shown in Figure 7. The three Democratic users all follow each other, and similarly many of the Republican members follow each other, but there is little inter-party following, with no Democratic users following Republicans, and only two Republicans following any of the Democrats. The interaction network **I**, restricted to these 16 users, is shown in Figure 8. Despite having few inter-party follows, Representatives Pelosi and Schiff have some of the highest interactions among these top users.

Figure 9 shows the average number of tweets on impeachment for these top 16 users, over several months and separated by party. Referencing table 1.3, we can understand the uptick in tweets around the formal impeachment inquiry announcement (09-24), and when the House formalized the impeachment inquiry (10-31). We see another dramatic uptick in discussion as the public hearings go on (from 11-13 through 11-21), and finally as the house releases its inquiry report (12-03). Though both Republican and Democratic users exhibit similar trends in this manner, the average number of tweets is typically higher for Republican users than for Democratic.

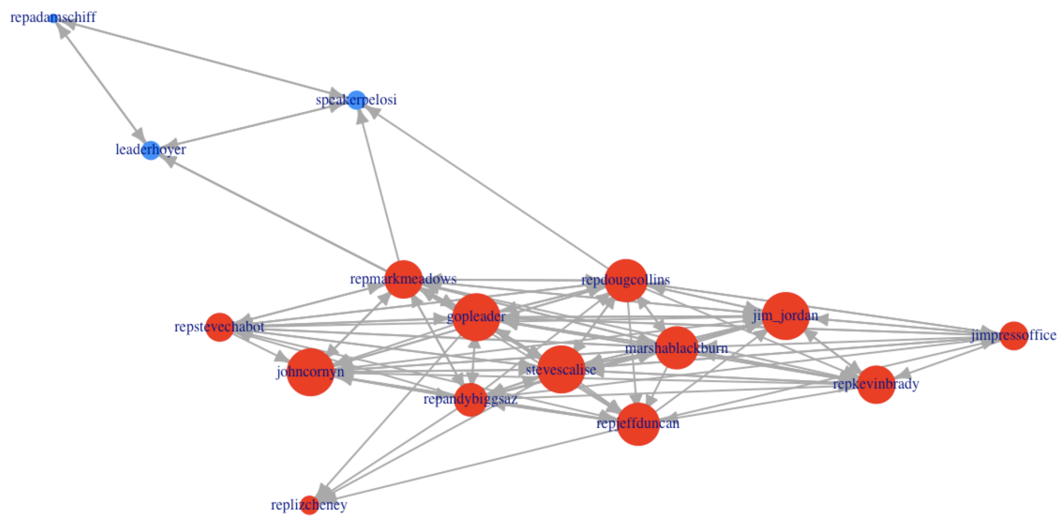


Figure 7: Follow network **F**, restriced to top 16 most influential users. Node color indicates party (Republicans in red, Democrats in blue), and node size is proportional in-degree

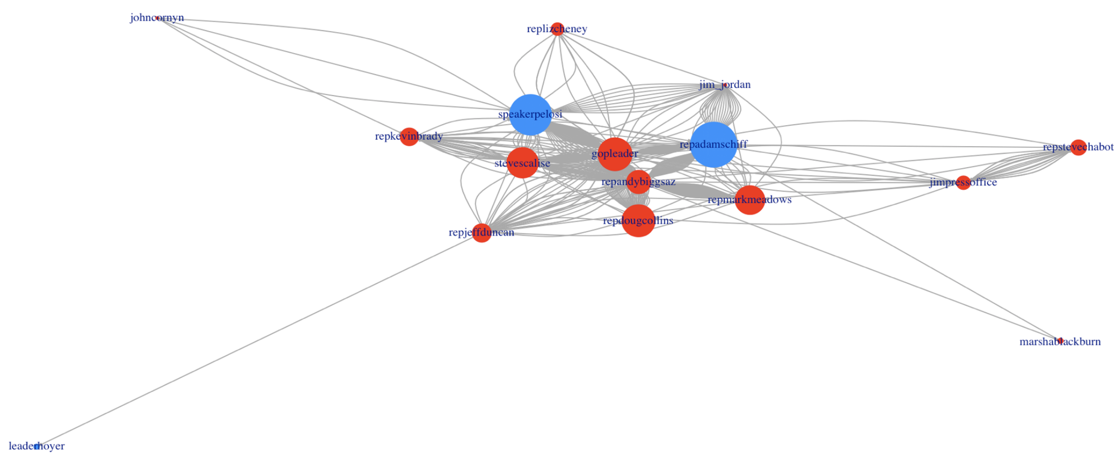


Figure 8: Interaction network **I**, restriced to top 16 most influential users. Node color indicates party (Republicans in red, Democrats in blue), and node size is proportional to number of tweets on impeachment.

