**Analysis of Twitter Interactions Between US Legislators Regarding SARS-CoV-2**

*Daniela Raygadas and Caroline Harrison*

**Literature Review:**

In recent years Twitter has become one of the most popular social media platforms and that allows individuals to interact with one another and create a dialog about current events and other important topics of conversation. It allows users to have direct contact with followers and share information to them in an efficient manner. Additionally, the news consumption and political dialog has shifted online and onto sites like Twitter and other social media platforms[11]. As a result, Twitter has become an important tool for governments as a way to reach its constituents.

With this increasing popularity within the political sphere, Twitter has become a major platform for government officials to express their political and personal ideas to the public and each other. Political science researchers have worked to dissect these networks to better understand different structural attributes and relational ties between government members, and between those members and their constituents. Previous studies have identified influential members within governmental bodies and political parties, such as Conover et al.[7] who identified political party clusters of twitter users during the 2010 midterm election,and Cherepnalkoski et al.[3] that broke down the European Parliament to identify influential members.

There are many ways individuals can directly interact with each other on Twitter, including by following, retweeting, liking, and mentioning, and therefore many ways to build networks based on the interaction between twitter users. Conover et al.[7] and Cherepnalkoski et al.[3] both concluded that networks built by retweets between users are the strongest indicator of influence on Twitter, as retweets tend to show more agreement with a user than other forms of interaction. In another study from Sadri et al.[8], identifying information sources and the flow patterns of information through Twitter during planned special events, such as sporting events and concerts, have used retweet networks as the relational ties in which they build their networks, and identified participating members through specific tweet entities, such as hashtags and mentions

For this case study, we propose a new perspective where we look at the legislative branch of the United States and the discussions regarding the novel COVID-19 pandemic amongst its members. The COVID-19 virus has had an unimaginable impact on the world, causing a pandemic that has dramatically disrupted our daily lives and impacted the world economy. On March 16, Google recorded coronavirus as the most searched term. It has dominated the news cycle, and governments have had to come up with ways to mitigate the impact and provide accurate information to their constituents about what is being done in their response.

We hope to take the ideas presented in building a network based on retweet and planned special events, and apply them to an unplanned event such as the COVID-19 pandemic, will uncover members who are critical to the spread of information during a time when new facts and legislative efforts are being discussed every day.

**Research Question and Motivations:**

The purpose of this project is to analyze the growth of conversation surrounding the COVID-19 virus over Twitter, within the legislative branch of the United States. Starting from when the first case was reported in the United States on January 18th, 2020 to April 30th, 2020. Members in this community often use Twitter as a platform to express official news and legislative efforts regarding the virus, as well as personal thoughts and opinions of those efforts. The focus of this project is to try and discover particular individuals from the members of Congress that have been more successful in communicating over Twitter information regarding the COVID-19 virus. To better understand the communities where the individuals show more influence, we also look at the explicit communities based on their political party affiliation and chamber membership. Then compare if these individuals who are considered influential regarding COVID-19, are also influential outside of COVID-19.

The proposed research questions are:

1. Who are the effective influencers regarding COVID-19 within these explicit communities?
2. Are these individuals also the effective influencers within these explicit communities outside of COVID-19 related information?

**Data Description:**

This data set comes from an ongoing data collection project in GitHub revolved around collecting the tweets of all members of the United States Congress – this includes members of the House of Representatives, members of the Senate, and all official committees currently run by those members. Each day the data is collected and then posted to the public GitHub repository. The available data goes back to June 2017. For each tweet, the information included in the available dataset includes the tweet\_id, the screen name and name of the member who tweeted, link to the tweet on twitter, full text of the tweet, the time that the tweet was sent, and the source (iPhone or other).

With this information we were then able to re-scrape Twitter in Python using the Twython package to gather additional information only for the tweets that happened between January 18th, 2020 to April 30th, 2020. This additional information would have been much more difficult to collect using only our original dataset, including entity information (hashtags, replies, embedded links), and retweet information (source of original tweet, tweet\_id, etc.).

Link to original dataset: <https://github.com/alexlitel/congresstweets/blob/master/data/2020-03-20.json>

**Network Description:**

The tweets collected in the above section were then used to build our retweet network. We built a weighted and directed graph where each node is a twitter user, and edges are generated by retweet activity and are pointed in the direction of the individual who retweeted the original tweet. Edges are weighted by the number of times a user retweeted another user in the network. The decision to build network edges on retweet data was based on research suggesting that retweets are a strong indicator of agreement and influence in Twitter[3,7].

