MY461: Social Network Analysis LT 2019

Take Home Exam

Student: 18126

Question 1

Density is defined as the ratio of the number of edges and the number of possible edges in a network. This concept was first used in Social Network Analysis by Clyde Mitchell, and he interpreted it as the completeness of the network (Scott, 2017, p.33). Usually, social networks with higher density tend to represent a more closely-knit community. For instance, Wellman's (1979) study of inner-city suburb showed that the densest network where the ones mainly composed by members of the same kin. (p.1217)

Average path length, on the other hand, is defined as the average of the geodesic distance between two nodes, for all nodes in a network. This metric can be used aa proxy to measure of how fast information can pass across the whole network. For instance, if the average path length is very low, it means that the network is well connected, and information can be spread quickly throughout the network.

Lastly, transitivity (sometimes called clustering coefficient) measures the probability that the neighbors of a node are themselves connected. High levels of transitivity indicate greater clustering, something usually found in real social networks rather than in randomly generated ones.

Table 1 – Network metrics

Model	Density	Average path length	Transitivity
Complete	1.0000000	1.0000000	1.0000000
Simple	0.2533489	1.859675	0.7622924
Erdos-Renyi	0.2536440	1.747348	0.2534074
Configuration	0.2533489	1.916400	0.6175796

As you can see from table 1, the complete network has all the previously defined metrics equal to 1. This means that the network has all possible edges, the geodesic path for any couple of nodes is a single edge, and for every node, all the neighbors are connected between each other. In other words, the graph is complete.

The simple network, instead, has features more similar to the ones produced by the small world experiment, which are relatively low average path length and high transitivity (Travers & Milgram, 1969).

The difference between the complete network and the simple network is probably due to the fact that often the bills have dozens of cosponsors, hence it is likely for senators to have at least one cosponsored bill in common with any other senator in a two-year period. However, they certainly have had much stronger relations with certain senators probably due to having similar political ideas. As a consequence, when we look only at the strongest relations between senators, there are fewer edges, which causes the lower density and transitivity and increases the path length.

When trying to simulate the simple network, the Erdős–Rényi and Configuration are both extremely accurate for the density and the path length. However, they are both quite far away from the transitivity value of the simple model. The configuration model is definitely much closer than the Erdös-Rényi, probably due to the fact that the configuration model maintains the exact degree distribution of the simple network for each node (in fact the density value is exactly the same). Instead, the Erdös-Rènyi model is randomly generated following a predefined probability.

Question 2

In this network influence can be defined as having cosponsoring relationships with other highly influential senators or with senators which might not have a strong relationship between them, Therefore, the two centrality measures that best describe influence are eigenvector and betweenness centrality.

Eigenvector centrality gives a score to each node depending on the relative importance of the scores of the neighbors. It can be used as a measurement of the social capital of a node, which Bourdieu (1986) defines as the aggregate of the actual or potential resources which are linked to possession of a durable network (p.244). The benefits of large social capital in the context of the simple network could be being able to have important senators cosponsoring one's bill or being able to tweak another senators bill to push one's agenda.

Betweenness centrality instead measures how often a node is part of the shortest path between other nodes. Therefore, a node with a high betweenness might have low degree, but it lies in such a position that many geodesic paths pass through it (Scott, 2017, p.98). High betweenness has been correlated in social networks with the position of broker between two different sides of a network (ibid., p.99). In the simple network this means that a senator with high betweenness could act as an intermediary between two senators that might have diverging political ideas.

I believe that the three most influential senators are: Hilary Clinton, Norm Coleman, and John Kerry.

Clinton has the highest eigenvector centrality and, as shown in table 2, the average and median eigenvector score is much higher in Clinton's ego network than in the simple one. Her large social capital is probably due to the two decades in the upper echelons of Washington as first lady and as senator. This makes her incredibly influential as she is one of the most well-known and well-connected figures in the Senate. Anyone with her cosponsorship could immediately use her social capital to give more standing, and possible more cosponsors, to the bill.

Table 2 - Eigenvector centrality summary statistics

Network	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
Simple	0.0002458	0.0191470	0.2088768	0.3192606	0.6210471	1.00000
Clinton	0.06148	0.29143	0.54375	0.51455	0.70742	1.00000
Coleman	0.008206	0.100273	0.504522	0.437937	0.681741	1.00000
Kerry	0.0513	0.2828	0.5384	0.5070	0.7031	1.00000

Coleman has the highest betweenness centrality, which is not unexpected considering he used to be a Democrat and was considered one of the most moderate Republicans in senate (GovTrack 2019). As shown in figure 1[1], he can be seen as an intermediary between the two parties having almost as many democratic as republican neighbors (37 Republican, 32 Democratic neighbors.) However, he only has the 27th highest eigenvector score and, as shown in table 2, 25% of the senators in his ego network have eigenvector centrality lower than 0.1. Therefore, his influence may also come from the willingness to cooperate with less prominent senators, a concept known as strength of weak ties (Mark Granovertter 1973).

