Influence, Social Structure and Cross-Party Politics

A Social Network Analysis on the Inner Workings of the 110th US Senate

Tony Hung
hunw00@vse.cz

Table of Contents

Introduction

The 110th United States Congress lasted from 3 January 2007 until 3 January 2009. During this period, it was also the last two years of George W Bush's presidency. In the 110th Congress, the Democratic party had majorities of both the house and the senate. Vice President Dick Cheney (R)[1] served as the Senate President and Robert Byrd (D) served as the Senate President pro tem.

The US Congress is comprised of two parts: the House of Representatives and the House of Senates. The number of representatives each state sends to the congress is dependent on the population of the state. California, being the most populous state, sends 53 representatives, while Wyoming only sends one representative. However, each state elects two senators. Senators in the US serve six year term and there are no limits as to how many terms the senator could serve. For this project, we will solely be looking at the senate branch of the congress (Walberg, 2017).

The goal of this project is to do a social network analysis on the 110th US Senate. With basic information of each individual senator and the list of bills the senators have co-sponsored, we hope to find some underlying relationships that are not publicly observed. In this paper, each senator will be our nodes, and the number of bills that they have co-sponsored are the edges that ties between them.

There are two data files used in this analysis. The first file (110_billspon.csv) contains all of the bill sponsorship by each senator. The rows are the bills, and the columns are the senators. If there is a value in the cell, it means that the senator co-sponsored the bill. The second file (110_sen.csv) contains all of the basic details of the 102 senators such as political party, gender, religion, education, etc. For the sake of clarity, both Independent Senators, Bernie Sanders and Joe Lieberman, are classified as Democrats. This is because both senators have caucused with the Democratic Party. Both files are obtained from Dr. Eleanor Power, from the London School of Economics and Political Science.

In this project, we hope to find some underlying relationships between the senators through their co-sponsorships of bills. We postulate that senators will only co-sponsor bills of their interest, whether it is personal, political, or public. With this, we can discuss and answer the following questions of interest:

1. Who was the most influential senator?
2. Does party membership influence bill co-sponsorship?
3. What are the social structures of the senate?
4. Which characteristics of the senators predict co-sponsorship?

These questions will guide us along the project and the project will contain the following sections: basic data and graph descriptions, our four main questions and a conclusion.

[1] For clarity of this project, if a (R) appears after a name, it signifies that the aforementioned senator is part of the Republican Party. If a (D) appears after a name, it signifies that the aforementioned senator is part of the Democratic Party.

Basic Data and Graph Descriptions

All of the basic information of each senator is included in the 110_sen.csv file. In this dataset, there are 102 observations (or 102 senators) and 11 variables of the following:

- *Name*: signifies the official name of each senator
- *Party*: denotes which political party the senator belongs to
- *Gender:* indicates whether the senator is a male or a female
- *Religion:* shows which religion the senator identifies with
- *Class:* indicates when their seat in the senate will be free for election again[2]
- *State:* signifies which state the senator represents in the senate[3]
- *Census Region:* shows which census region the senator represents according to the United States Census Bureau[4]
- *Prior Experience:* shows what jobs the senator held before entering the US Congress
- *Education:* indicates the highest level of education
- *First Took Office:* denotes the year the senator first took office
- *Born:* shows the year when the senator was born

At first glance of this dataset, we see that both the Democratic Party and the Republican Party has equal membership of 51 senators, but please keep in mind that two of the Republican senators were replaced (see footnote 3), thus effectively, the Republicans only had 49 senators. There are 16 females and 86 males in the 110th senate. 25 senators identify with the Roman Catholic faith, leading the majority. 13 senators identified as Jewish, and 13 more senators identified as Methodist. There are 6 senators who were state attorney generals before entering the senate. The longest serving senator in this meeting is Robert Byrd (D), who started in 1959. Coincidentally, the oldest senator in this meeting is also Robert Byrd (D), who was born in 1917. The youngest senator in this meeting is John Sununu (R) who was born in 1964. The average age of the senators in 2007 was 63 years old.

The 110_billspon.csv file contains all of the bills each senator sponsored. The rows in the dataset represents the bills and the columns represents the senators in alphabetical order. In the 110th meeting of the Senate, there were 10327 bills sponsored.

For the network of the senators, we will 'split' the data in two networks for a better analysis. The first network shall remain unchanged and will be called the full network. The second network is a split from the full network, where only the senators who co-sponsored the top 75% of the bills will be extracted. This network will be called the simplified network. By extracting the top 75%, we are essentially only keeping the strong edges between the senators.

---

[2] In this meeting of the senate, there are three classes. The first class is 2009. The second class is 2011. The third class is 2013. These classes do not represent any hierarchy or social standing. It merely represents when the seat will be free for election again. For example, John McCain (R) was a class 2 senator and this means that his seat in the senate will be free in 2011.

[3] In the 110th US Senate, there are two senators who was replaced. Craig Thomas (R) passed away in April 2007 and was replaced by John Barrasso (R). Trent Lott (R) resigned in December 2007 and was replaced by Roger Wicker (R). For our analysis, all four senators will be included.

[4] There are currently 9 division of census regions in the United States. More information here.

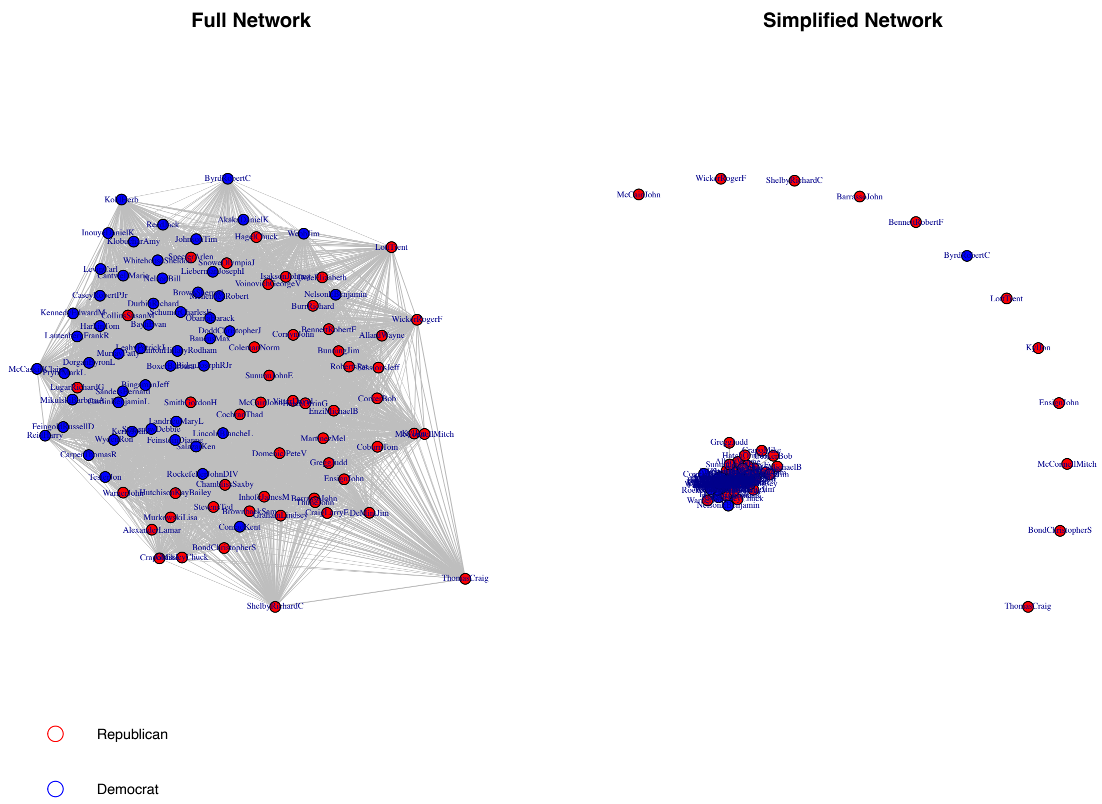


Figure 1: Plot of the full network and the simplified network

From the two graphs above, the red vertices mean that the senator is from the Republican Party and the blue vertices mean that the senator is from the Democratic Party. By looking at the full network, we can see that the senators all seem very well connected to each other. This means that all senators have co-sponsored with one another at least once, and no one is left out of the network (meaning no isolates).

On the other hand, looking at the simplified network gives us a different picture. The simplified network represents the top quartile of the full network. With this differentiation, we immediately see a few isolate who did not make this inclusion. The isolates in the simplified network are: Richard Shelby (R), Roger Wicker (R), Craig Thomas (R), John McCain (R), Trent Lott (R), Mitch McConnell (R), Jon Kyl (R), John Ensign (R), John Barrasso (R), Robert Bennett (R), Chris Bond (R), and Robert Byrd (D).