Our two base networks contained all members of the Houses of Congress and retweet activity between those members. The first network was built using all retweet activity between those members, while the second network was built using only COVID-19 related retweet activity. Retweet edges were only included if there was some entity in the tweet (url, hashtag) that contained a reference to COVID-19, or any social or legislative responses that were directly related to COVID-19. Entities such as hashtags and urls have been used in previous studies to identify topic related tweets, rather than text processing tweets.[8]

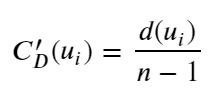
It has been established that there tends to be more polarization in retweets than with other forms of interaction between members of communities on Twitter[7]. From this, we can assume that within our network, explicit communities based on political affiliation and chamber membership will show a more complete network as they are more likely to interact with each other than with outside members of their explicit communities. As with our base networks, the explicit communities based on an actor’s chamber membership (House or Senate), and political party affiliation (Democrat or Republican) were each built using both all the retweet activity, and only COVID-19 related retweet activity. Breaking down these base networks into different explicit communities may be helpful to better understand the party or chamber communities where they show more influence.

We implemented our network using the Networkx Python Packages and all visualizations and calculations in this analysis were facilitated using this package. Please see Figures 1 - 5 in the appendix for visualizations of each of the community networks built.

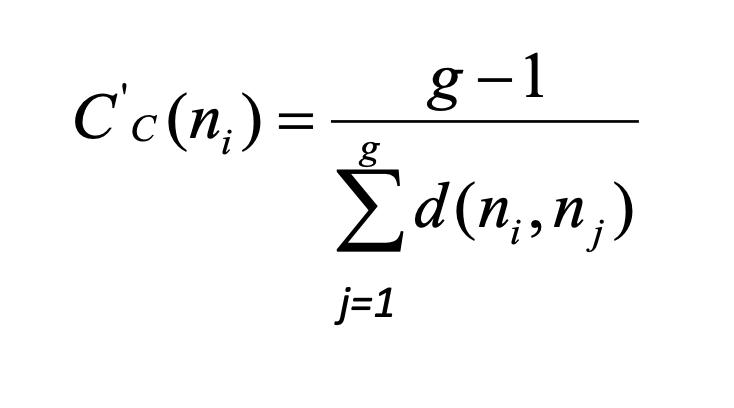
**Applied Analysis Methods:**

We chose to calculate several measures of centrality to better understand the structure of our networks and understand how actors played different roles within their network. Centrality measures identify actors in the network who occupy critical positions, with different measures of centrality highlighting different critical areas, including how many links a node has to other nodes in the network, how close a node is to all other nodes in the network, and how important the node is to the flow of information[6]. In directed networks such as this, strong centrality measures indicate a level of prestige as well as influence within the network, as there is a choice of interaction between nodes[1]. The measures we used for this study were out-degree centrality, closeness centrality, betweenness centrality, and PageRank centrality. For this discussion, we used the PageRank centrality to ultimately decide which actors were most central in their network, as it focuses on how popular and visible an actor is to other influential members in the network, showing us that they stand out within a network of people who are all striving to be influential themselves[2]. Due to the scale and large number of networks built for this analysis, all calculations were performed using the Networkx Python packages centrality algorithms. Please see Appendix 6 for code relating to the use of Networkx in our centrality calculations.

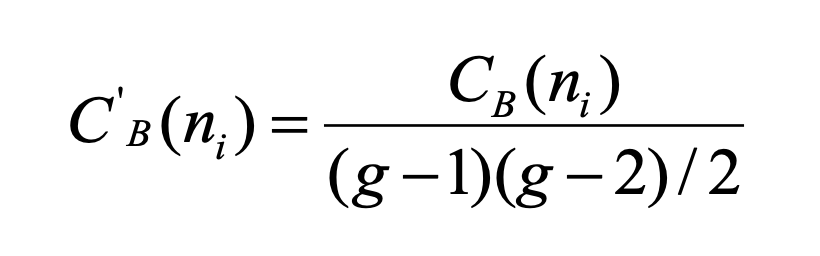
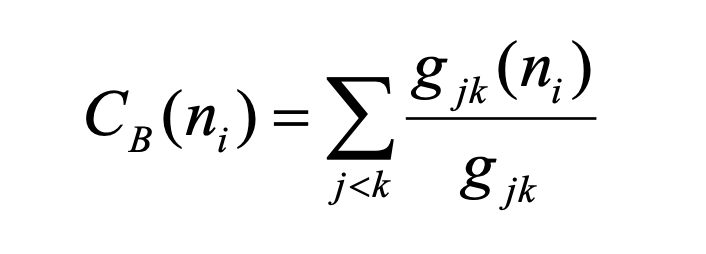
Out-degree centrality measures the number of out-degree edges from each node in the network and can indicate how popular an actor is. The choice to use out-degree centrality rather than degree centrality or in-degree centrality was based on the retweet network structure, as an out-degree edge represents an actor having retweeted you, and an in-degree edge represents you retweeting another actor in the network. In this case, the out-degree centrality would be a stronger indicator of prestige than the in-degree centrality. Having a high out-degree centrality measures indicates that the actor is prominent and that they are highly visible to other actors in the network[1]. In order to compare the out-degree centrality measure across networks, it needs to be standardized. The standardized out-degree formula[1] is:



The second centrality measure we computed was closeness centrality, which measures a node’s average geodesic length between all other nodes in the network. This centrality measure was selected for our analysis as it indicates when an actor is close to all other actors in the network. In a directed network, the measure of closeness between ni to nj may not be equal to that of nj to ni[1]. Nodes with a high closeness centrality can reach most other nodes in the network relatively quickly. Closeness based centrality is also an indicator of independence, suggesting that if a node is central it does not need to rely on other nodes to spread information. A node that is close to all other nodes has a more direct and efficient impact on other actors in the network because there are fewer intermediaries involved in the flow of information presented by that actor[5]. The formula for calculating closeness centrality was standardized, and was also scaled by the ratio of the fraction of the actors in the group for whom the node is reachable. This scale factor was proposed as a way to mitigate the effects on this centrality measure for nodes who do or do not reach every other node in the network[1]. This formula[1] is presented here:

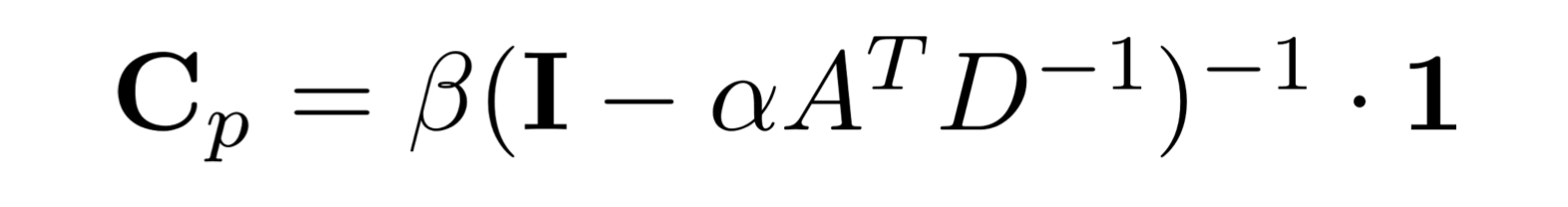


The third centrality metric we measured was betweenness centrality, which is a measure of how many geodesics a node in the network lies on. Nodes with a high betweenness centrality measure are often crucial for the flow of information. Betweenness centrality is also indicative of the contribution that actor *j* has in transmitting the information presented by actor *k* to other nodes in the network[4]. The standardized formula[1] for computing a node’s betweenness centrality value is here:



a ) Betweenness Centrality Formula b) Standardized Betweenness Centrality Formula

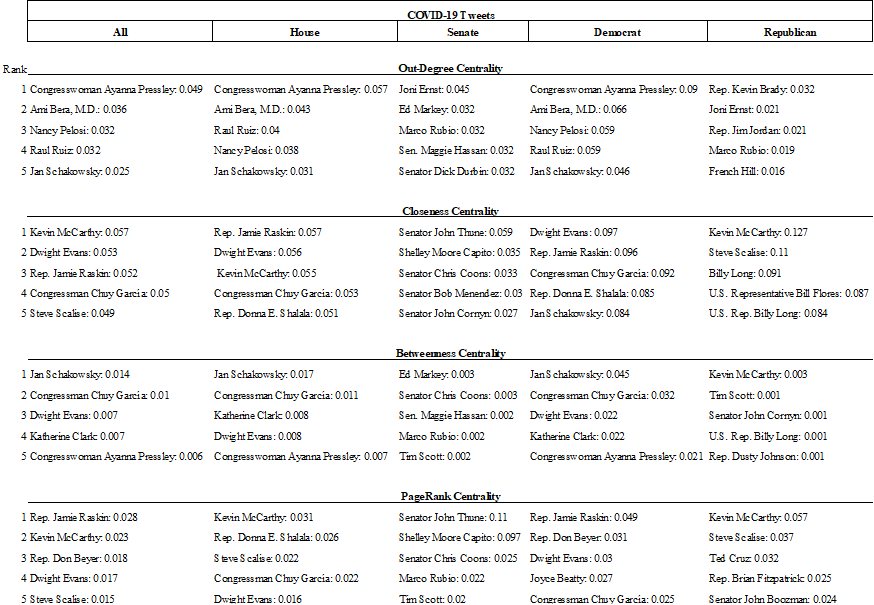
The final centrality measure we computed was PageRank centrality, which is a standardized centrality measure that incorporates not only the degree centrality value of the node in question, but the degree centrality value of the node’s neighbors. This then assumes that actors who interact with other influential members of a network must be influential themselves. PageRank improves upon two other centrality measures, Eigenvector centrality and Katz centrality, to better handle directed networks and reduce the effects of having potentially one highly influential member in the directed network[2]. For each node, PageRank divides its value of passed centrality by it’s out-degree centrality, so that each adjacent node gets a fraction of that node’s centrality. The formula for PageRank centrality is here[2]:



**Results and Discussion:**

The results are split into two tables, based on our two base networks of all retweet activity and COVID-19 related retweet activity. Each column in the tables represents the explicit communities that we broke our networks out into, and contains the four centrality measures with the top five ranking actors.

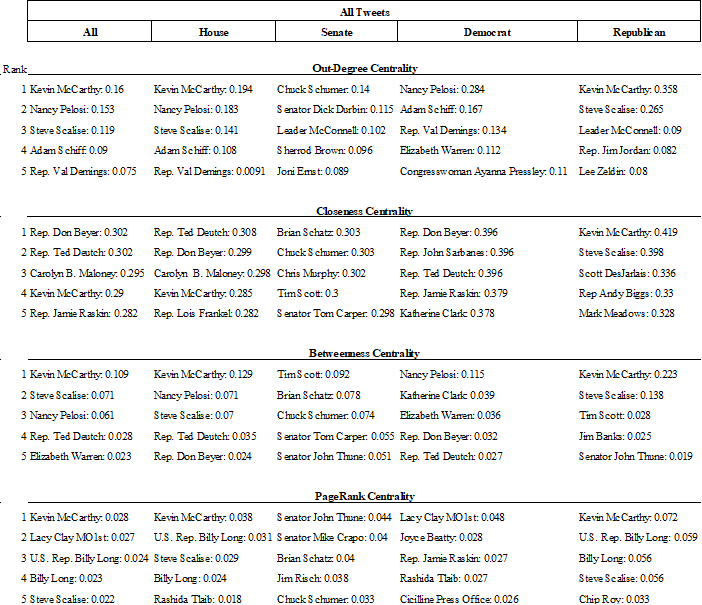
Table 1: Results from COVID-19 Retweet Activity Network



The results from Table 1 show that in the COVID-Retweet network, the members with the highest PageRank score in their communities were Jamie Raskin (All - 0.028, Democrat - 0.049), Kevin McCarthy (House - 0.031, Republican - 0.057), and Senator John Thune (Senate - 0.11).

Kevin McCarthy and Jamie Raskin’s PageRank scores were higher in the party affiliation communities than they were for the overall network, showing that they are not as influential to members outside of his party as they are within, indicating that they may be a more polarizing figures within the network[7]. Individuals that had a high PageRank score within the Senate community, did not have a high Page Rank score in their party affiliation communities or in the overall network. Showing that, the members of the Senate community are not influential to members outside their chamber. One interesting thing we noticed was that influential members who had a high PageRank centrality score did not necessarily have a high Out-Degree centrality score, meaning that people with a high Out-Degree may have been connected to more peripheral actors. Examples of members who had a high Out-Degree centrality score who did not rank high in other areas were Congresswoman Ayanna Pressley, and Nancy Pelosi. We also noticed that members who were highly ranked for Closeness and Betweenness centrality tended to have a higher overall PageRank centrality scores, which can be seen from members such as Dwight Evans and Congressman Chuy Garcia, indicating that these may be more correlated with PageRank centrality than Out-Degree centrality. It is also interesting to see that the Speaker of the House, Nancy Pelosi, and Majority Leader of the Senate, Mitch McConnell, did not appear to be highly central figures in their communities, as may be expected by their positions of leadership within Congress.

