Examining the 114th Congress

Kira Lowe

Introduction

Upon reading any given issue of any given newspaper, one is almost certain to be confronted with stories asserting that partisan gridlock is stalling progress in the U.S. House and Senate; it often seems as if nobody can work together. It is true that ideological extremity measured in roll-call voting is significant, especially among younger legislators, Republicans, and legislators who won a relatively high vote share to secure their election (Ladewig 2010). Does this same ideological extremity in roll call voting behavior isolate them in legislative actions?

In 2006, James Fowler published a study that shows us that a better measure of the true alignment and impact of our representatives is in measuring the networks that are developed between them in sponsorship and cosponsorship of legislation, a measurement he calls social "connectedness." Increased connectedness, larger networks of cosponsors of legislation within the House and Senate, increases likelihood of success in passing legislation (Fowler 2006a) (See also Harward and Moffett 2010). Because the long-accepted cornerstone of democracy is that voters hold representatives responsible for their performance via elections, there should be a strong correlation between connectedness and reelectability, expressing that we expect our legislators to be passing legislation. However, partisan ideology has been known to skew opinion and therefore electoral actions (Tilley and Hobolt 2011).

Given that all this information, I decided to examine the most recent full term of congress and perform a social network analysis to visualize their connectedness. I will explore additional visualizations within this social network to see if there is any obvious cause or effect resulting from each member's connections.

Process

<u>Data Collection</u> The first step in performing a congressional social network analysis is collecting data on each bill's sponsor and cosponsors (if any). At the time of analysis, the 115th congress

had not yet ended. Instead, I examined the 114th congress which ran from January 2015 through January 2017. Fowler has maintained a website containing data from his original papers in 2006, recently augmented with data on congressional bills through the 114th Congress provided by GovTrack (Fowler 2017). From Fowler's site, I was able to download information on the bills from that congress in the form of sponsorship and cosponsorship data in a CSV file. Because I also wanted to examine personal information about each legislator such as their political party, gender, and other statistics, I needed to gather additional data.

Fowler's website did not contain data on each congressional member, so I turned to GovTrack, the original source of data on the bills of this particular congress. GovTrack produces a report card for each congress, containing statistics and analysis such as committee positions, laws enacted, and leadership and ideology scores (GovTrack 2017). I downloaded and aggregated this data into an Excel spreadsheet containing eleven different characteristics of each member of the Senate and the House of Representatives.

<u>Initial Attempt at Analysis</u> Because I am not yet adept at either the level of statistical analysis or the programming skills required to perform a social network analysis, Professor Karl Ho assisted me by providing some code for analysis using R. After receiving the code, I first reviewed it to seek understanding of the process and what it required. The first observation I made was that there should be two data separate files. The first file needed to contain data on the bills and in what way they connect to each legislator; this file will be referred to as the "link" file going forward. The second file needed to contain any potentially relevant information about each legislator; this file will be referred to as the "nodes" file going forward. I also observed that I would need to format any text data as zeros or ones. Both of these adjustments were made in the data cleaning process. The code proceeds to weight and normalize the data, establish a

sponsorship network based on the information in the links file, and then begins to visualize the network using data from both the links file and the nodes file.

After I gathered an understanding of what the process would be in R, I began to run the code. From the first attempt, I was able to assign my CSV files to the links and nodes in the code. However, I immediately noticed that I needed to make some changes to my files before I would be able to proceed. I cleaned my data, attempted to run it in R, and repeated the process many times before I had success in producing a visualization.

<u>Data Cleaning Process</u> The first step in my data cleaning process was to adjust the data files to fit my initial observations about the code. I already had established two separate link and node files, but I did not have numerical values in place for text values. For example, the "party" column contained the letter "R" to represent Republican legislators, "D" for Democratic, and "I" for Independents. I changed the column header to reflect one party, then, using conditional logic in Microsoft Excel, I changed all values matching that party to the number one, and all values that did not match that party to the number zero. I repeated this process for the chamber and gender columns. I tried running these cleaner files in R, but I encountered another error.

The next error that needed to resolution through data cleansing sounds elementary but was not an intuitive step for a beginner; I needed to remove any duplicate records and either populate missing fields or remove them as well. This error stemmed from the fact that there were several legislators whose names were spelled inconsistently throughout the original data I downloaded from GovTrack. I had to standardize the spelling and delete the duplicates. After this error was resolved, I was able to get much further in running the code in R.

The final major hurdle I encountered in cleaning and formatting the data was in the lines of code that develop a network between legislators. The original data file had rows of bill names and one column per row with the name of a legislator. This format did not provide a clear directional relationship. In order to develop a connection between a sponsor and their cosponsors, I needed to reorganize the data into columns that connected a sponsor ("from") to a cosponsor ("to"). I used Excel's pivot tables function to rearrange the data and connected any remaining data via Excel's lookup function. My last step was to separate each of my data files by chamber, resulting in two files for the House of Representatives and two files for the Senate. Once these steps were complete, I was able to proceed to the visualization steps in R.

Visualization and Observation After cleaning the data files, I began working on visualizations of the Senate using the igraph package in R (see figure 1 below). The first visualization showed every link between every member of the Senate. All the links are represented by grey arrows and all the Senators are represented by yellow nodes. The image did not clearly show any relationships or even clear directionality.