Table 1: Table with measures of number of edges, number of nodes, graph density, average path length, transitivity, and diameter for the full and simplified network

	Number of Edges	Number of Nodes	Density	Average Path Length	Transitivity	Diameter
Full	5151	102	1.0000000	1.000000	1.0000000	74
Simplified	1305	102	0.2533489	1.859675	0.7622924	507

From the full network, there are 5151 edges. The density, average path length, and transitivity in the full network are all 1. The density represents the percentage of actual edges to all possible edges (Newman, 2018). This means that the full network has exhausted all possible edges in the network. The average path length tells us the mean number of steps from one node to another (Newman, 2018). On average, the number of steps from senator to another senator is only 1. Transitivity indicates how likely two nodes are connected to a third node, thus making a triangular relationship (triad) (Newman, 2018). In this case, there is a 100% chance that two senators are themselves connected with a third senator, which means that the network is completely composed of triads. The diameter represents the size of the network (Newman, 2018). In the full network, the diameter is 74. This means that the longest path in the network takes 74 steps. With all of these statistics in mind, it reveals that information in the full network travels extremely fast and it also means that each senator must have co-sponsored at least with another senator once. This is not exactly surprising given small population of senators and a high amount of legislation proposed.

On the contrary if we look at the simplified network, we see that the number of edges decreased to 1305. The density is 0.253, which means that only 25.3% of the network is utilized. The average path length increased to 1.859. This means that if one senator would like to communicate to another senator, it would take, on average, 1.859 steps to get another senator. The transitivity is 0.762 in the simplified network. This means that 76.2% of the senators form a triad. The higher the transitivity, the more likely the senators will co-sponsor bills with the same group of senators. Given the low density, but the high levels of transitivity and short average path length, the simplified network is an example of the Small-world network (Barrat & Martin, 2000).

Question 1: Who was the most influential senator?

What does the most influential senator mean in our scenario? Does it mean the senator with the most ties? Or does it mean the senator with the most unique ties? Or perhaps the senator who can bridge the gap between different groups? All of these different questions could be answered by different centrality measures. Degree centrality measures the number of edges each node has. Betweenness centrality measures the number of unique edges each node has. Eigenvector centrality is similar to degree centrality, but it takes into account of all relative importance in the proximity. Closeness centrality measures how far a node is from other nodes on average. Drawing upon a very similar study by Fowler (2006), these centrality measures will shed light as to which senator is indeed the most influential in the 110th Senate.

For this question, we will solely be looking at the simplified network. This is due to the completeness of the full network where every node is connected to every other node, thus calculating the centrality measures are redundant[5].

[5] Because everyone is connected to everyone, thus the degree centrality for each node will be 101. The closeness centrality is the same for all nodes. The betweenness centrality is zero. Therefore, there is no actual significance to compute the centrality measures.

Table 2: Table describing the senator who has the most in each measure

	Degree Centrality	Eigenvector Centrality	Closeness Centrality	Betweenness Centrality
Simplified Network	69 Norm Coleman	1 Hillary Clinton	7.502e-04 Norm Coleman	771.229 Norm Coleman

From the scores, we see that Norm Coleman (R) is indeed the most influential senator in the simplified network. The degree centrality for Norm Coleman is 69. This means that Coleman has in total 69 co-sponsorships. The closeness centrality for Coleman is near zero. This means that Coleman is very close to all other senators. The closeness centrality could also be thought as the shortest (geodesic) paths between nodes. This means Norm Coleman is the node that lies on the majority of the shortest paths in this network. The betweenness measure for Coleman is 771.229. The betweenness centrality captures the idea that a person is a bridge to other people. In reality, Coleman, was once a Democrat and then later turned to Republican. This unique characteristic made Coleman the broker in the senate, who can overcome structural holes and bridge between different groups (Burt, 2004).

What about Hillary Clinton (D)? The eigenvector centrality gave a perfect score of 1 for Clinton. This means that she is the senator with the most unique edges. She has the most influential ties which made her stand out. However, the degree centrality for Clinton is 61 and the betweenness centrality is 102.267. Both measures are lower than Coleman's, especially the betweenness centrality. This suggests that the pitfall of Clinton's influence in the senate is her inability to broker between social groups.

Question 2: Does party membership influence bill co-sponsorship?

To answer this question, we will first calculate assortativity based on political membership. Assortativity is the tendency where nodes of the same characteristics will tend to group together. Then, we will perform community detection for the two networks. The approach will be slightly different for the two networks due to the completeness of the full network. For the full network, we will perform the Louvain Community Detection, created by Blondel et al. from the University of Louvain (Blondel, et al., 2008). For the simplified network, we will perform stochastic blockmodel to detect if party membership do influence bill co-sponsorship. Each of these community detection methods rely on different theoretical approaches. Louvain Community Detection computes by simulating the nodes in different groups until optimization is reached (Blondel, et al., 2008). The stochastic blockmodel approach is based on calculating the probability of an edge between nodes, and thus determine which nodes belongs to which group (Doreian, et al., 2005).

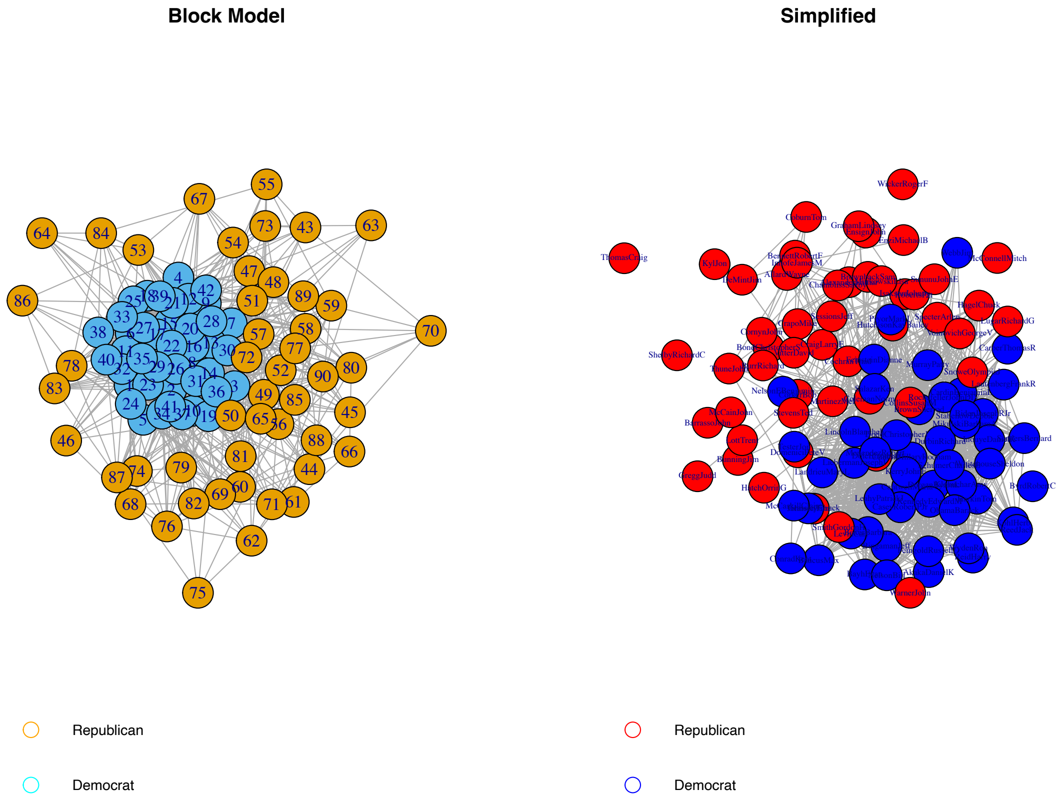


Figure 2: Comparison between the blockmodel approach and the simplified network plots

Starting with the simplified network. From the graph above (figure 2) we can see that aesthetically, the two graphs look dissimilar. We must keep in mind that the blockmodel approach is a model based on the simplified network, according to party affiliation. Upon closer inspection, we can see that almost all of the nodes in the blockmodel are connected to each other, while in the simplified network, there are noticeable isolates. In the blockmodel, there is a bit of misspecification on party membership. The orange nodes should be Republicans, and the light blue nodes should be democrats, but this is different when we compare it to the simplified model. It is also worth pointing out that the blockmodel graphed the Democrats as a tightly knitted circle in a crowd of Republicans.

The assortativity of the simplified network is 0.336. This means that in the simplified network, 33.456% of the people who are in the same political party will tend to assort together. The assortativity for the blockmodel is 0.339. This means that 33.909% of the senators who are in the same political party will tend to assort together. These two statistics are very similar to each other, which means that the blockmodel could maybe be a good proxy for the simplified network.

Table 3: Probability of a co-sponsorship link using the blockmodel approach

	Democratic	Republican
Democratic	0.6533333	0.1218762
Republican	0.1218762	0.1215686

From table 3, the blockmodel suggests that a Democratic senator has a chance of 65.333% to co-sponsor with another Democratic senator. A Democratic senator has a 12.188% chance of co-sponsoring with a Republican senator and vice versa. A Republican senator has a 12.157% chance of co-sponsoring with another Republican senator. This suggests that the Democrats has a high homophily and the Republicans has neither a high a homophily nor a high heterophily. This suggests the possibility of other attributes that could explain the low homophily and heterophily of Republican senators.