Table 2: Results from All Retweet Activity Networks



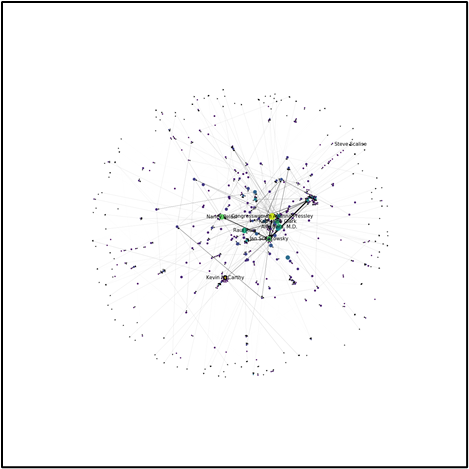
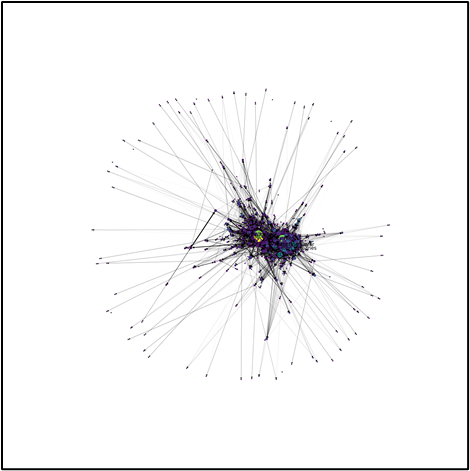
The results from Table 2 show that in the All-Retweet network communities, the members with the highest PageRank scores in their communities were Kevin McCarthy (All - 0.028, House - 0.038, Republican - 0.072), Senator John Thune (Senate - 0.044), and Lacy Clay (Democrat - 0.048). Two overlapping figures from the COVID-Retweet network here are Kevin McCarthy and Senator John Thune. It can be concluded that Kevin McCarthy is one of the most central members of this network, as he appeared as the top PageRanked member for all explicit communities he belongs to, as well as a top ranking individual for the other measures of centrality calculated here to better understand the structural properties of these networks. Apart from Kevin McCarthy, similar structural trends that appeared in the COVID-Retweet networks also appeared in the All-Retweet network, including seeing members with a high out-degree centrality score have lower PageRank centrality scores. Similar to what we saw in the COVID-Retweet networks, Kevin McCarthy and Lacy Clay’s PageRank scores were also higher in the party affiliation communities than they were for the overall network, showing that they are not as influential to members outside of his party as they are within, indicating that they may be a more polarizing figures within the network[7].

Based on these results, we can state that the effective influencers regarding COVID-19 information within these explicit communities were Jaime Raskin, Kevin McCarthy and Senator John Thune. From these individuals, the only members that remained a central figure outside of COVID-19 related information was Kevin McCarthy and Senator John Thune. Other interesting discoveries through this analysis were that members who held positions of leadership and influence in Congress, such as Speaker of the House, Nancy Pelosi, and Leader of the Senate, Mitch McConnell, did not necessarily translate that influence onto Twitter. And that central members who had high PageRank scores in political party affiliation networks did not have as high of a PageRank score in their overall networks, indicating that those people may be more polarizing political figures within Congress and hold less influence within the overall networks.

**Future Work:**

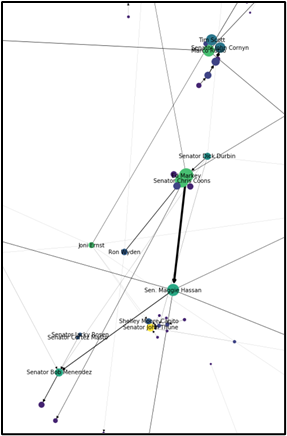
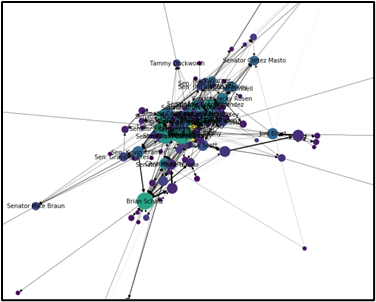
Just as previous studies[3,9] have used sentimental analysis to uncover the attitude of different communities towards various issues, we would like to apply sentimental analysis on the tweets by influential members to see how these explicit communities feel towards COVID-19, and understand what relevant information is being shared to the people.