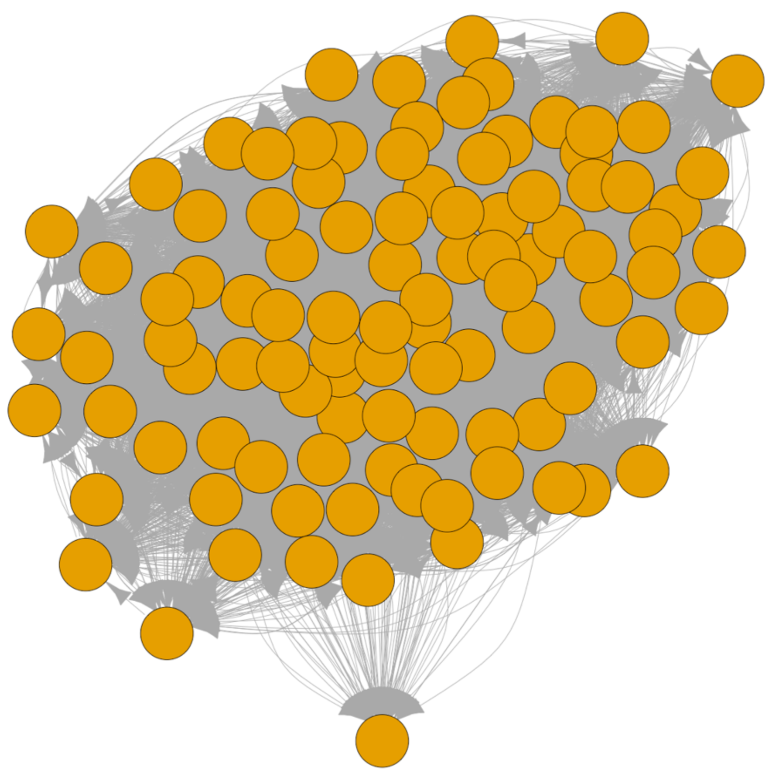

Figure 1

Additional steps had to be taken to produce a more intelligible graph. First, multiple ties between each legislator were removed; for example if Senator A sponsored multiple bills on which Senator B was a cosponsor, the tie between them would only be represented graphically one time. Next, colors were assigned to each senator based on their party's traditional colors; red for Republicans and blue for Democrats. The one Independent represented in the image was classified as a Democrat for these purposes because that is the party with which that Senator caucuses. Finally, size was assigned to each node dependent on the size of their networks, with larger nodes representing Senators with larger networks. The resulting image is much more

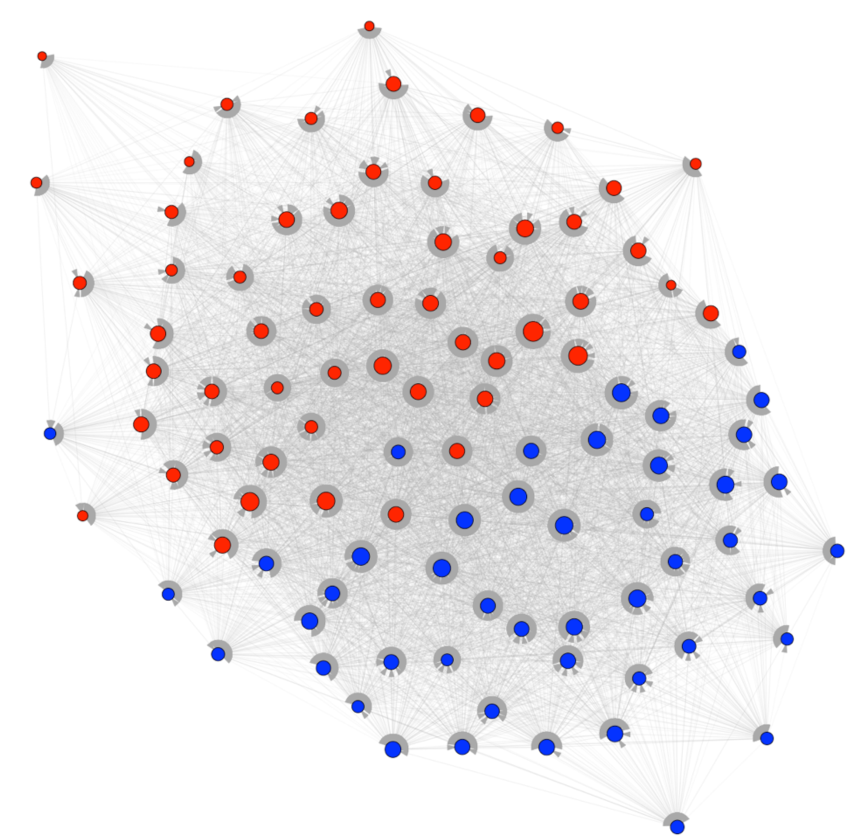

Figure 2

straightforward (see figure 2 above). In this graph, a number of new observations can be made. First, there is a clear divide between parties with very little crossover. Second, we can see that generally, but not always, the larger nodes representing more connected Senators tend to be more central to the graph. Third, there are some nodes that are distanced away from the main cluster indicating fewer connections with nodes central to the graph.

In any directional network graph, it is important to try to achieve the most straightforward image possible using a force-directed layout algorithm. A force-directed layout algorithm can increase the clarity of a graph in different ways; they can minimize crossover of ties (called edges), minimize the tension between nodes by bringing them as close together as possible without causing more edge overlap, or aim to evenly distribute nodes throughout a graph (Ognyanova 2016). For this project, I tried using three different algorithms. First, I applied the Fruchterman-Reingold layout (see figure 3 below).