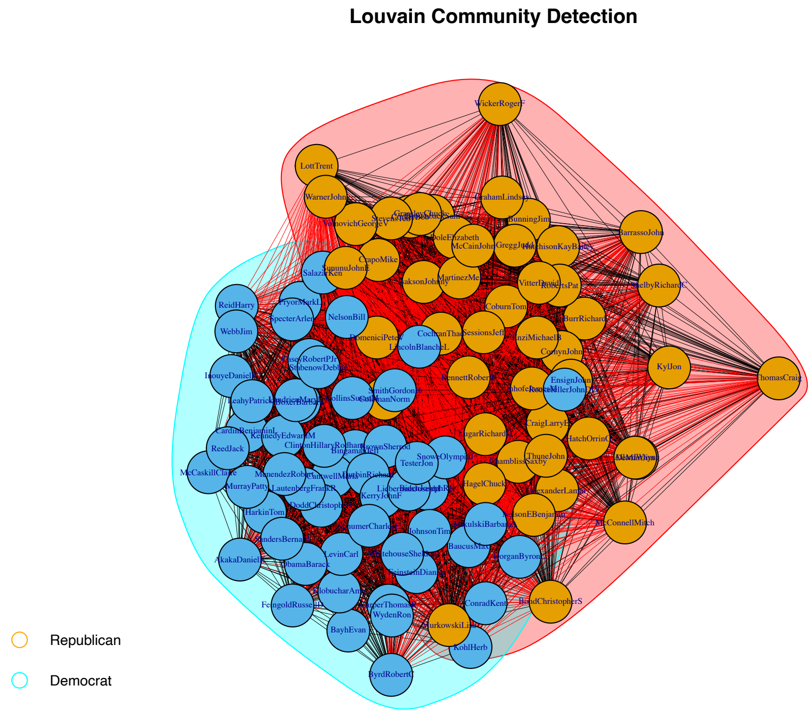


Figure 3: Plot of the Louvain Community Detection algorithm

From Louvain model (figure 3), we see that the algorithm identifies two groups in the full network. The red group mainly consists of Republican senators and the blue group mainly consists of Democratic senators. There is a bit of overlap between the groups.

Table 4: Comparison between results of party membership of the Louvain Community Detection

	Community 1	Community 2
Democratic	1	50
Republican	47	4

The Louvain algorithm correctly identified 50 Democrats, but mismatched 1 senator. On the other hand, the algorithm correctly identified 47 Republicans, but mismatched 4 senators.

To see what other attributes could closely align to the Louvain communities, all other attributes have been compared to see what other characteristic could possible influence co-sponsorships.

Table 5: Table detailing other attributes in comparison to the Louvain Community Detection algorithm

Attributes	Comparison Scores against Louvain Algorithm
Party	0.73048434
Class	0.02997089
Census Region	0.13010106
Gender	0.05722225
Religion	0.14404113
First Took Office	0.07656904
Born	0.10265503

According to this table, the party attribute has a 73.048% similarity with the Louvain model. Class attribute has a 2.997% similarity with the Louvain model. Census Region has a 13.010% similarity with the Louvain model. Gender has a 5.722% similarity with the Louvain model. Religion has a 14.404% similarity with the Louvain model. First took office has a 7.657% similarity with the Louvain model. Born has a 10.266% similarity with the Louvain model. From this, we can see that the party attribute is the most similar to the Louvain model. Class, gender, and first took office are rather different to the Louvain model.

In essence, we can say that party affiliation is a strong predictor for bill co-sponsorship. We see that the Democrats will more likely to co-sponsor each other and this could possibly because the Democrats at that time will use their combined powers to push for different objectives under the leadership of a Republican president. The Republican senator are harder to predict, but this suggests that there could be other attributes which could influence bill co-sponsorship.

Question 3: What are the social structures of the senate?

To examine the social structure, of the 110th Senate, a few operations needs to be performed. We would have to cross examine the results of the dendrogram and then impart these results on to the full network. These methods will solely be based on the full network, because we would like to examine the senate as a whole.

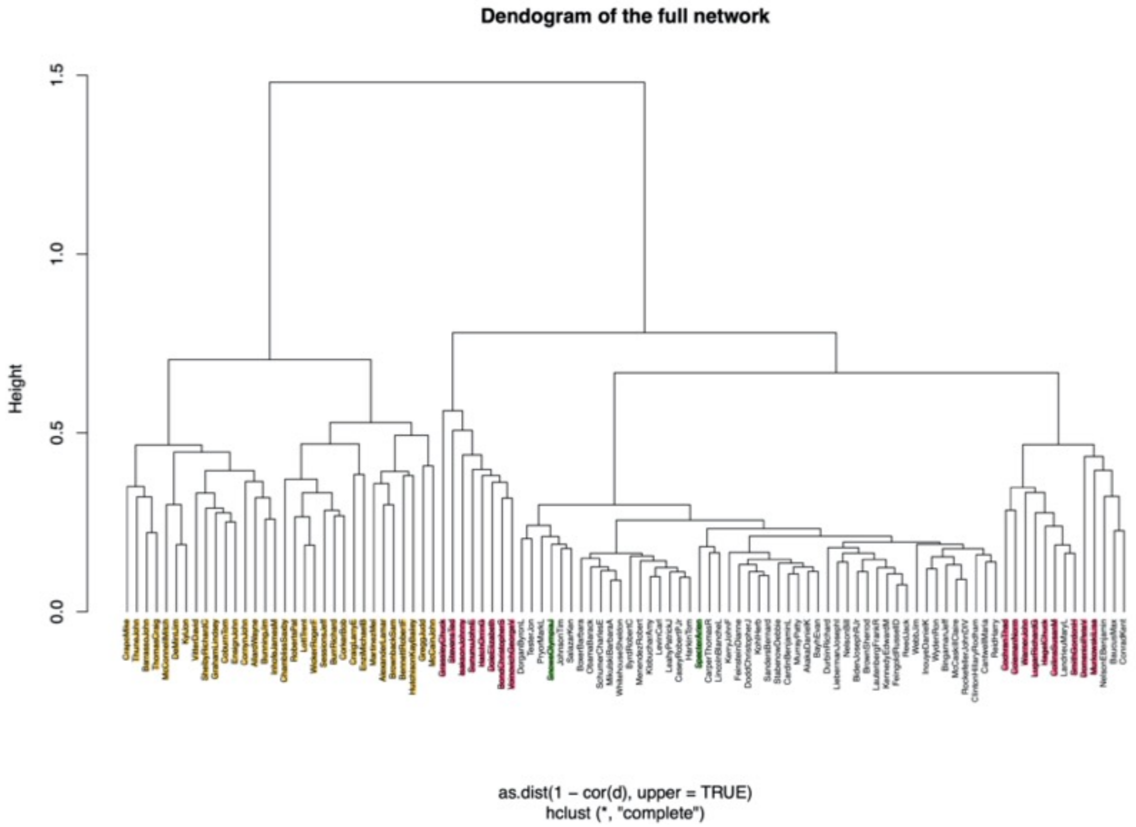


Figure 4: Dendrogram using the complete algorithm, representing the full network

From the figure above, we see that with the complete algorithm, the algorithm identified two main groups in the senate and then later split the network further into four subgroups. The complete algorithm is a type of agglomerative hierarchical clustering, where it maximizes the complete linkage of nodes (Everitt, et al., 2011). If we refer back to question 1, our most influential senator Norm Coleman (R), the most similar senator to him is Thad Cochran (R). Cochran's centralities are: degree 27, eigenvector 0.275, betweenness 3.025, closeness 7.236e-04. The only centrality that is similar to Coleman's is the closeness centrality. The height or the closeness of the clustering between Coleman and Cochran only occurred near 0.4. In comparison to the other clusters, the clustering between Coleman and Cochran is not immediate.

All of the unhighlighted senators in the dendrogram are Democrats and all of the highlighted senators are Republicans. Interestingly, Olympia Snowe and Arlen Specter, both Republicans, are classified with the clusters of Democrats. This could possibly be explained due to the fact that between 1965 to 2009 Arlen was a Republican, but all other years of his political activity, Arlen was a Democrat (CNN, 2009). Snowe was often voting in line with the Democrats, during this meeting of the senate, thus Snowe could be seen as a 'Democrat' in during this period (Weinger, 2012).

It is also very interesting to see that Mary Landrieu, Benjamin Nelson, Max Baucus and Kent Conrad are all Democrats placed in the same cluster with Republicans. This is because all four of these senators often vote in opposition against their own party and was considered as a conservative Democrat (Grim, 2010) (Wing, 2011) (Bouie, 2013) (Raju, 2011).