**Appendix:**

 ****

a) COVID-Retweet Network b) All Retweet Activity

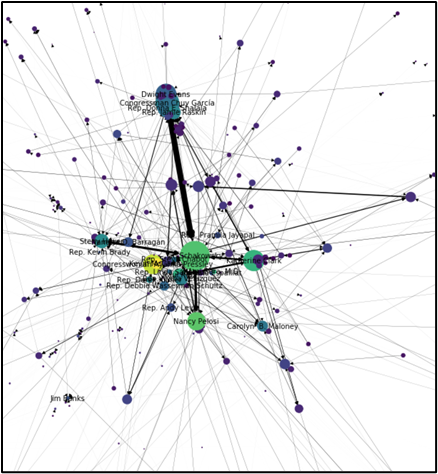
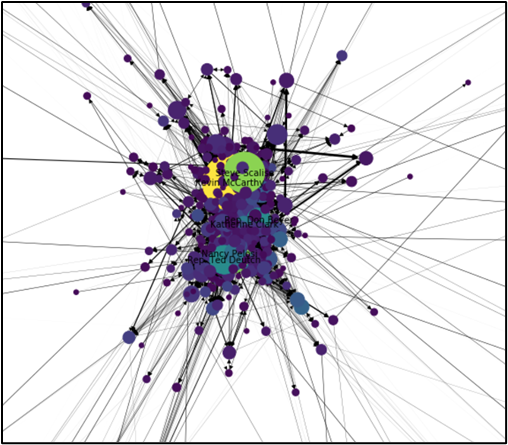
Figure 1. All Legislative Network for COVID-19 Retweet Activity and All Retweet Activity

** **

a) COVID-Retweet Network b) All Retweet Activity

Figure 2. Senate Network for COVID-19 Retweet Activity and All Retweet Activity

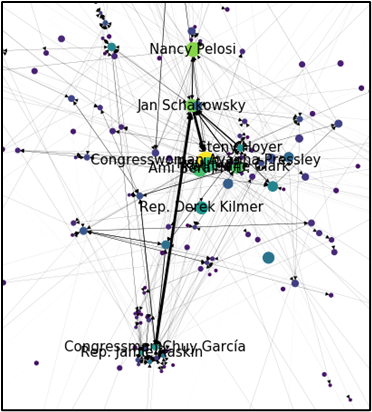
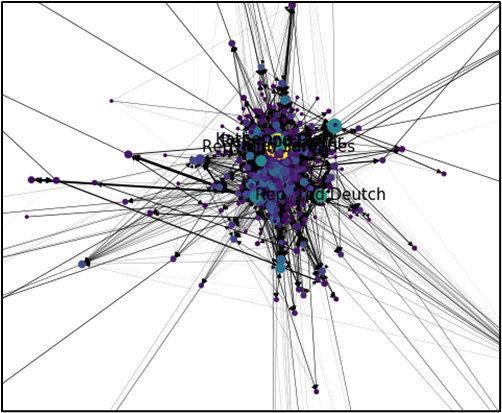
Focused on Central Members