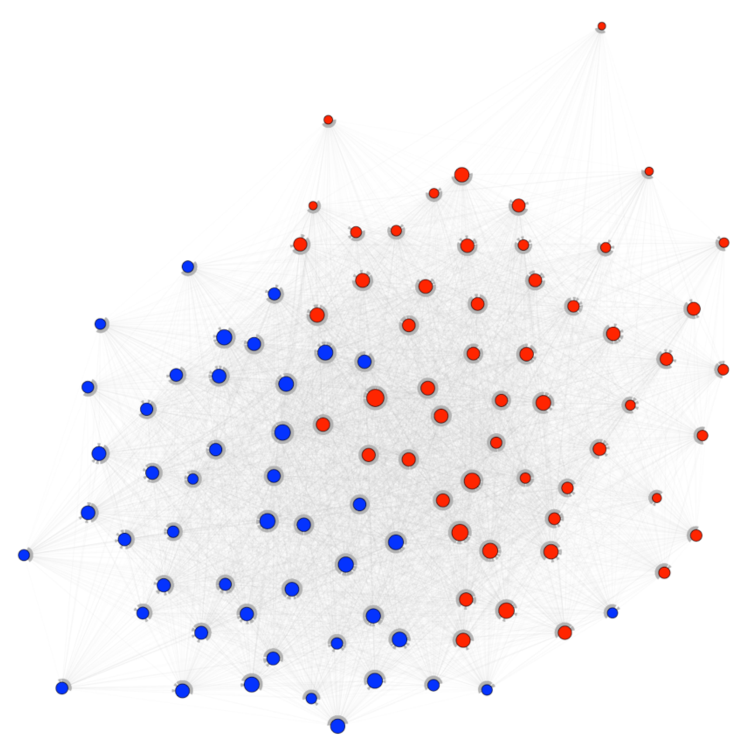

Figure 3

At first glance, application of this layout algorithm did not appear to cause a significant change, however there were some small but positive differences. Not only did the number of crossovers appear to diminish slightly, but also the size of each node became more discernable. This layout is an improvement, but the other force-directed layout algorithms I tried were not as successful. Next, I applied the Large Graph Layout that is intended to give clarity to large networks (see figure 4 below). Again, this layout resulted in subtle differences, but not necessarily an improvement upon the Fruchterman-Reingold layout. As compared to Fruchterman-Reingold, this algorithm shows an increase in edge overlap as well as an increase in crossover between platforms, with a pocket of Republican red nodes deep into the Democratic blue side of the party

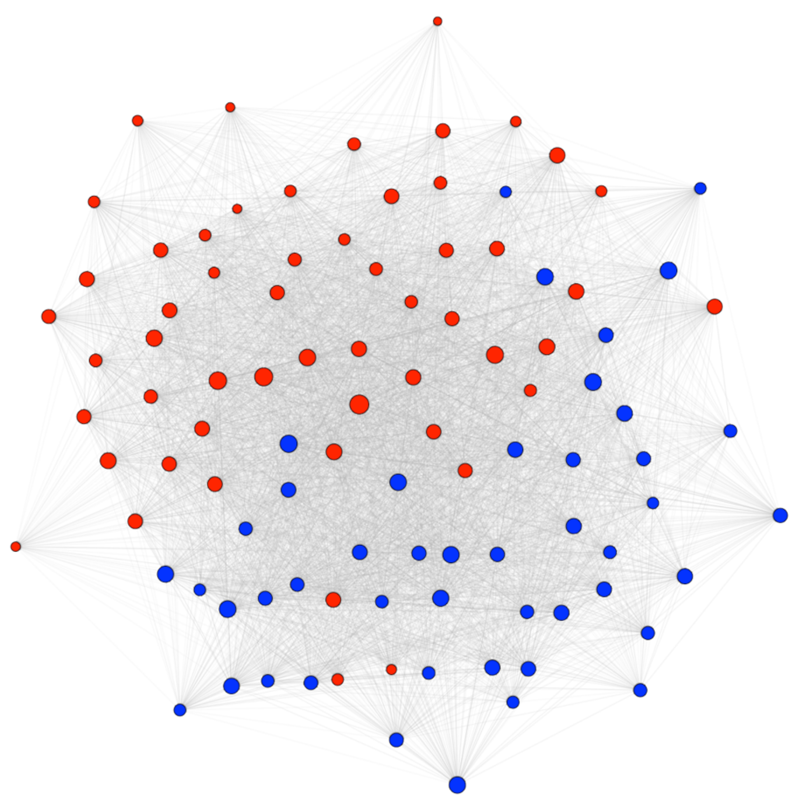

Figure 4

divide. There are only 100 Senators being compared in this analysis which is likely too small a size for the Large Graph Layout.

Finally, I applied the Kamada Kawai layout algorithm (see figure 5). This graph seems to be the most evenly distributed across the plane, however there is a discernable increase in overlap from the Fruchterman-Reingold layout in figure 3. Though all three force-directed layout algorithms were easily discernable, the most sensical among all the visualizations seen thus far is the Fruchterman-Reingold layout because it displays the least amount of edge overlap and maximum clarity in features such as node size and distance between nodes.