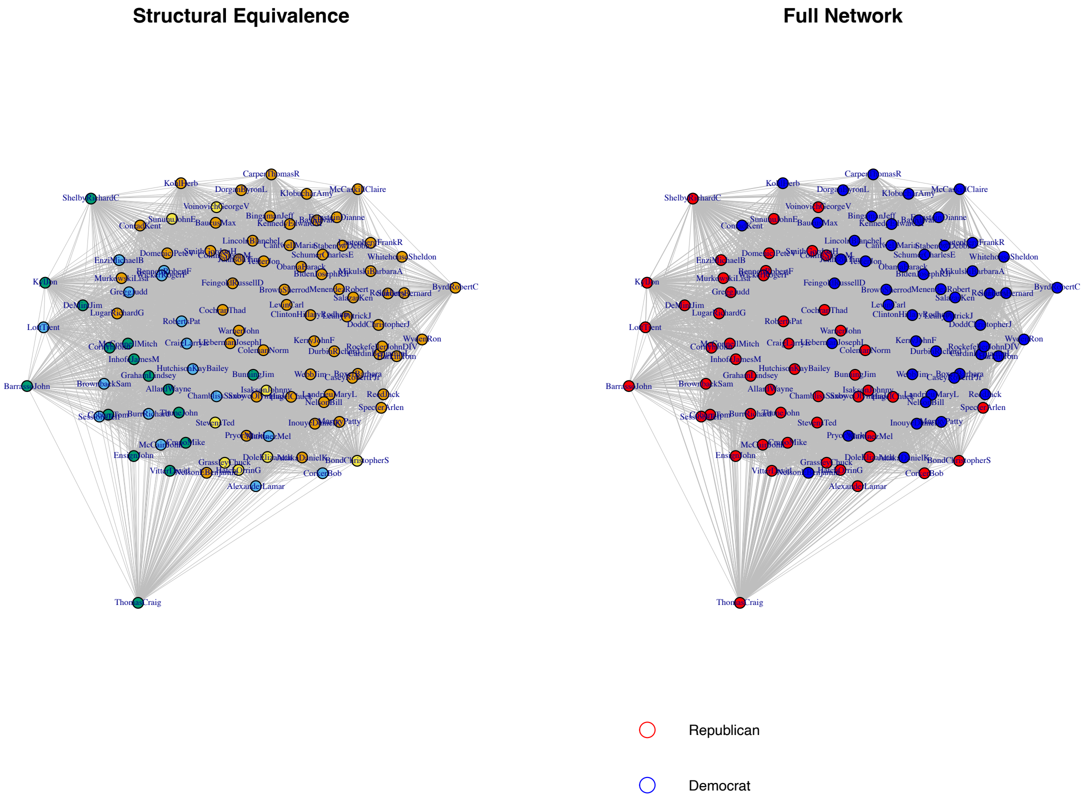


Figure 5: Comparison between structural equivalence and the full network

By using the results from the dendrogram, we can plot the results on to the full network and compare. We see that the orange colored nodes in the structural equivalence graph are generally Democrats. Coincidentally, some of the green colored nodes are the isolates identified from simplified network.

Table 6: Results of the structural equivalence classes by party membership

	Democratic Party	Republican Party
Class 1	51	11
Class 2	0	16
Class 3	0	16
Class 4	0	8

From the Table above, we see that class 1 consists of all 51 Democratic senators and the Republican senators are more spread out to the all four classes. This means that the algorithm could correctly identify the Democrats as a group and we saw this graphically in figure 2 where the Democrats were a cluster inside a cloud of Republicans. This could further mean that the Republican party has different cliques within their party due to bill co-sponsorship history, and perhaps, the Republican senators have a less homophily and assortativity.

From the Louvain model, the other attributes that could be representative of the senate are Census Region and Religion. It is unclear if geography influences politics. Glaeser and Ward (2006) argues that even though in the United States, it is common to label each state as either red, blue, or swing state, it does not represent the total population. For this project, we will continue to analyze with census region as a factor. Glaeser and Ward also mentioned that religious politics have become common (2006). Studies have shown that religion influences politics in the United States (Wald & Calhoun-Brown, 2014), but the inclusion of religion as a factor on the co-sponsorship in the senate is beyond the scope of this project.[6]

Table 7: Table showing the frequency of each census region and structural class by political party

	Census Region 1		Census Region 2		Census Region 3		Census Region 4		Census Region 5		Census Region 6		Census Region 7		Census Region 8		Census Region 9	
Class 1	8	2	5	1	8	1	7	2	8	1		1	3		5	1	7	2
Class 2		1						2		3		5		1		4		
Class 3								1		2		3		4		6		
Class 4		1				1		2		2						1		1

Note: the blue numbers denote number of Democratic senators and the red numbers denote number of Republican senators. Blank cells are zero. For more information regarding census regions in the United States, please refer to footnote 4.

From this table, we can see that the spread of the Republican senators across the various regions. A significant amount of Republican senators comes from region 8, 6, 5, and 4. These states are commonly referred to the 'South' in the US and tend to have more rural / suburban characteristics. It is important to note that census region 6 is solely comprised of Republican senators. The majority of Democratic senators comes from regions 1,3,4,5,2, and 9. Census region 4 and 5 have equal amounts of senators from both parties. This supports the argument from Glaeser and Ward (2006), where a simple label of color does not fully represent a state.

To conclude, there are four groups in the senate and these groups are classified this way possibly due to their history of co-sponsorships with other senators. We have discovered that there are Republican senators who are more in line with Democrats (Arlen Specter and Olympia Snowe) and Democratic senators who are more in line with Republicans (Mary Landrieu, Benjamin Nelson, Max Baucus and Kent Conrad). We have also seen that not all states are either blue or red, some are more purple (the mix of blue and red).

[6] Religion is a sensitive and complex subject. Without the complete knowledge of religious studies, it is rather difficult to analyze the religion attribute of the network. This is mainly due to too many religion categories thus for better efficiency, it is better to group similar faiths together. But without a complete understanding of each of the faith listed in the variable, it is impossible to categorize the faiths. I have thought that since the biggest religions in the senate are Roman Catholics and Jewish, so I thought maybe it could be Catholics=1, Jewish=2, Others=3, but this means that there might be important information lost in the process due to this overgeneralization.

Question 4: Which characteristics of the senators predict co-sponsorship?

Exponential Random Graph Model is a statistical model specifically for modelling networks. In this section, we will look at building a model which could predict the likelihood of a bill being co-sponsored by using some of the characteristics. In essence, we are building a logistic regression for networks. Due to the relational properties of networks, standard statistical models are not recommended. The relations between nodes violates one of the core properties of statistical modelling: independence, as outlined by Contractor et al. (2006). This section will be based on the simplified network, as the completeness of the full network will not result in any modelling.

```
==========================
Summary of model fit
==========================

Formula:   qnet ~ edges

Iterations:  5 out of 20

Monte Carlo MLE Results:
      Estimate Std. Error MCMC % z value Pr(>|z|)
edges -1.08083    0.03204      0  -33.74   <1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 7141  on 5151  degrees of freedom
 Residual Deviance: 5831  on 5150  degrees of freedom

AIC: 5833    BIC: 5839    (Smaller is better.)
```

Figure 6: Output of model 1

In the first model, I have modelled the relationship between the simplified network as a whole and the existing edges. To interpret the coefficient of the estimate, we would need to exponentiate. This means that in the simplified network, any two senators are 0.339 times as likely will co-sponsor together. We can see that this ERGM model required Monte Carlo simulations. The edges term is statistically significant at all levels. A way to think of the edges term is that it functions around the same way as a constant in linear regression.

```
========================
Summary of model fit
========================

Formula:   qnet ~ edges + nodefactor("party") + nodematch("party") + nodefactor("gender") +
    nodematch("gender") + nodematch("census") + absdiff("first") +
    gwesp(0.6, fixed = TRUE)

Iterations:  8 out of 20

Monte Carlo MLE Results:
                    Estimate Std. Error MCMC % z value Pr(>|z|)
edges              -6.915860   0.723122      0  -9.564   <1e-04 ***
nodefactor.party.2  -1.199017   0.033548      0 -35.740   <1e-04 ***
nodematch.party      1.861290   0.066948      0  27.802   <1e-04 ***
nodefactor.gender.2 -1.126577   0.133594      0  -8.433   <1e-04 ***
nodematch.gender     0.673041   0.151189      0   4.452   <1e-04 ***
nodematch.census     0.661768   0.107159      0   6.176   <1e-04 ***
absdiff.first       -0.006935   0.003689      0  -1.880   0.0601 .
gwesp.fixed.0.6      3.810799   0.372957      0  10.218   <1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 7141  on 5151  degrees of freedom
 Residual Deviance: 4041  on 5143  degrees of freedom

AIC: 4057    BIC: 4109    (Smaller is better.)
```

Figure 7: Output of model 2

In model 2, I have included several other variables, such as party, gender, census and first. All terms are statistically significant besides absdiff.joined, which is a term on the absolute difference when the senators first took office. In the following paragraphs, I will be including the variable names in brackets to avoid confusion.

According to this model, a republican senator (nodefactor.party.2) is 0.301 times as likely to have ties with other senators. A male senator (nodefactor.gender.2) is 0.324 times as likely to have ties with other senators. This means that a republican senator tends to have around 70% less ties with other senators and a male senator tends to have 67.6% less ties with other senators

If two senators are in the same party, they are 6.432 times as likely to have ties, than in different parties (nodematch.party). If two senators have the same gender, they are 1.960 times as likely to have ties, than different genders (nodematch.gender). If two senators come from the same census region, they are 1.938 times as likely to have ties(nodematch.census). If two senators took office further apart, they are 0.993 times as likely to have ties (absdiff.first). This means that as the years of two senators’ took office increases by one, then those two senators are 0.993 times as likely to have ties with every additional year of difference.

This generally means that there is a high homophily between senators, where senators tend to assort between similar attributes. This means that, generally speaking, if two senators are from the same political party or same gender or the same region, then they are likely to have a tie exist between them.

Suppose that alpha=0.6 gives us the best model (gwesp.fixed.0.6), then this means that we are not discounting as much with subsequent triangles formed (transitivity). Thus the term gwesp.fixed.0.6 means that two senators who already have a tie between them are 45.187 times more likely to also know a third senator whom they both have a tie with. This means that there is a high transitivity in this model, which corresponds to the transitivity calculated in table 1.