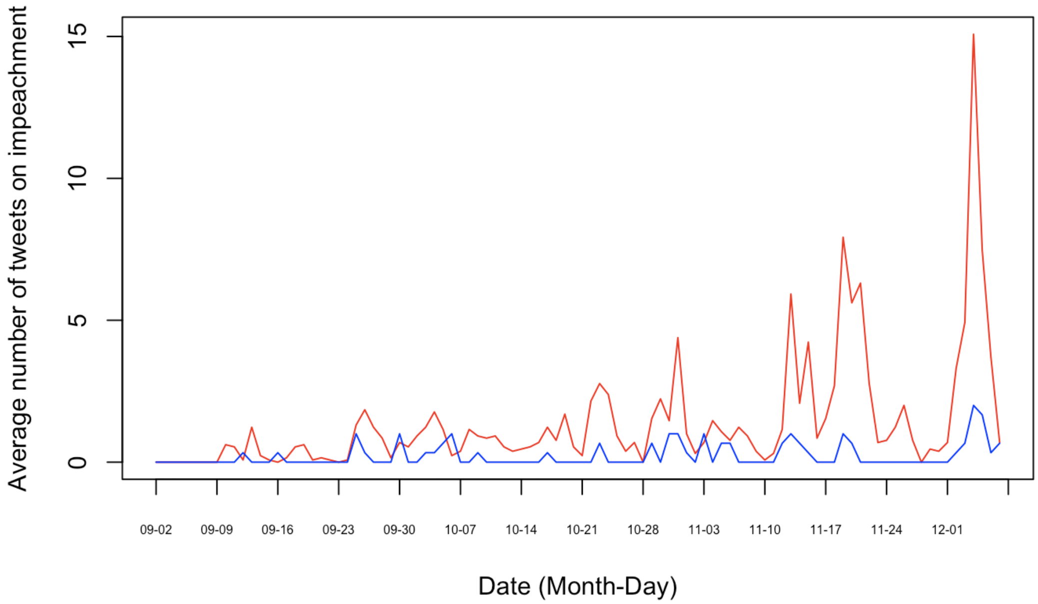


Figure 9: Average number of tweets on impeachment per day for Republican (red) and Democratic (blue) members of the top 16 users (Table 3.1)

3.2 Polarized Nature of Discussions

Figure 10 shows the density over time of the interaction network $\mathbf{I}$, along with the subgraphs $\mathbf{I}_R$ and $\mathbf{I}_D$ restricted to Republican and Democratic users, respectively. Again referencing table 1.3, we see that the first peak, largest for $\mathbf{I}_R$, is towards the beginning of the Ukraine scandal (9-09-19 through 9-18-19), as investigations began, news of the whistleblower complaint surfaced, and subpoenas were issued. On 10-03-19, when Donald Trump admitted he had hoped Ukraine would investigate the Bidens, and suggests China do the same, a large peak occurs in $\mathbf{I}_D$, and a small dip occurs in $\mathbf{I}_R$. On 10-12-19, when reports state that Sondland will testify "no quid pro quo" from Trump, $\mathbf{I}_D$ dips towards 0, while $\mathbf{I}_R$ reamins steady. Over the course of public hearings, and immediately afterward, $\mathbf{I}_R$ remains steadily higher than $\mathbf{I}_D$. We can conclude that key political events coincide with increases in party-specific increases or decreases in intercommunication. One could speculate that upticks in $\mathbf{I}_D$ correspond to events politically damaging to the president, decreases in $\mathbf{I}_D$ correspond to events damaging to the impeachment case, and upticks in $\mathbf{I}_R$ correspond to Republican outrage over the legality of impeachment proceedings.

Using the methodology described in Section 2.2 on our interaction network $\mathbf{I}$, we obtain a polarization measure of $P = 0.92$ (Figure 11). We conclude that discussion on impeachment is a highly divisive topic, as other common policy terms like "education" and "healthcare" produce lower values of $P = 0.62$ and $P = 0.67$, and even controversial terms like "abortion" and "immigration" only get values of $P = 0.77$ and $P = 0.78$, respectively.

4 Conclusions

In this project, I've developed a methodology to extract key users from dynamic networks built around Twitter interactions and follow relationships. Restricting the analysis to the topic of im-