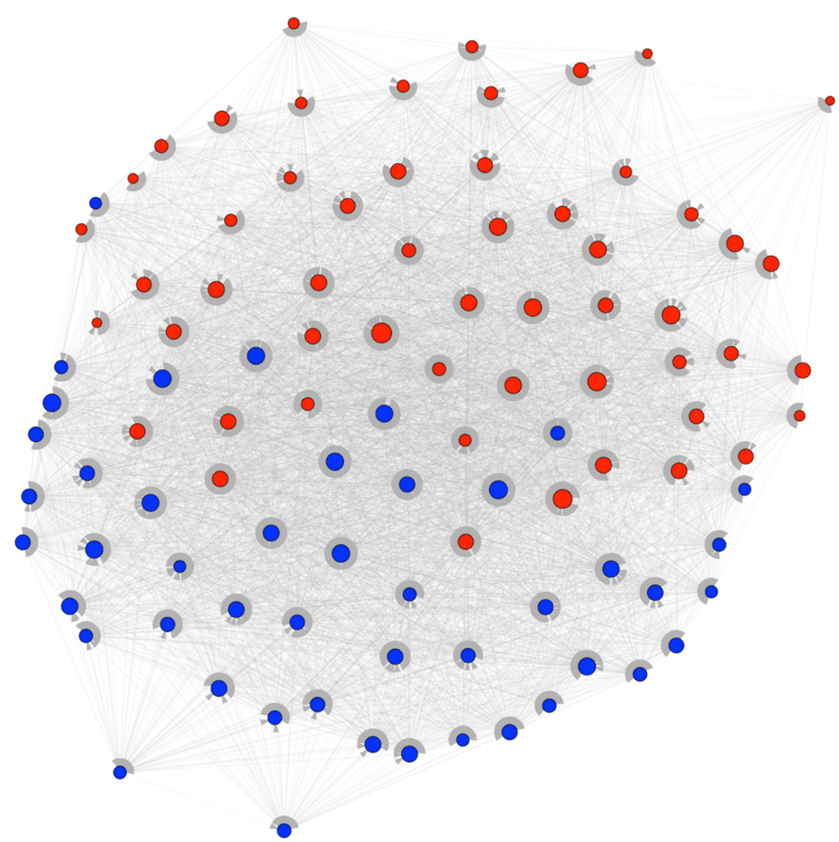

Figure 5

The differences between all three force-directed layout algorithm applications was relatively minimal for the Senate data; however, when applied to the House data, the variations were much more significant.

The initial visualization of networks in the House was far less clear than that of the Senate, even after multiple ties had been removed and color assigned to each party (see figure 6 below). There are a few reasons for this discrepancy in initial clarity; the most significant is that there are a small quantity of Representatives who only served a fraction of their term because they resigned or were replacing a member who had resigned. Those Representatives are pictured far away from the rest of the House because of their relatively few connections resulting in small networks that

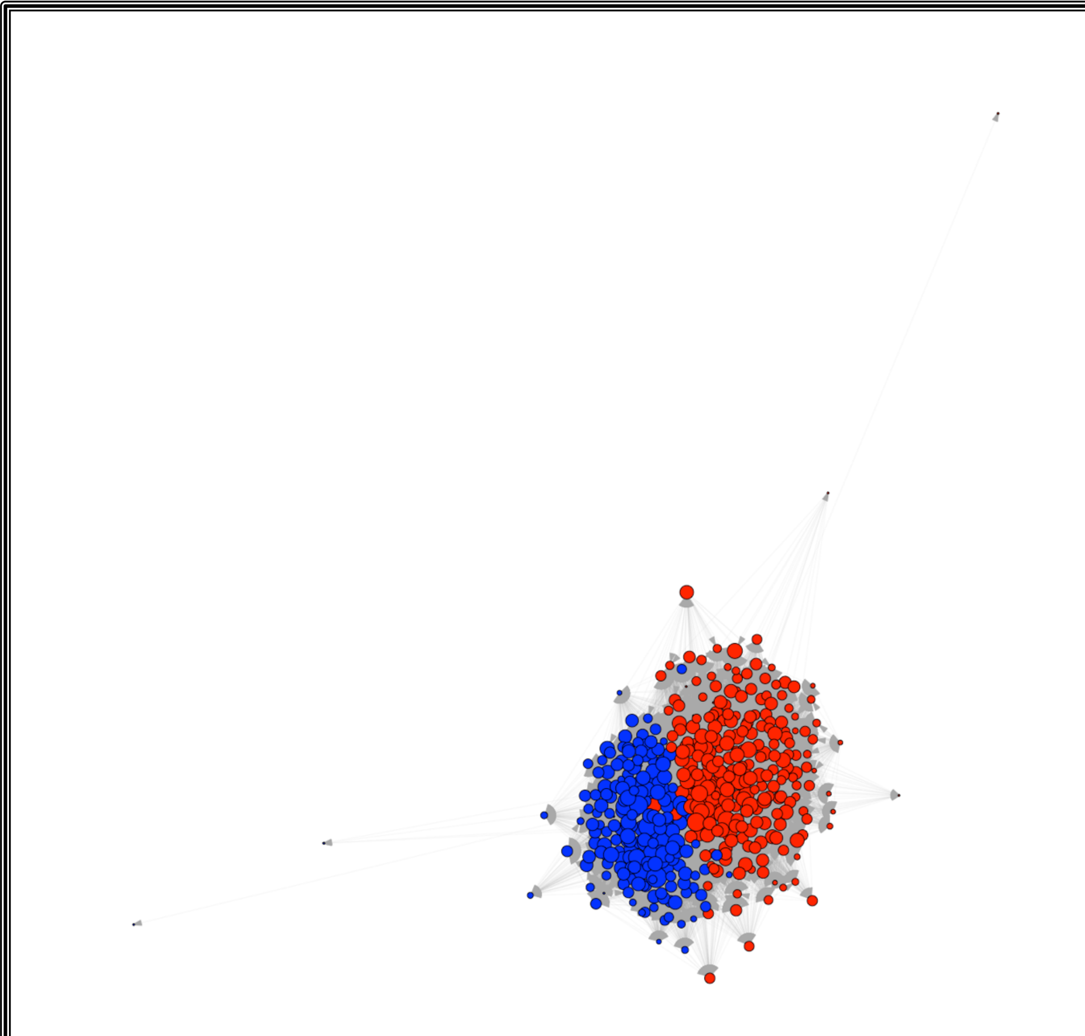

Figure 6

did not have time to develop. Because they were part of the network of other Representatives who did serve the entire term, they were not removed for this paper. Forcing these nodes so far outside the central cluster caused the graph to be compressed and, consequently, each feature becomes less distinguishable.

Although the Fruchterman-Reingold layout algorithm produced the best visualization of the Senate network, application of this algorithm to the House network did not result in significant improvement (see figure 7 below). The outlying nodes are still too far from the central cluster and, while there is minimal additional clarity, this is clearly not an ideal layout.