Suppose that there are two women from the same region, who joined in the same year, and who have both cosponsored a bill with one other senator, once, and they are both Democrats. The probability that these two women will form a tie with each other is 52.101%. Conversely, given the same conditions, but the two women are both in the Republican Party, then the probability of these two women forming a tie is 8.998%. The

Conclusion

The United States Senate is often a focal point in politics and to understand the inner workings of the senate, it often requires some extensive investigation. To hopefully uncover some of the true nature of the senate, this project used each senator as nodes, and the bills each senator has co-sponsored as edges. This project started with basic descriptions on the characteristics of the senators, for example age, party affiliation, gender, religious affiliation, class, when the senator first took office, etc. Because of the completeness of the full network, a second graph was extracted based on the top quartile of bill co-sponsorships. Through this, we have identified Norm Coleman (R) as the most influential senator based on centrality measures. We have also modelled both of the simplified and full networks and determined that the party membership characteristic influences the bill co-sponsorship through the Louvain model and the blockmodeling approach. We have also looked at the social structure of the 110th Senate through the dendrogram and evaluated which other characteristics are in line with the results. Finally, we have built an ERGM model which could help us predict bill co-sponsorships given certain characteristics.

However, there work with these two networks are unfinished. One idea for further analysis is to categorize the religions more effectively, and use it to examine the social structure of the senate. Another idea is to use different methods of clustering and then compare results. The ERGM could be improved with better variables and perhaps a more in-depth analysis on why the Republican senators are more spread out, compared to the Democrats.

Works Cited

Newman, M., 2018. *Networks.* 2nd Edition ed. s.l.:Oxford University Press.

Barrat, A. & Martin, W., 2000. On the properties of small-world network models.. *The European Physical Journal B-Condensed Matter and Complex Systems,* 13(3), pp. 547-560.

Fowler, J. H., 2006. Connecting the Congress: A study of cosponsorship networks. *Political Analysis,* 14(4), pp. 456-487.

Burt, R. S., 2004. Structural holes and good ideas. *American Jounal of Sociology,* 110(2), pp. 349-399.

Blondel, V. D., Jean-Loup, G., Renaud, L. & Etienne, L., 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment ,* Volume 10, p. P10008.

Doreian, P., Batagelj, V. & Anuska, F., 2005. *Generalized blockmodeling.* s.l.:Cambridge University Press.

Everitt, B. S., Landau, S., Leese, M. & Stahl, D., 2011. *Cluster Analysis.* 5th Edition ed. s.l.:Wiley.

CNN, 2009. *Longtime GOP Sen. Arlen Specter becomes Democrat.* [Online]
Available at:
https://edition.cnn.com/2009/POLITICS/04/28/specter.party.switch/index.html
[Accessed 04 04 2020].

Weinger, M., 2012. *9 Snowe votes that angered the GOP.* [Online]
Available at: https://www.politico.com/story/2012/02/9-snowe-votes-that-angered-the-gop-073449
[Accessed 04 04 2020].

Grim, R., 2010. *Mary Landrieu: 'Obama-McConnell Plan' Is 'Almost Morally Corrupt'.* [Online]
Available at: https://www.huffpost.com/entry/mary-landrieu-obamamcconn_n_793272
[Accessed 04 04 2020].

Wing, N., 2011. *Ben Nelson Is Senate Democrat Most Likely To Vote Against His Party: Analysis.* [Online]

Available at: https://www.huffpost.com/entry/ben-nelson-vote-democrats_n_804101
[Accessed 04 04 2020].

Bouie, J., 2013. *Max Baucus's mixed legacy.* [Online]
Available at: https://www.washingtonpost.com/blogs/plum-line/wp/2013/04/23/max-baucus-mixed-legacy/
[Accessed 04 04 2020].

Raju, M., 2011. *Conrad's last-chance budget deal.* [Online]
Available at: https://www.politico.com/story/2011/04/conrads-last-chance-budget-deal-053680
[Accessed 04 04 2020].

Walberg, T., 2017. *HOW CONGRESS WORKS.* [Online]
Available at: https://walberg.house.gov/about/how-congress-works
[Accessed 04 04 2020].

Glaeser, E. L. & Ward, B. A., 2006. Myths and Realities of American Political Geography. *JOURNAL OF ECONOMIC PERSPECTIVES,* 20(2), pp. 119-144.

Wald, K. D. & Calhoun-Brown, A., 2014. *Religion and Politics in the United States.* s.l.:Rowman & Littlefield.

Contractor, N. S., Wasserman, S. & Faust, K., 2006. Testing Multitheoretical, Multilevel Hypotheses About Organizational Networks: An Analytic Framework and Empirical Example. *Academy of Management Review,* 31(3), pp. 681-703.

Appendix A: R Codes Used in this Project

```
#### 110th US Senate Project #####
# Tony Hung

# Note: some figures and outputs might be different to the ones mentioned in the word
doc. This is because some of the figures or models are simulated, thus each simulation
would produce different results.

# Basic Set-up
#install.packages("intergraph")
#install.packages("igraph")
#install.packages("gplots")
#install.packages("sna")
#install.packages("netcluster")
#install.packages("ape")
#install.packages("ergm")
#install.packages("dplyr")
#install.packages("network")

require(intergraph)
require(igraph)
require(gplots)

sen <- read.csv("~/Desktop/Work/MY461/110_sen.csv", header=TRUE)
bills <- read.csv("~/Desktop/Work/MY461/110_billspon.csv", header=TRUE,row.names = 1)

#### Descriptive Statistics ####
summary(sen)

#### Full and Simplified Networks ####
bill <- as.matrix(bills)
bill1 <- graph.incidence(bill, weighted=TRUE, directed = FALSE)
bill2 = bipartite.projection(bill1)
bill_adj <- as_adjacency_matrix(bill2$proj2, attr = 'weight')
full <- graph.adjacency(bill_adj, mode = 'undirected', weighted = TRUE)
summary(full)
par(mfrow=c(1,1))
dlayout = layout.fruchterman.reingold(bill2$proj2)
V(full)$label.cex = 0.5
full <- set.vertex.attribute(full,"party",index = V(full), value = (sen$Party))
summary(full)
V(full)$color = ifelse(V(full)$party=="1","blue","red")

full <- set.vertex.attribute(full,"gender",index = V(full), value = (sen$Gender))
full <- set.vertex.attribute(full,"religion",index = V(full), value = (sen$Religion))
full <- set.vertex.attribute(full,"class",index = V(full), value = (sen$Class))
```

```
full <- set.vertex.attribute(full,"state",index = V(full), value = (sen$State))
full <- set.vertex.attribute(full,"census",index = V(full), value = (sen$CensusRegion))
full <- set.vertex.attribute(full,"prior",index = V(full), value = (sen$PriorExperience))
full <- set.vertex.attribute(full,"educ",index = V(full), value = (sen$Education))
full <- set.vertex.attribute(full,"first",index = V(full), value = (sen$FirstTookOffice))
full <- set.vertex.attribute(full,"born",index = V(full), value = (sen$Born))

summary(E(full)$weight) # After obtaining the summary of the edge weights, we can see that the cut-off for the 3rd quartile is 103
quarter <- subgraph.edges(full, E(full)[E(full)$weight>=103], del=FALSE)
V(quarter)$label.cex = 0.5

# Figure 1
par(mfrow=c(1,2))
plot(full,layout=dlayout,edge.width=E(full)/max(E(full)),vertex.size=5,edge.color="grey",main="Full Network")
legend("bottomleft", c("Republican","Democrat"), pch=21, col=c("red","blue"), pt.cex=2, cex=.8, bty="n", ncol=1)
plot(quarter,vertex.size=5,edge.color="grey",main="Simplified Network")
par(mfrow=c(1,1))

ecount(full)
vcount(full)

# Table 1: basic graph statistics
table1 <- data.frame(c('Full','Simplified'),
              c(ecount(full),ecount(quarter)),
              c(vcount(full),vcount(quarter)),
              c(graph.density(full), graph.density(quarter)),
              c(average.path.length(full), average.path.length(quarter)),
              c(transitivity(full), transitivity(quarter)),
              c(diameter(full, directed = TRUE),diameter(quarter, directed = TRUE))
              )
colnames(table1) <- c('Number of Edges', 'Number of Nodes','Network', 'Density', 'Average Path Length', 'Transitivity','Diameter')
table1

#### Question 1: Who was the most influential senator? ####

# Table 2: centrality measures
max(degree(quarter)) # Degree Centrality.Norm Coleman 69
max(evcent(quarter)$vector) # Eigenvector Centrality. Hillary Clinton 1
max(betweenness(quarter, weights = 1/(E(full)$weights))) # Norm Coleman 771.2286177
max(closeness(quarter, weights = 1/(E(full)$weights))) # Norm Coleman 7.501875e-04

#### Question 2: Does party membership influence bill co-sponsorship? ####