** **

a) COVID-Retweet Network b) All Retweet Activity

Figure 3. House Network for COVID-19 Retweet Activity and All Retweet Activity

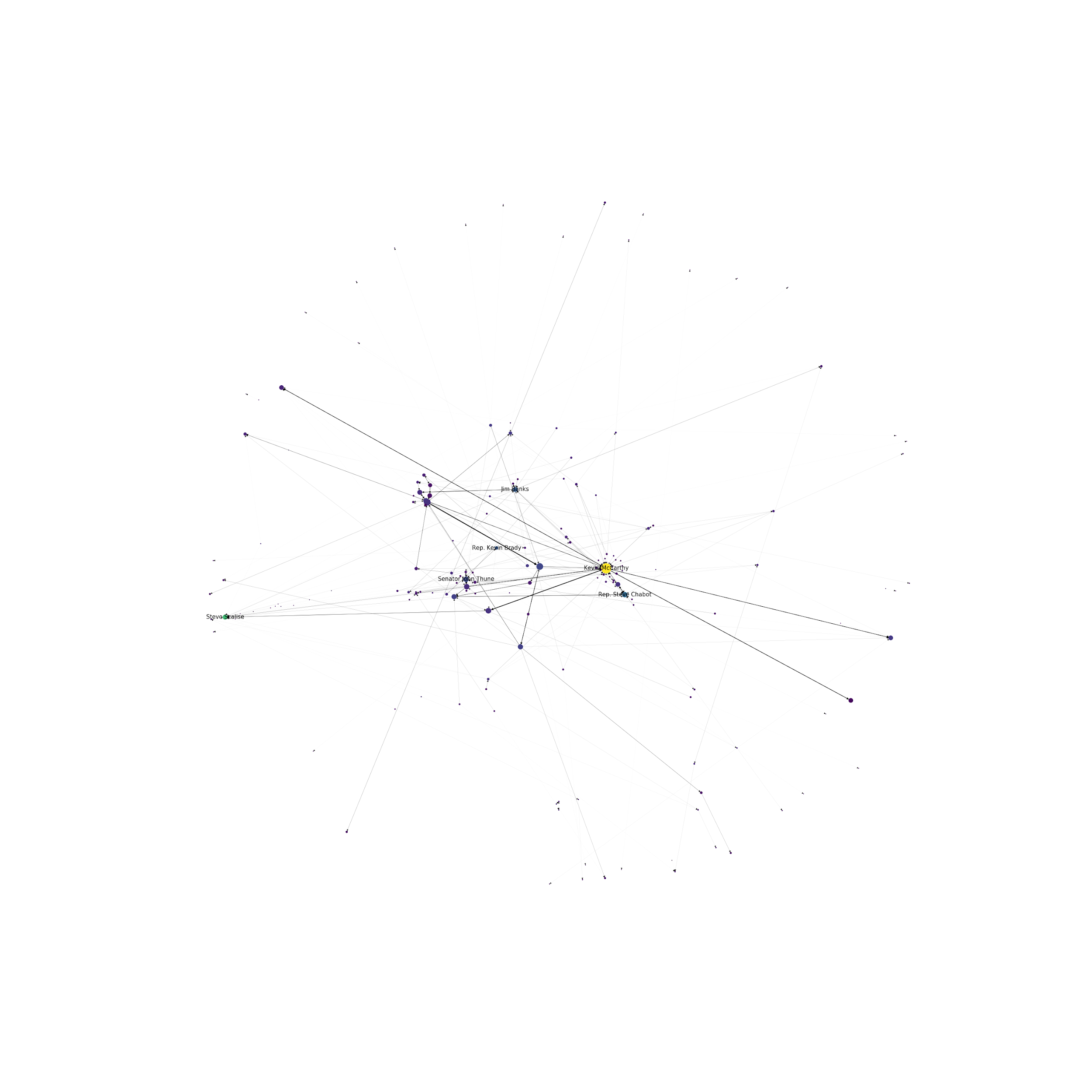
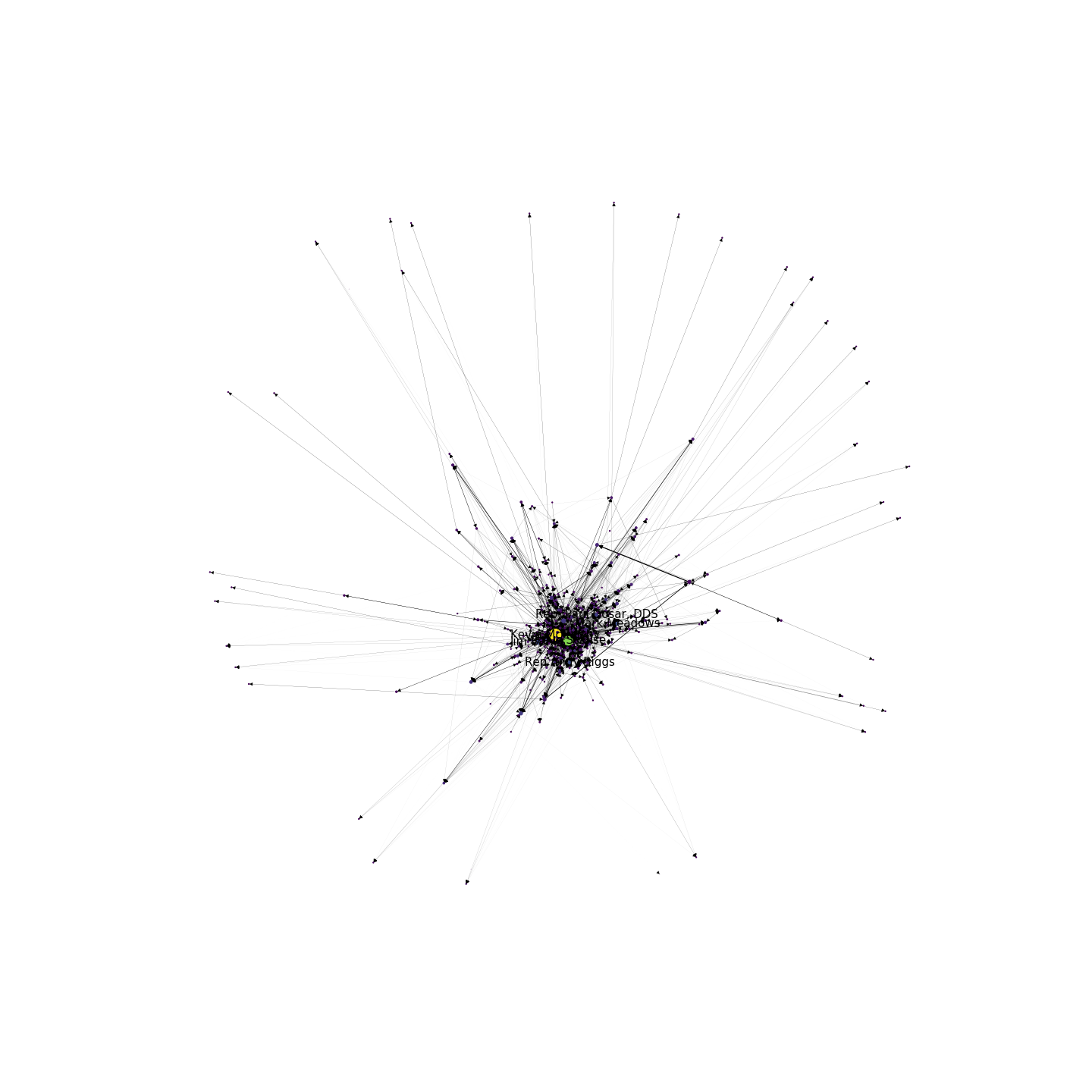
Focused on Central Members