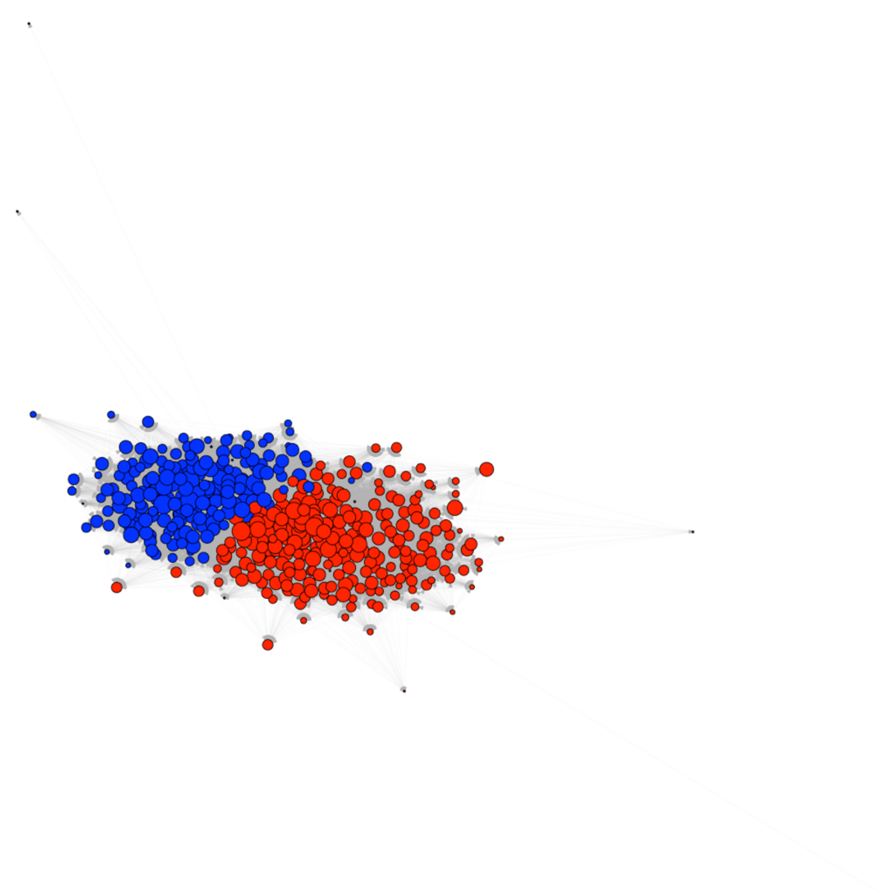

Figure 7

Because there are over four times as many Representatives as there are Senators, the House data files were much larger than those from the Senate. I decided to apply the Large Graph Layout algorithm to this larger data set in an attempt to get a more distinct graph (see figure 7 below). Indeed, this layout does provide far more clarity for the central cluster of nodes. Even though there is a very dense web of edges, the size and placement of each node is much more discernible in this layout. In this graph, as was seen in the Senate graph, a clear divide between parties can be observed as well as what appears to be a loose correlation with node size and graph centrality.

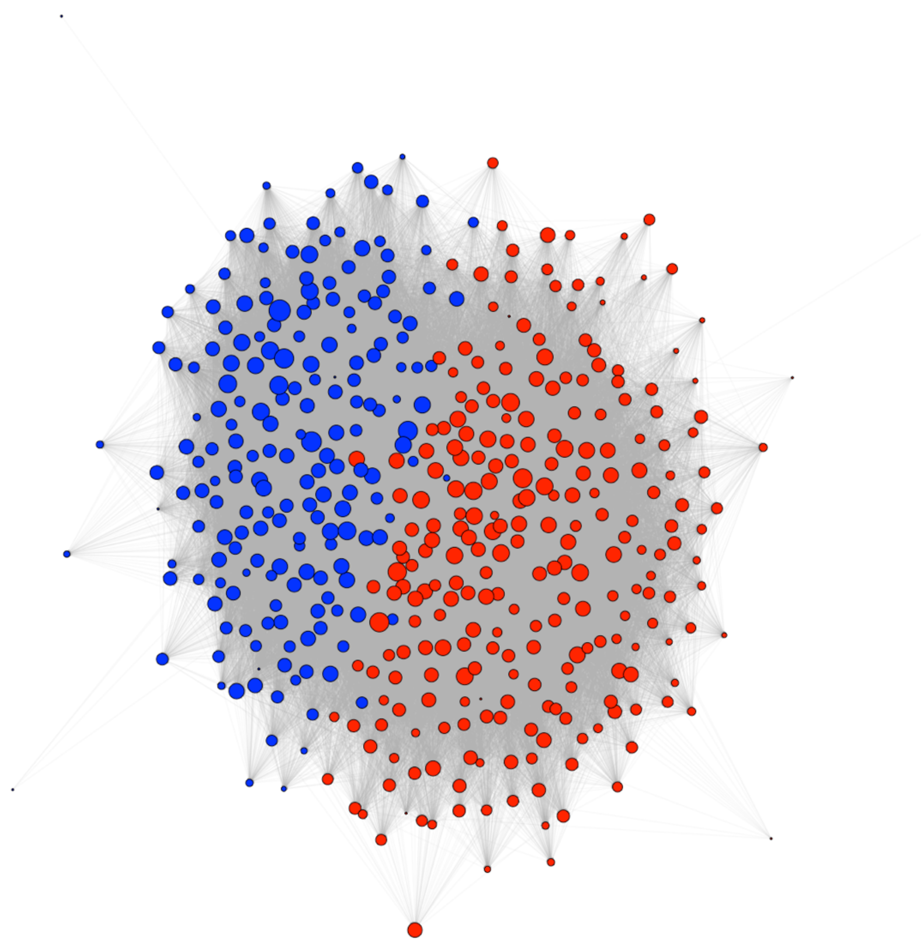

Figure 7

Though the Large Graph Layout appears to have provided the needed clarity in the House graph, I applied the Kamada Kawai layout algorithm to see if it would produce an even more direct graph (see figure 8 below). Obviously, it did not. Using this algorithm, the network of edges became even more dense, the node placement became inconsistent and any relationships are undiscernible. It is clear that the Large Graph Layout format is the best presentation of House data.