```

```
# Block Model (simplified) and then Louvain Model (full)

# Block Model for the simplified network
require(intergraph)
detach(package:igraph)
require(sna)

qnet<-asNetwork(quarter)
qbm = blockmodel(qnet,ec=qnet %v% "party",rlabels = c("Democratic",
"Republican"))$block.model

summary(qbm)

# fnet<-asNetwork(full)
# blockmodel(fnet,ec=qnet %v% "party",rlabels = c("Democratic", "Republican"))


#Going back to igraph
detach(package:sna)
detach(package:intergraph)
require(igraph)

qkar <- sample_sbm(90, pref.matrix=qbm, block.sizes=c(42,48), directed=FALSE)

# Figure 2: blockmodel vs simplified
par(mfrow=c(1,2))
plot(qkar,vertex.color=c(rep(2,42),rep(1,48)), main='Block Model')
legend("bottomleft", c("Republican","Democrat"), pch=21, col=c("orange","cyan"), pt.cex=2,
cex=.8, bty="n", ncol=1)
plot(quarter, main='Simplified',layout=dlayout)
legend("bottomleft", c("Republican","Democrat"), pch=21, col=c("red","blue"), pt.cex=2,
cex=.8, bty="n", ncol=1)

dev.off()

assortativity(quarter,factor(V(quarter)$party))
# 0.3345591
assortativity(qkar,c(rep(1,42),rep(2,48)))
# 0.3390892

# Figure 3: Louvain Community Detection: Full Network
ml <- cluster_louvain(full, weights = E(full)$weight)
plot(ml, full, node.size=0.01, main='Louvain Community Detection',edge.width =
1/E(full)$weight, vertex.label.cex = .5)
legend("bottomleft", c("Republican","Democrat"), pch=21, col=c("orange","cyan"), pt.cex=2,
cex=.8, bty="n", ncol=1)
```

```
## Note: This section below is not included in the word document, as it is more of my own
'amusement'
# If we use Louvain on the simplified network, it will detect one big chunk, and each isloates
individually, thus defeating the purpose... Unless we remove the isloates to do Louvain?

qq <- subgraph.edges(full, E(full)[E(full)$weight>=103], del=TRUE)
V(qq)$label.cex = 0.5
qml <- cluster_louvain(qq, weights = E(qq)$weight)
plot(qml, qq, node.size=0.01, main='Louvain Community Detection',edge.width =
1/E(qq)$weight, vertex.label.cex = .5,layout=dlayout)
# Very interesting. There are technically three groups, but the overlaps are way too similar...
Baucus is an isolate, meaning that he did not get included in any group.
## Note: This section above is not included in the word document, as it is more of my own
'amusement'

# Table 4
table(V(full)$party,membership(ml))

# Comparing the Louvain against other attributes

# full network based on other attributes
fullc <- graph.adjacency(bill_adj, mode = 'undirected', weighted = TRUE)
fullc <- set.vertex.attribute(fullc,"class",index = V(fullc), value = (sen$Class))
fullcr <- graph.adjacency(bill_adj, mode = 'undirected', weighted = TRUE)
fullcr <- set.vertex.attribute(fullcr,"cr",index = V(fullcr), value = (sen$CensusRegion))
fullg <- graph.adjacency(bill_adj, mode = 'undirected', weighted = TRUE)
fullg <- set.vertex.attribute(fullg,"g",index = V(fullg), value = (sen$Gender))
fullr <- graph.adjacency(bill_adj, mode = 'undirected', weighted = TRUE)
fullr <- set.vertex.attribute(fullr,"r",index = V(fullr), value = (sen$Religion))
fullf <- graph.adjacency(bill_adj, mode = 'undirected', weighted = TRUE)
fullf <- set.vertex.attribute(fullf,"f",index = V(fullf), value = (sen$FirstTookOffice))
fullb <- graph.adjacency(bill_adj, mode = 'undirected', weighted = TRUE)
fullb <- set.vertex.attribute(fullb,"b",index = V(fullb), value = (sen$Born))

# Table 5
table3 <- data.frame(c('Party','Class','Census Region','Gender','Religion','First Took
Office','Born'),
                     c(compare(V(full)$party,ml,method="nmi"),
                       compare(V(fullc)$class,ml,method="nmi"),
                       compare(V(fullcr)$cr,ml,method="nmi"),
                       compare(V(fullg)$g,ml,method="nmi"),
                       compare(V(fullr)$r,ml,method="nmi"),
                       compare(V(fullf)$f,ml,method="nmi"),
                       compare(V(fullb)$b,ml,method="nmi")))
colnames(table3) <- c('Attributes','Comparison Scores against Louvain')
table3
```

```
#### Question 3: What are the social structures of the senate? ####
# Week 4 seminar

require(NetCluster)
require(gplots)
require(ape)
require(sna)
require(network)

d = as.matrix(bill_adj)

# Figure 4: Dendrogram
cor(d) # Another way to look at
as.dist(1-cor(d),upper=TRUE)
comp<-hclust(as.dist(1-cor(d),upper=TRUE),method="complete")
plot(comp, cex=0.5, main='Dendogram of the full network',hang=-1)

# Figure 5: Structural Equivalence
#cutting it into four
dev.off()
par(mfrow=c(1,2))
plot(bill2$proj2,edge.width=1/E(bill2$proj2)$weight,vertex.color=cutree(comp, k=4),
vertex.label.cex = .5, main='Structural
Equivalence',layout=dlayout,vertex.size=5,edge.color="grey")
plot(full,layout=dlayout,edge.width=E(full)/max(E(full)),vertex.size=5,edge.color="grey",mai
n="Full Network")
legend("bottomleft", c("Republican","Democrat"), pch=21, col=c("red","blue"), pt.cex=2,
cex=.8, bty="n", ncol=1)

dev.off()

table_class_equivalency<-data.frame(ClassID=cutree(comp, k=4))
row.names(table_class_equivalency)<-sen$Name

eq_class = table_class_equivalency$ClassID
sen=cbind(sen,eq_class)

require(dplyr)
# Table 6
sen %>% group_by(eq_class,Party) %>% tally()
# sen %>% group_by(eq_class,Gender) %>% tally()

# Table 7
q =  sen %>% group_by(eq_class,CensusRegion,Party) %>% tally()
t(q)
```

```
#### Question 4: Which characteristics of the senators predict co-sponsorship? ####

# Week 10 ERGM

# Again, the full network cannot be simulated due to the completeness of the graph. If everyone is connected to everyone, then we cannot predict anything anymore.

# https://rdrr.io/cran/ergm/man/ergm-terms.html
# https://stats.stackexchange.com/questions/149502/regression-model-and-social-network-analysis
# https://www.r-bloggers.com/ergm-tutorial/
# http://badhessian.org/2012/09/lessons-on-exponential-random-graph-modeling-from-greys-anatomy-hook-ups/

require(ergm)

table(qnet %v% "religion")

m1 <- ergm(qnet~edges)
summary(m1)
exp(m1$coef)

# This takes a really long time
m2 <- ergm(qnet ~ edges + nodefactor("party") + nodematch("party") + nodefactor("gender") + nodematch("gender") + nodematch("census") + absdiff("first") + gwesp(0.6, fixed = TRUE))
summary(m2)

#options(scipen=999)
exp(m2$coef)

# calculating the fitted probability
# two women from the same region, who joined in the same year, and who have both cosponsored a bill with one other senator; once, when they are both Democrats
# edges + nodematch.census + absdiff.joined + gwesp.fixed.0.6 + nodematch.party
r = -6.915860 + 0.661768 + 3.810799 + 0.673041 + -0.006935 + 1.861290
exp(r)/(1+exp(r))

# same conditions, but Republicans
r = -6.915860 + 0.661768 + 3.810799 + 0.673041 + -0.006935 + 1.861290 + 2*-1.199017
exp(r)/(1+exp(r))
```

Appendix B: Graphs used in this Project

All of the figures included in this document are reproduced in bigger size in this appendix.