a) COVID-Retweet Network b) All Retweet Activity

Figure 4. Democrat Network for COVID-19 Retweet Activity and All Retweet Activity

Focused on Central Members

a) COVID-Retweet Network b) All Retweet Activity

Figure 5. Republican Network for COVID-19 Retweet Activity and All Retweet Activity

Focused on Central Members

Figure 6. Code for Calculating centrality scores using Networkx Python Package

def centrality\_measures(Network, count):

out\_degree\_centrality = sorted([(v, c) for v, c in dict(nx.out\_degree\_centrality(Network)).items()], key = lambda x: x[1], reverse = True)

closeness\_centrality = sorted([(v, c) for v, c in dict(nx.closeness\_centrality(Network, wf\_improved = True)).items()], key = lambda x: x[1], reverse = True)

betweeness\_centrality = sorted([(v, c) for v, c in dict(nx.betweenness\_centrality(Network, weight = True, normalized = True, endpoints = True)).items()], key = lambda x: x[1], reverse = True)

pagerank\_centrality = sorted([(v, c) for v, c in dict(nx.pagerank(Network)).items()], key = lambda x: x[1], reverse = True)

centrality\_list = [out\_degree\_centrality,closeness\_centrality,betweeness\_centrality, pagerank\_centrality]

centrality\_name\_list = ['Out-Degree Centrality','Closeness Centrality','Betweenness Centrality', 'PageRank Centrality']

for c,n in zip(centrality\_list, centrality\_name\_list):

print(n)

for i in range(count):

print(f'{i+1}. {c[i][0]}: {c[i][1]: 0.3f}')

print('\n')

**References:**

1. Wasserman, Stanley, and Katherine Faust. *Social network analysis: Methods and applications*. Vol. 8. Cambridge university press, 1994.

2. Zafarani, Reza, Mohammad Ali Abbasi, and Huan Liu. *Social media mining: an introduction*. Cambridge University Press, 2014.

3. Cherepnalkoski, Darko, and Igor Mozetic. "A retweet network analysis of the European Parliament." *2015 11TH International Conference On Signal-Image Technology & Internet-Based Systems (SITIS)*. IEEE, 2015.

4. Freeman, Linton C. "Centrality in social networks conceptual clarification." *Social networks* 1.3 (1978): 215-239.

5. Friedkin, Noah E. "Theoretical foundations for centrality measures." *American journal of Sociology* 96.6 (1991): 1478-1504.

6. Valente, Thomas W., et al. "How correlated are network centrality measures?." *Connections (Toronto, Ont.)* 28.1 (2008): 16.

7. Conover, Michael D., et al. "Political polarization on twitter." *Fifth international AAAI conference on weblogs and social media*. 2011.

8. Sadri, Arif Mohaimin, et al. "Analyzing social interaction networks from twitter for planned special events." *arXiv preprint arXiv:1704.02489* (2017).

9. Sluban, Borut, et al. "Sentiment learning of influential communities in social networks." *Computational social networks* 2.1 (2015)

10. Landherr, Andrea, Bettina Friedl, and Julia Heidemann. "A critical review of centrality measures in social networks." *Business & Information Systems Engineering* 2.6 (2010): 371-385.

11. Conover, Michael D., et al. "Predicting the political alignment of twitter users." *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*. IEEE, 2011.

12. Kwak, Haewoon, et al. "What is Twitter, a social network or a news media?." *Proceedings of the 19th international conference on World wide web*. 2010.