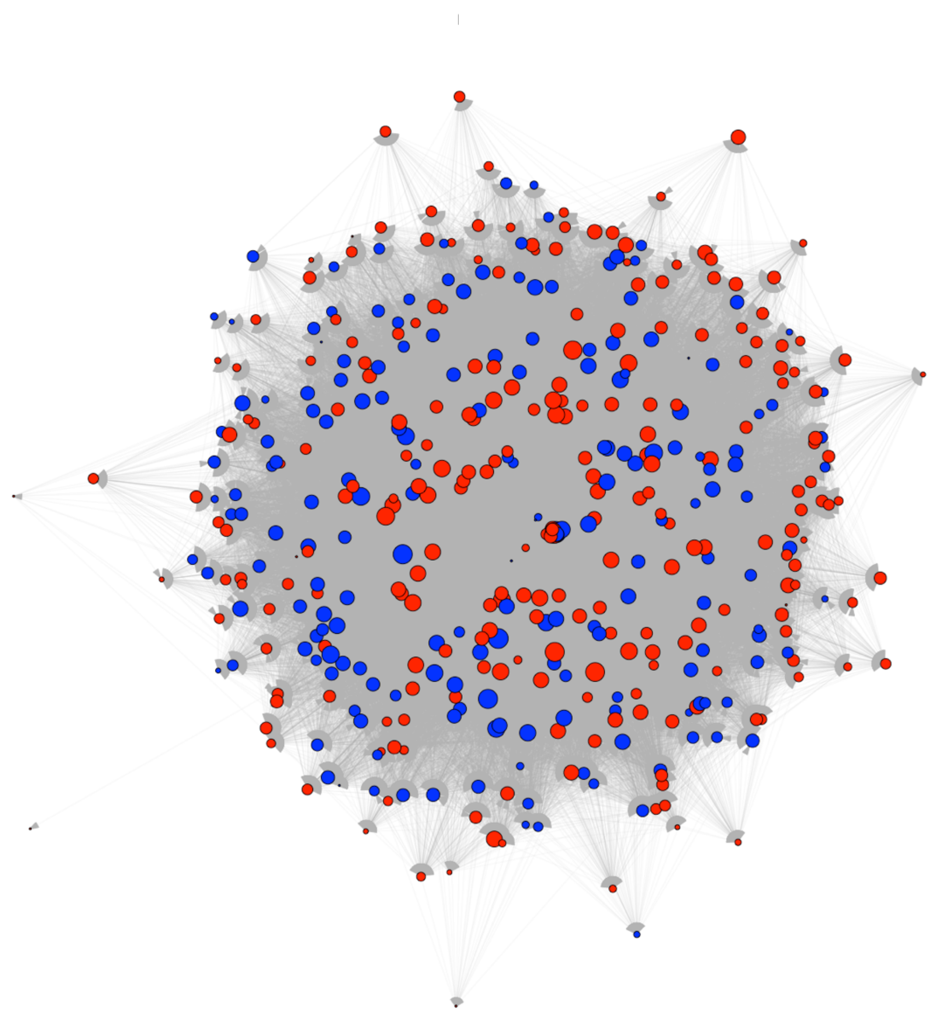

Figure 8

After establishing the best layout for each network and making some initial observations based on network size and party, I began to use color (SAPE 2018) to visualize based not just on party, but also gender and election results for both 2016 and 2018 elections. These visuals will be presented in the next section.

Observations

There were several visuals in particular that I wanted to create in order to see if there was any observable relationship between a legislator's characteristics and their connectedness. These segmentation choices were driven by popular news stories as well as trends in recent elections.

Party First, as illustrated in figures two through seven above, I examined party to see if the partisan divide is as severe as it is purported to be. In both chambers of the 114th Congress, there is a clear, observable separation between Republican and Democratic legislators. While there are a small number of legislators from each party who seem to be more willing than others to network across party lines, with very few exceptions both Republicans and Democrats appear to be creating and operating within networks consisting primarily, or in some cases exclusively, of their own party.

Related to this is another observable difference; both parties have members who are further on the fringes of the graph and members who are more central. The legislators who are more central to the graph are more willing to network across the aisle, whereas those on the edges of the graph are less likely to do so. The resulting inference is that the legislators who are more polar on the graph are also more polar ideologically. Finally, there seems to be some correlation, even if a weak one, that the more ideologically extreme a legislator is, the smaller their network will be. This makes intuitive sense because one would expect a legislator who deviates from the status quo ideologically to have a more difficult time securing cosponsors for their legislation and be less willing to cosponsor legislation.

Gender Another factor I chose to visualize in context of legislative networks is gender. In theory, Americans elect legislators who network effectively in order to pass legislation. After the recent 2018 midterm election, more women were elected to congress than any time in American history. I colorized the networks to see if there was any correlation between a representative's gender and their connectedness. In figure 9 below, I colorized female members of the Senate with a coral color and male Senators with a light blue color. This graph shows that there is little correlation with gender and connectedness or ideology. However, it does provide a clear illustration that there are far more female Senators from the Democratic Party (left side of graph) than from the Republican Party (right side of graph).