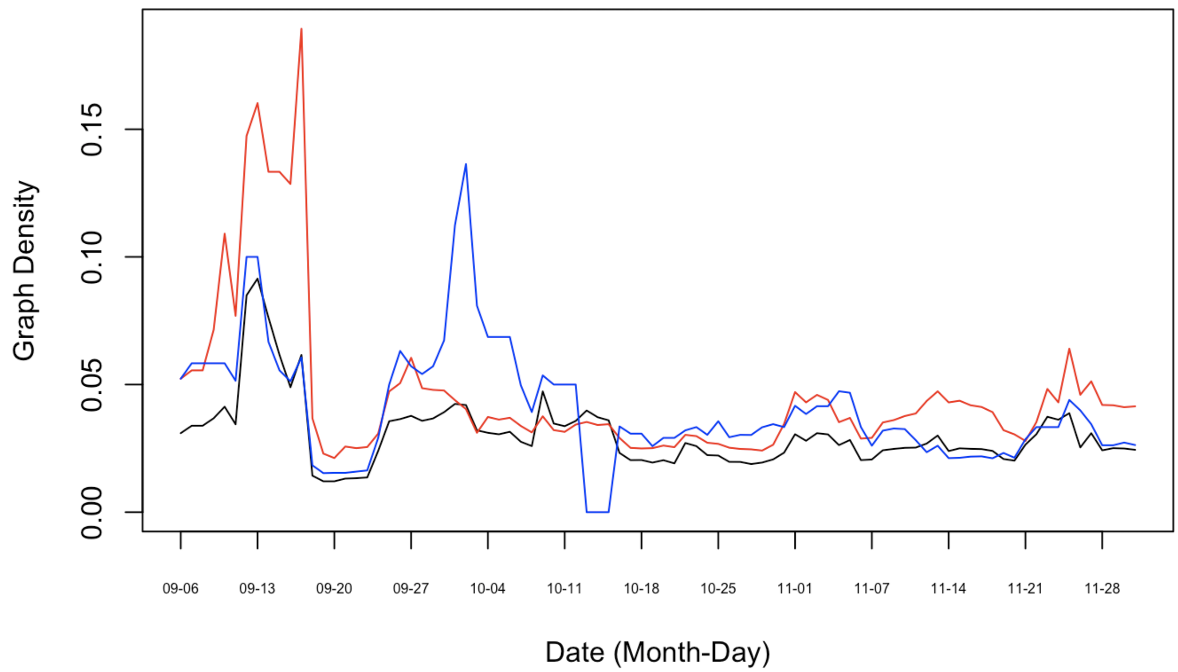
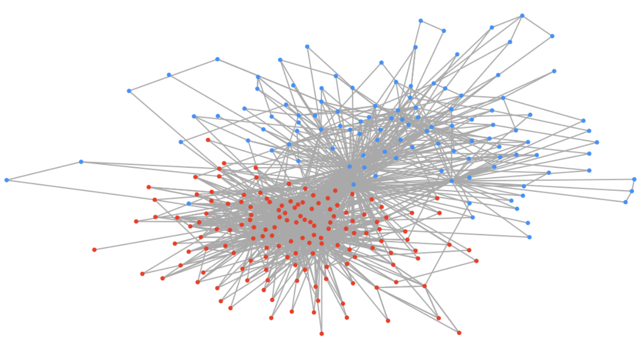
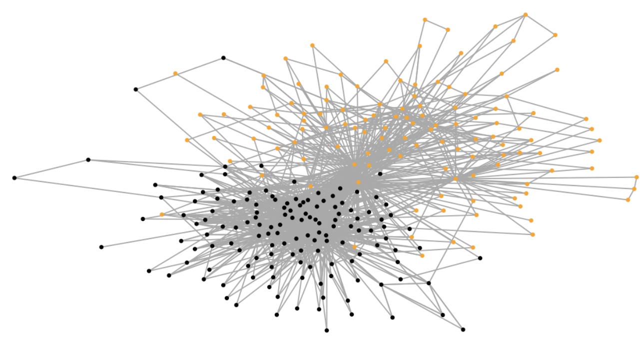


Figure 10: Network density over several months for $\mathbf{I}$ (black), $\mathbf{I}_R$ (red), and $\mathbf{I}_D$ (blue).



(a) Node color indicates party.



(b) Node color indicates group from fastgreedy clustering algorithm.

Figure 11: Interaction network $\mathbf{I}$, clustering on keyword "bipartisan"

peachment, I successfully isolated several key players in the impeachment process, and analyzed their Twitter behavior over time. I conclude that the overall discussion centers around members of the House rather than the Senate, and that Republicans appear to have a larger presence in the Twitter conversation. The quantity of tweets from these top users is coupled with key impeachment events, as we see increases in tweets on impeachment shortly after these events.

I've also build a method to understand party-based polarization in the discussion of impeachment on Twitter. The P measure represents how polarized along party lines a topic is, based on the interaction network and party makeup of its nodes. I found that impeachment is a particularly polarized topic, with little inter-party communication. I also found that inter-party discussion, estimated by density of the network subgraphs, seems to increase and decrease over time depending on the political favorability of events.

Future work on this topic could take into account tweet likes, an interaction type I did not include in this analysis. One could also include the president as a node, and analyze how different parties interacted with him over time. Finally, one could study hashtags, urls, and overall content and sentiment of tweets to better understand the relationships users have to this topic.

References

[1] M. A. Smith, L. Rainie, B. Shneiderman, and I. Himelboim, "Mapping twitter topic networks: From polarized crowds to community clusters," Feb 2014.

[2] K. et al., "Exploring twitter communication dynamics with evolving community analysis," 2017.

[3] A. R. M. Teutle, "Twitter: Network properties analysis," in *2010 20th International Conference on Electronics Communications and Computers (CONIELECOMP)*, pp. 180–186, Feb 2010.

[4] A. Lamb, "Twitter network analysis: identifying influencers and innovators," Oct 2013.

[5] F. B. d. Sousa and L. Zhao, "Evaluating and comparing the igraph community detection algorithms," Oct 2014.

[6] Rachael, K. Cheney, E. Johnson, E. Johnson, A. Isenstadt, A. Restuccia, and N. Cook, "Meadows would give trump a skilled brawler in the white house," Dec 2018.