Figure 1

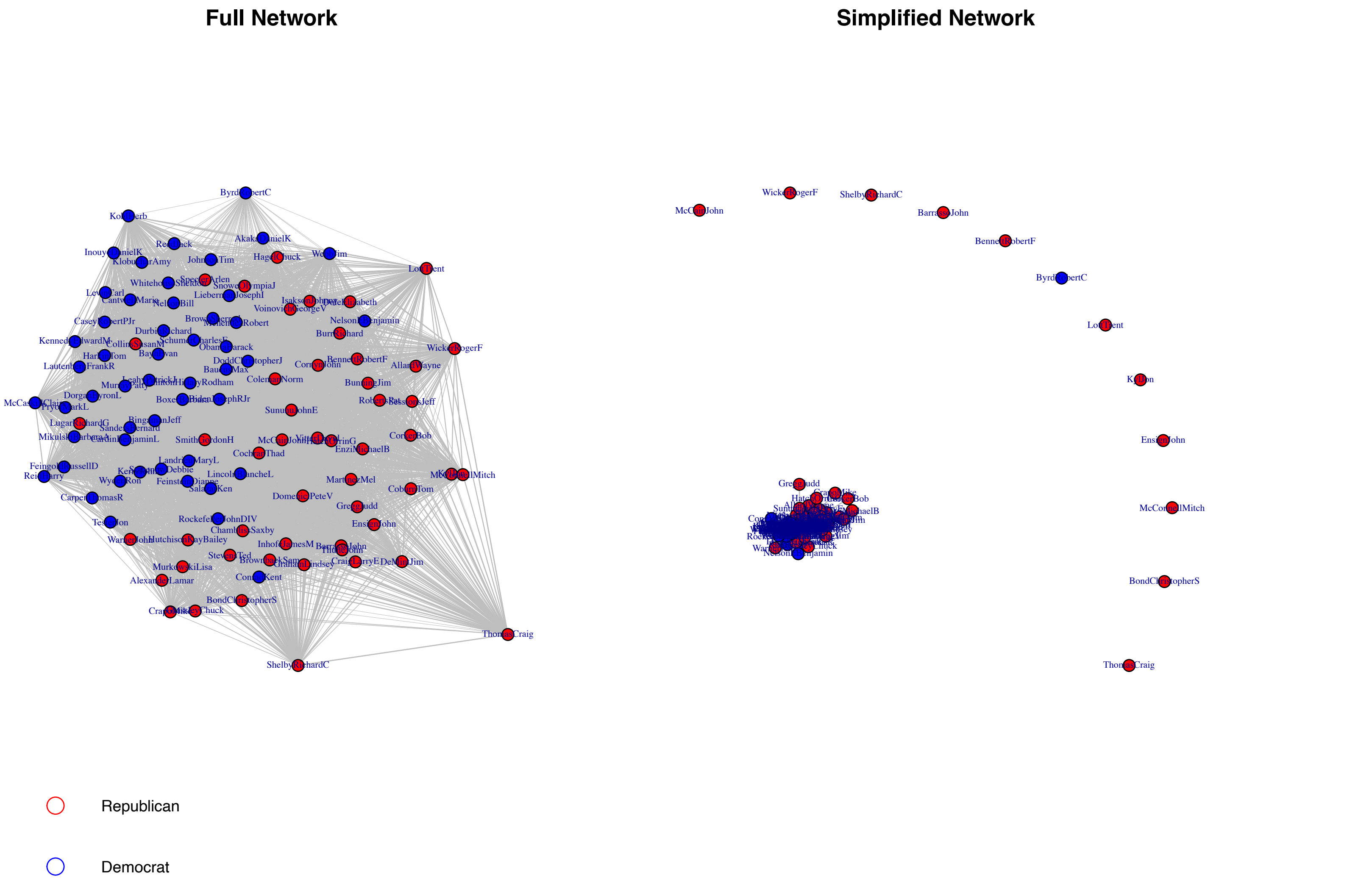


Figure 2

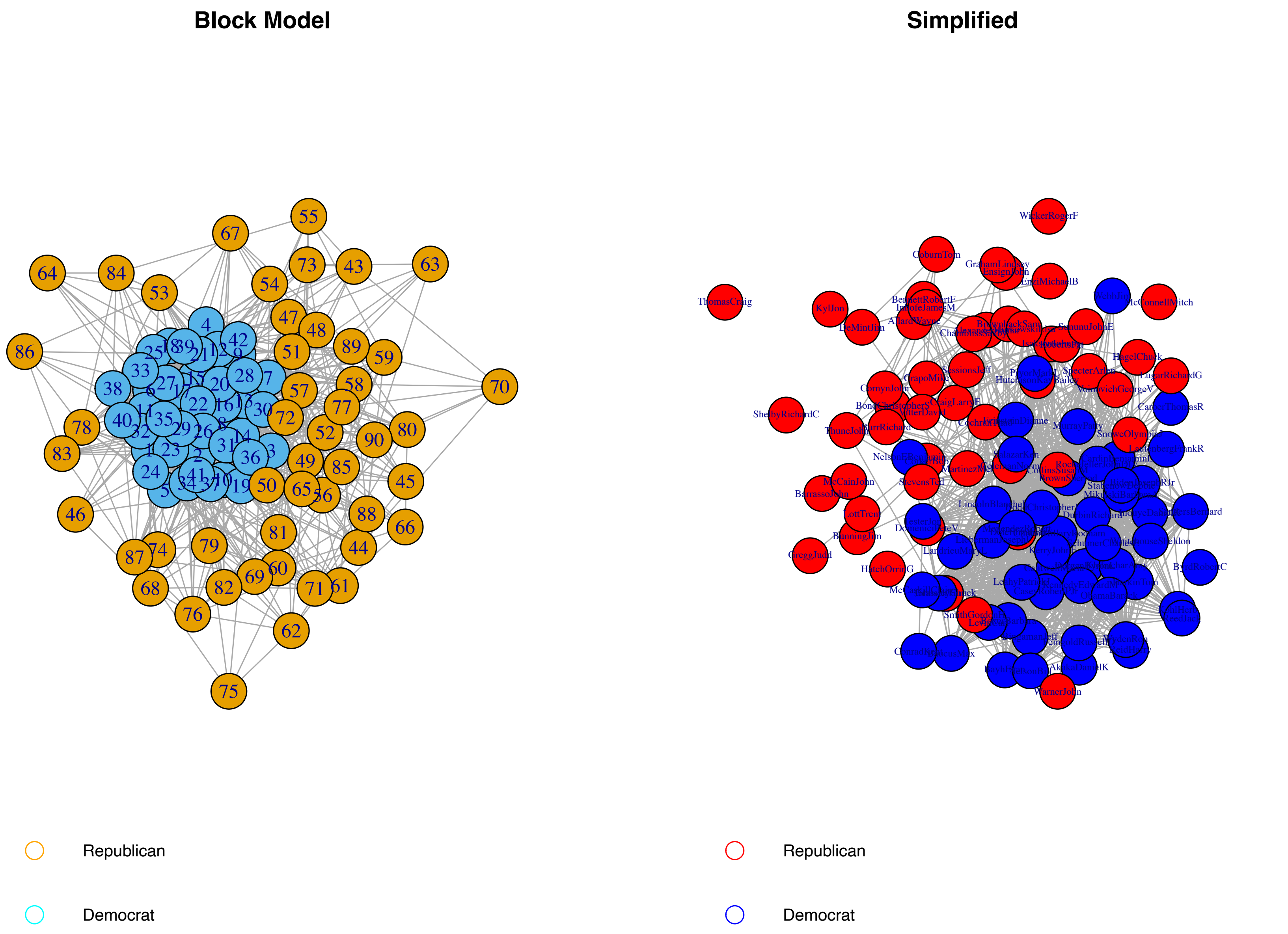

Block Model
Simplified
Republican
Democrat
Republican
Democrat

Figure 3

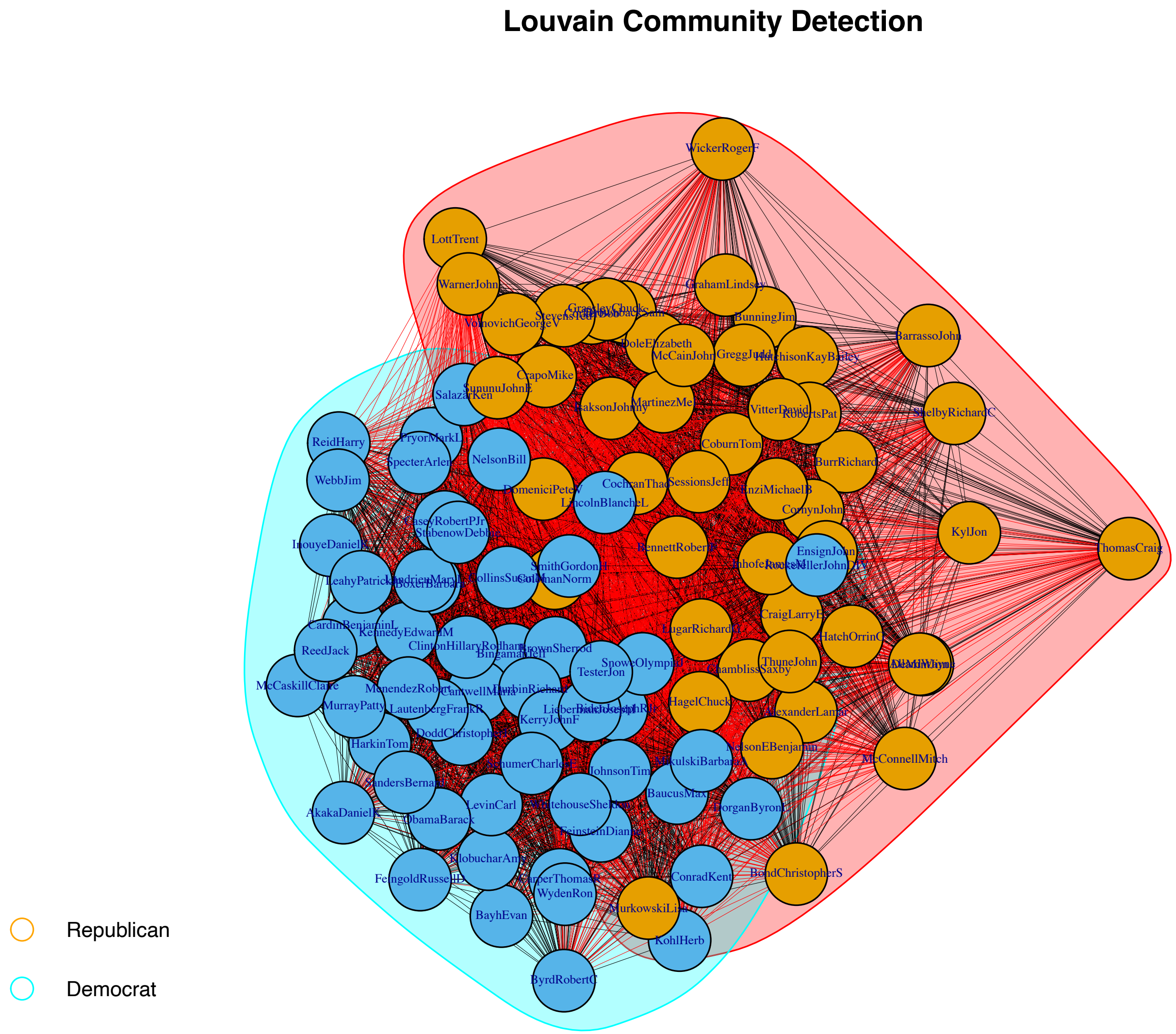

Louvain Community Detection
WickerRogerF
LottTrent
WarnerJohn
BarrassoJohn
KylJon
ThomasCraig
McConnellMitch
BondChristopherS
ByrdRobertC
BayhEvan
KohlHerb
ConradKent
ReidHarry
WebbJim
SpecterArlen
NelsonBill
SalazarKen
ReedJack
MurrayPatty
HarkinTom
AkakaDanielK
ObamaBarack
LevinCarl
TesterJon
HagelChuck
CrapoMike
CoburnTom
BurrRichard
SessionsJeff
McCainJohn
LincolnBlancheL
JohnsonTim
BaucusMax
DorganByron
ThuneJohn
WydenRon
Republican
Democrat