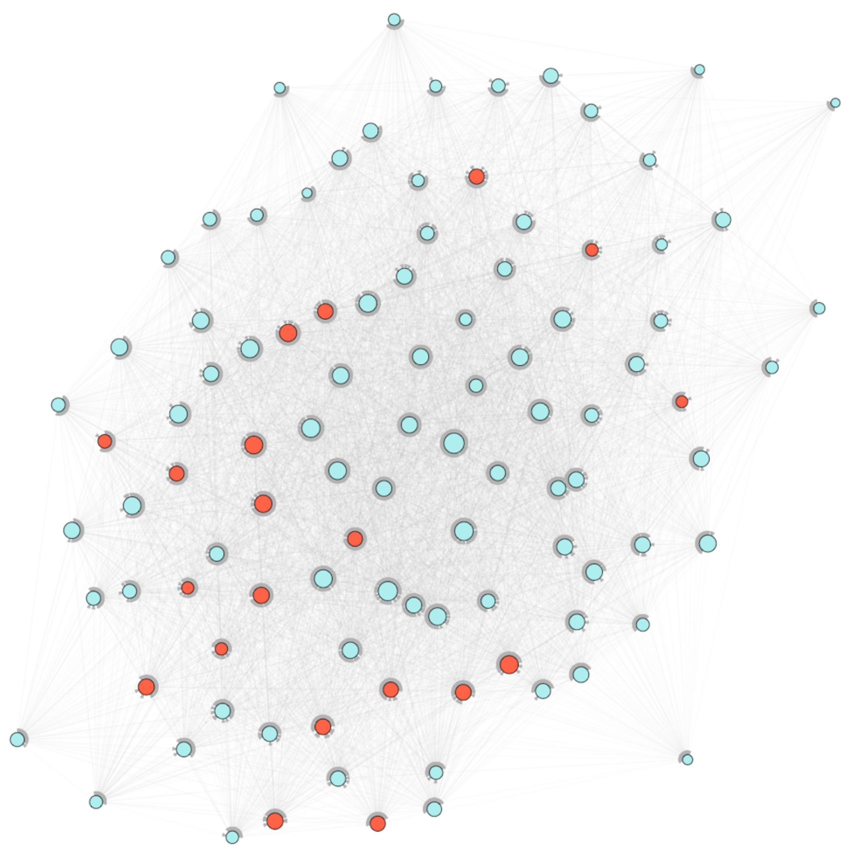

Figure 9

Electrability Finally, I wanted to see if there is any observable correlation between a legislator's connectedness and whether or not they were reelected. Because members of the House of Representatives are up for reelection every two years, I chose to colorize members of the House as light blue if they were reelected or dark blue if they either resigned or were defeated in reelection. I examined results from both the 2016 election (see figure 10 below) which took place during the 114th Congress as well as the 2018 election (see figure 11 below) which resulted in a number of House seats changing hands.

The 2016 election graph (see figure 10 below) did not show much, if any, correlation between connectedness and reelectability. The vast majority of Representatives ran for reelection and also

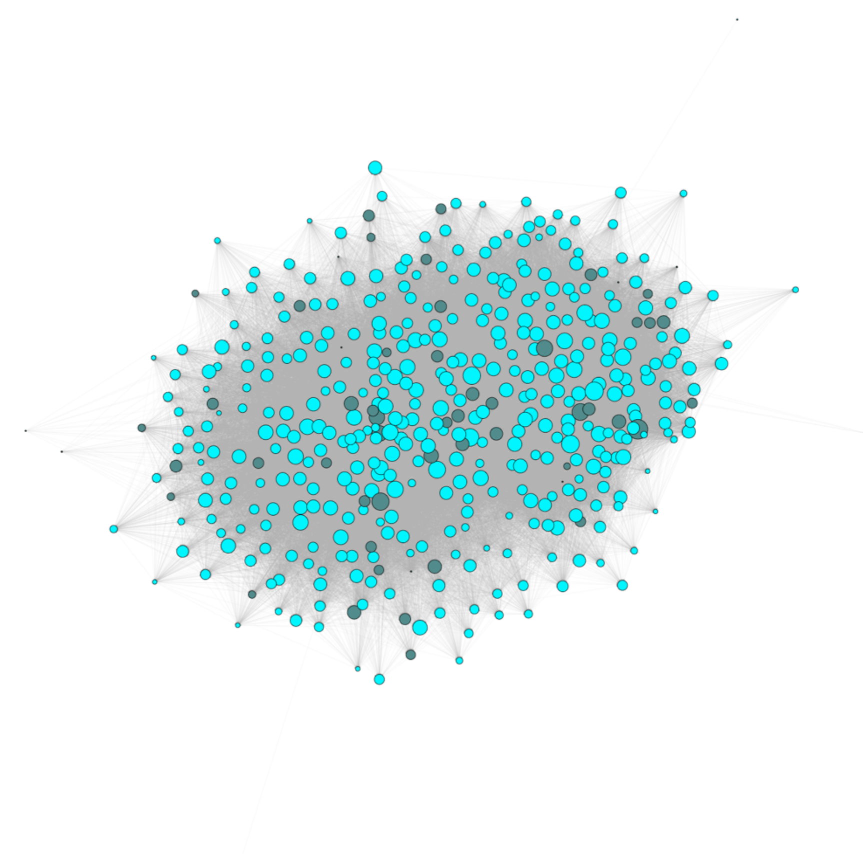

Figure 10

kept their seats. Those who did not are relatively evenly distributed in terms of size and placement.

The 2018 election graph (see figure 11 below) showed significantly more turnover, especially on the Republican side of the graph (the right side of the graph). This is consistent with the election results in which control of the House shifted from Republicans to Democrats. While it appears that Republicans who were more ideologically polar and also had smaller networks left office more than similarly polar Democrats, more analysis must be done to determine whether there is a true correlation.

