**Network Analysis Methodology and Conclusions**

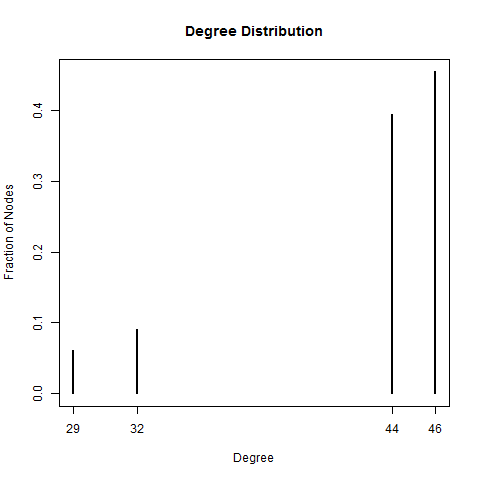
It is clear how the most straight forward implementation of a network to our data, or maybe the most obvious, might be bi-modal. Such a network would have nodes belonging to both candidates and industries. In particular, the visualization of such a network might make clusters of candidates much clearer. However, it is also clear that multimodal networks, even if the network under consideration is just a bimodal network, quickly become difficult to interpret and manage.[[1]](#footnote-1) With this in mind, and in accordance with the assignment, we decided to consider the network of candidates and industries, but converted to the unimodal domain.

In order to make a suitably small matrix, we used a subset of our data to build it. The subset was determined by taking all of the winning candidates for senate seats in 2014, and their two largest sponsors. What we were hoping to see was something interesting about the primary industries that supported those candidates. Perhaps the Republicans and Democrats would be in two clusters, or perhaps different industries focused on different candidates by region of the United States. What we ended up finding was that it looks as though industries hedge their bets. Using the top two industries, as we did here, we connect all of the winning senators a remarkable amount. There is a strong possibility that this is because of the binning strategy applied to the industries. However, an analysis of the results follow below.

The resulting dataset after the subsetting above results contains 33 senators. These senators each have two rows with industries that could connect them directly with other candidates. Keep in mind the following summary statistics for the senators’ dataset:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Industry** | Consumer Staples | Energy | Financials | Not for profit | Not publicly traded |
| **Count** | 1 | 1 | 18 | 30 | 16 |

After converting our network using a strategy by Solomon Messing found in his blog, we applied the example from Blackboard and implemented our network using the R igraph package.[[2]](#footnote-2) Since the original dataset had 33 senators, we end up with a network with 33 nodes. Amazingly, with just the top two industries per candidate, the candidates can connect to one another with a mean degree of 42.9 (we also tried this statistic after removing multiple edges, finding a mean degree of 31.6). Of course, the only way to get more than 32 degrees is to connect to a candidate through both of the top two industries. Only 5 of the 33 candidate failed to connect to every other candidate. Please find a degree distribution for the network with multiple edges, below.

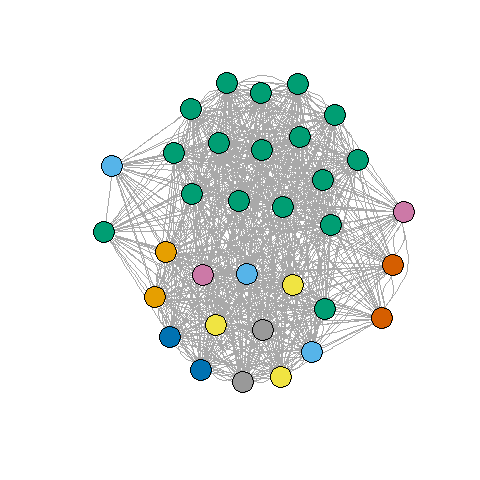


Unsurprisingly, this was also born out in the betweenness statistic, so it didn’t provide any additional value. The 5 that didn’t connect to every other candidate had betweenness scores of 0 since the other candidates don’t need to pass through any nodes to get to everyone. Every other candidate had a betweenness score of 0.2142857 due to the 5 being able to cross through them 1 time to get to any that they can’t get to otherwise.

The mean clustering coefficient was 0.56. For the 5 less connected candidates, in two cases there was a coefficient of 1, and for the other 3 it was close to 1 at 0.88. All of the other were around 0.5. This shows that 2 of the 5 less connected candidates were the only winning candidates for one of their sponsoring industries (of those where that industry was one of the top two sponsors). That would be the only way to get a coefficient equal to 1. The other three, however, share whatever industry it is that they have in common, bringing down the coefficient.

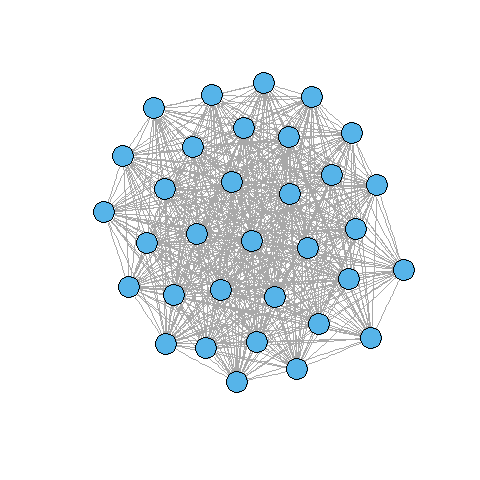
Given the above statistics, we already have a pretty clear idea of the remaining statistics. The graph density is greater than 1 due to the multiple edges at 1.34. The graph diameter is equivalent to 2, which indicates that even the 5 candidates that stand out have a first connection the same network. The number of connected components is equivalent to 1 (If just one industry is picked for each Senator – essentially turning the network into clustering – there are four components). Finally the largest k-core was 39. These are all indicators of just how many of the senators share both top supporters.

Please find the edge-betweenness network, below.



As you can see, and as we guessed based on the network statistics, the network is very intertwined. We did take a look at and tried to analyze the green cluster in particular, and this appears to be associated primarily with candidates that had both “Not for Profit” and “Financials” as their supporting industries. It is not clear why this cluster was able to snag the two dots not in its prime area. It is not clear why the bottom cluster was not also colored the same, as it must represent the combination of “Not for Profit” and “Not Publicly Traded”. The five that are more separated from the rest are clearly our 5 disparate points, with the two loners on the left holding “Consumer Staples” and “Energy” as one of their two top supporters, respectively, and the three on the right sharing “Financials” and “Not Publicly “Traded” in common.

Please find the modularity network, below. For this network, multiple edges were removed since that was a requirement of the greedy algorithm to run on this network.



Not surprisingly, the greedy algorithm determined that all of the nodes were part of the same community. Once multiple edges are removed it is very hard to see any significant pattern in the visualization of the network.

1. Source: Scott Weingard, “Networks Demystified 9: Bimodal Networks”, http://www.scottbot.net/HIAL/?p=41158 [↑](#footnote-ref-1)
2. Source: blog, https://solomonmessing.wordpress.com/2012/09/30/working-with-bipartiteaffiliation-network-data-in-r/ [↑](#footnote-ref-2)