Figure 4

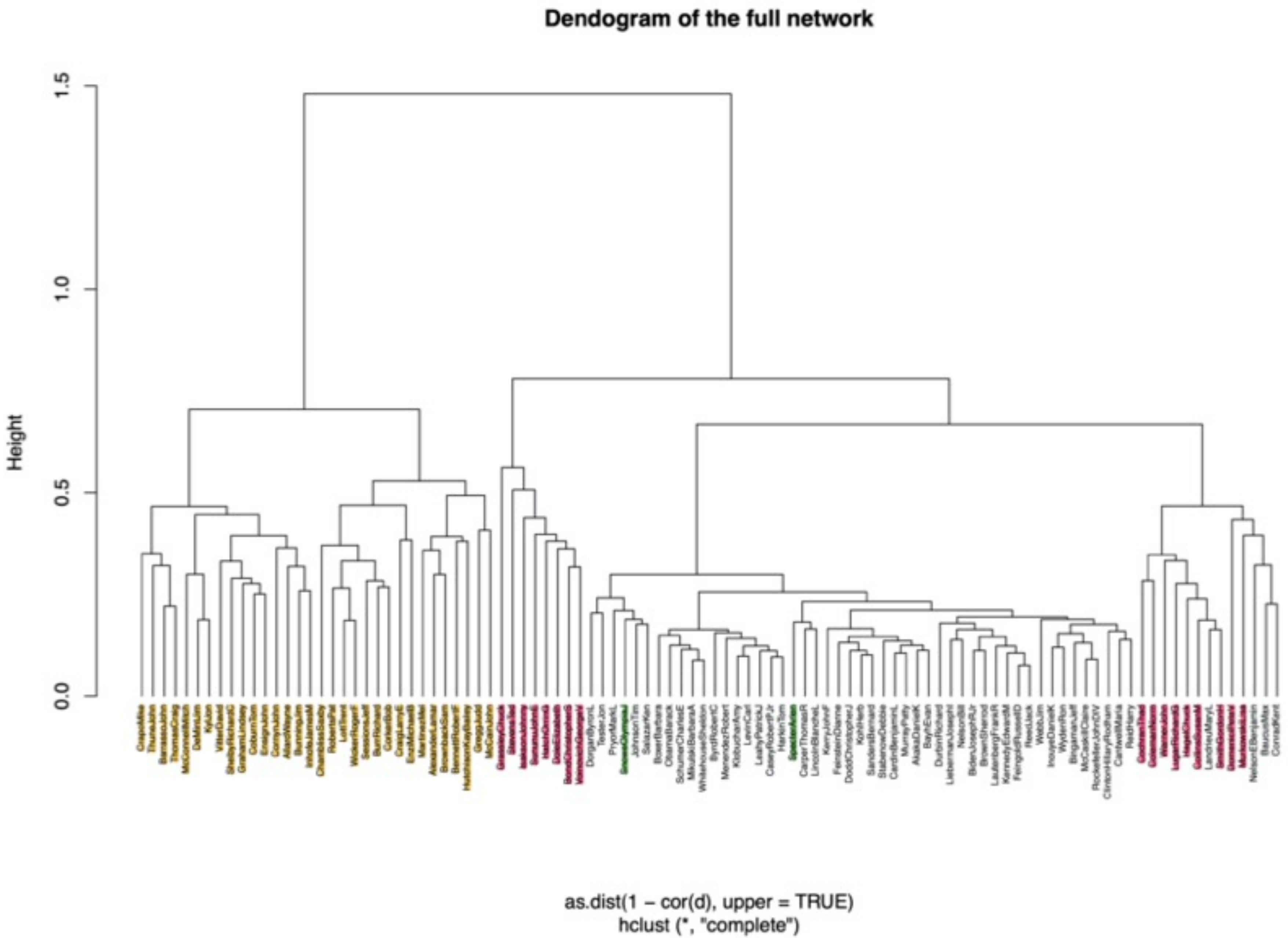
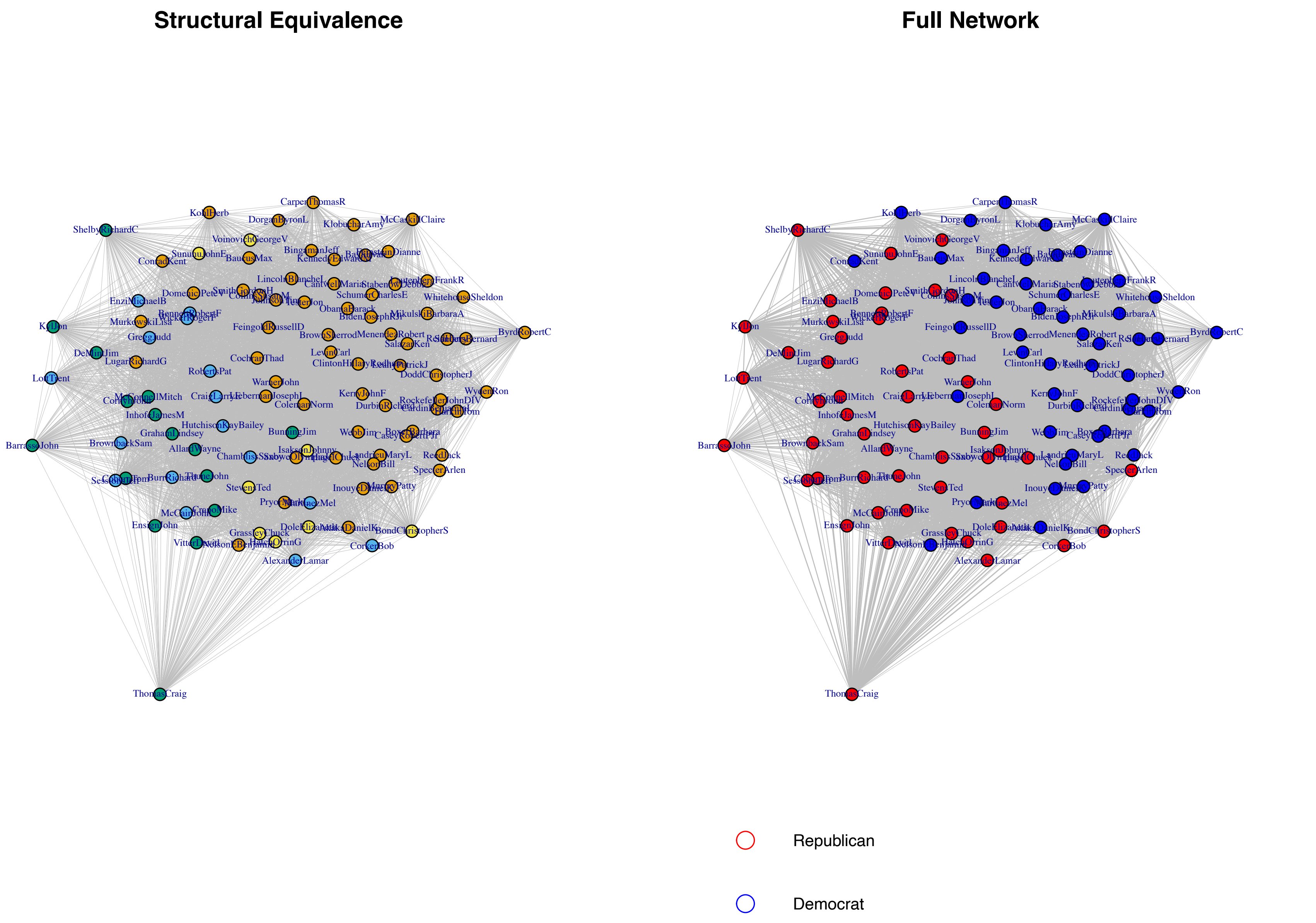

Dendogram of the full network
Height
0.0
0.5
1.0
1.5
as.dist(1 − cor(d), upper = TRUE)
hclust (*, "complete")

Figure 5



Structural Equivalence
Full Network
Republican
Democrat