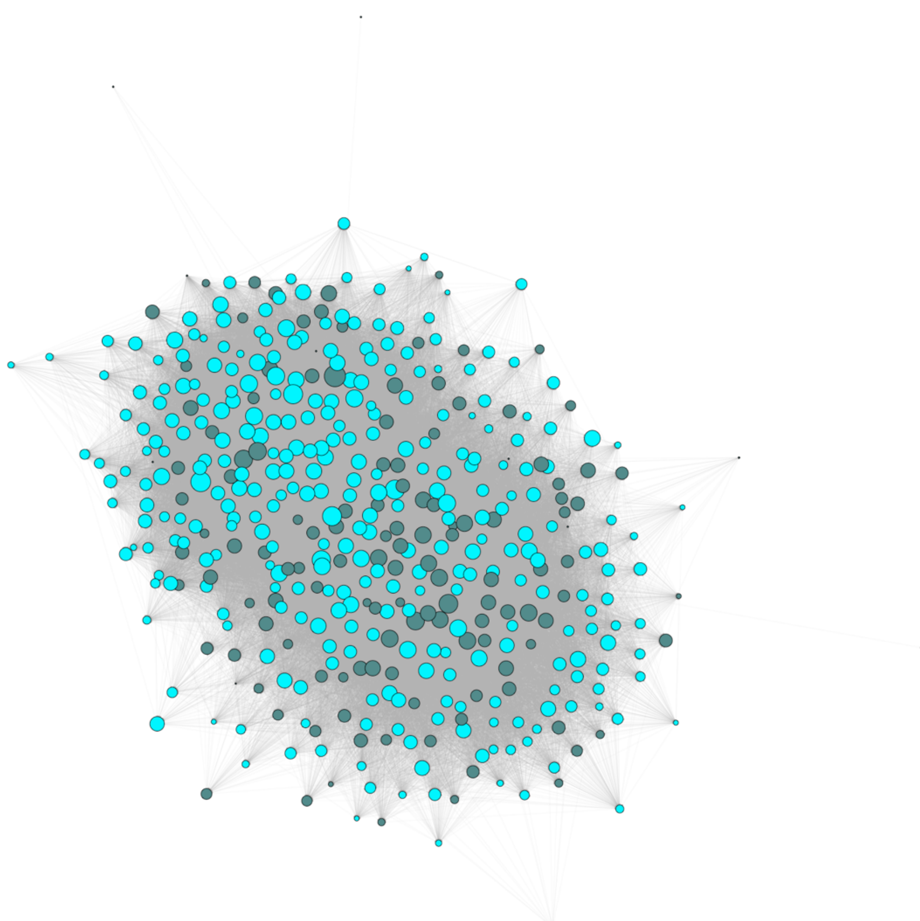

Figure 11

Conclusion

I was able to produce several different visualizations of the social networks in the 114th Congress. These illustrations did show a significant partisan divide that supports many of the news headlines about partisanship in Congress. There is very little crossover between parties, and there are a number of legislators who are relatively extreme in their ideologies and therefore are unlikely to work with members of the opposing party. While many of the legislators who have a high level of connectedness are located near the party line indicating at least some willingness to work across the aisle, not all well-connected legislators are as willing. Additionally, it does not appear that gender has any correlation to connectedness, indicating that more female candidates were elected in 2018 not because of their potential for developing networks within Congress, but due to other factors. Neither the size of a legislator's network nor their ideological extremity appears upon visual examination to have much impact on their electability. If a further correlation analysis were to support this visual analysis, it would confirm that either that the public does not view connectedness as a leading reason for reelection, or that they are not informed of how connected their legislators are.

In the future, this analysis could be expanded to such factors as age, length of term, geographic region, or committee membership to identify any other potential correlations. Furthermore, in the cases where a visual did not definitively rule out correlation, a statistical analysis should be done as a final determinant. Finally, it would be interesting to pair these results with a survey measuring the knowledge and attitudes of Americans regarding congressional connectedness.

References

Fowler, James H. 2006a. "Connecting the Congress: A Study of Cosponsorship Networks." *Political Analysis* 14, no. 4: 456–487.

Fowler, James. H. 2006b. "Legislative Cosponsorship Networks in the U.S. House and Senate." *Social Networks* 28, no. 4: 454-465.

Fowler, James H., Waugh, Andrew Scott, and Sohn, Yunkyu. 2017. "Cosponsorship Network Data." http://jhfowler.ucsd.edu/cosponsorship.htm (December 12, 2018).

GovTrack. 2017. "2016 Report Cards." https://www.govtrack.us/congress/members/reportcards/2016 (December 12, 2018).

Harward, Brian M., and Kenneth W Moffett. 2010. "The Calculus of Cosponsorship in the U.S Senate." *Legislative Studies Quarterly* 35, no. 1: 117–143.

Ladewig, Jeffrey W. 2010. "Ideological Polarization and the Vanishing of Marginals: Retrospective Roll-Call Voting in the U.S. Congress." *The Journal of Politics* 72, no. 2: 499–512.

Ognyanova, Katherine. 2016. "Network analysis with R and igraph: NetSci X Tutorial." Presented at the NetSciX 2016 School of Code Workshop in Wroclaw, Poland. Retrieved from www.kateto.net/networks-r-igraph.

Software and Programmer Efficiency Research Group. 2018. "ggplot2 Quick Reference: colour (and fill)." http://sape.inf.usi.ch/quick-reference/ggplot2/colour (December 12, 2018).

Tilley, James, and Sara B Hobolt. 2011. "Is the Government to Blame? An Experimental Test of How Partisanship Shapes Perceptions of Performance and Responsibility." *The Journal of Politics* 73, no. 2: 316–330.