Kerry has both the second highest betweenness and eigenvector centrality. Therefore, he has both types of influence that Coleman and Clinton have, probably making him the most influential senator in the network. Yet, he might lack Coleman's level of strength of weak ties and Clinton's level of notoriety.

[1] All figures in appendix

Question 3

Party assortativity in the simple network is relatively low but positive, with a value of 0.2919754. This means that there is a 29.20% chance that two members of the same party have a strong cosponsorship relation between them. As a consequence, we can conclude that party affiliation has a small degree of influence on bill cosponsorship, but it does not absolutely guarantee that same party senators will cosponsor the bill.

Table 3 – Stochastic blockmodel matrix

	Democratic	Republican
Democratic	0.6188235	0.1364860
Republican	0.1364860	0.1262745

However, this result is an average of the behavior of the whole network. In fact, table 3 clearly shows very different behaviors from the senators of the two parties: The probability of ties within the Democratic party is quite high (61.88%), while ties within the Republican party have a similar probability (12.63%) as ties between members of different parties (13.65%). This is evidence that the Democratic senators are much more likely to cosponsor the bills of members of the same party, probably due to the fact that having a majority and they do not necessarily need the support of Republican senators to pass a bill. Instead, the Republican senators do not really seem to have a preference for cosponsoring the bills of their own party. The result actually suggests a marginal preference for members of the Republican party to cosponsor bills with Democratic senators but, being the difference in probability only around 1%, we can consider the Republican senators as having no preference when it comes to cosponsoring bills.

This might be due to a variety of reason: maybe they decide on a case by case basis instead of strictly cosponsoring bills along party lines. This is not surprising behavior for a party in the minority since they necessarily need to collaborate with a member of the majority party to enact legislation. Another possibility is that the Republican party includes members of a very wide political spectrum, which might not necessarily agree with the contents of another party member's bill. This seems to be supported by the ideology analysis conducted by UCLA (figure 6) which shows a very spread out Republican party both in terms of economic and social ideology.

Lastly, the Louvain community algorithm identifies two communities, and these communities align quite well with party membership, as proven by the comparison via normalized mutual information, which produces a result of 0.6475944, the highest of all network attributes. The results are even more clear in figure 2. Party affiliation is such a close match that only 7 nodes are classified differently in the Louvain communities. This seems to show that also single cosponsoring relations are affected by party membership, maybe even more strongly than the simple network. However, these results are to be taken with a pinch of salt as the comparison results shows that the communities are not perfectly equivalent another community detection algorithm might produce very different results.

Question 4

Figure 3 shows that there are no structurally equivalent nodes, or in other words, there are no two senators that have exactly the same connections. However, as the heatmap (figure 4) shows there seems to be one group - the one forming the red rectangle in the bottom right - that has a relatively homogeneous behavior.

As you can see from figure 5, after splitting the dendrogram into four classes it seems pretty clear that this group – Class 1 – is composed by Democrats and very liberal Republicans such as Snowe, Collins or Coleman (Lewis, 2019). This is probably caused by the Democratic party having control of the Senate, which means that we could identify this group as being in the position of working majority of the Senate. As such, they do not necessarily need the support of the other classes, to pass bills. It is hard to find a leadership figure in such a large class. However, I assume that, that Clinton or Kerry could be seen as leaders, as Question 2 determined that they were the most influential senators in the network.

Class number 2 is mostly composed by Republican senators that could be defined as moderate conservatives such as Lamar, Gregg or Bennett (Govtrack, 2019). However, this class also has Brown, a very liberal Democratic senator that seems to have very little in common with the others. Interestingly, he was also classified as a Republican by the Louvain community detection algorithm, meaning that he might be potentially showing a cosponsorship behavior more akin to a moderate Republican

Class 3 is probably the most clearly defined of the three opposition classes. They are the only class that is 100% composed by Republican, Christians males and their most prominent senator, and also minority leader for the Republicans, is Mitch McConnell, a politician famous for his very conservative views on topics ranging from foreign politics to abortion and drugs (Zengerle, 2013). Therefore, the role of the senators of Class 3 could be described as extreme conservatives. In fact, it contained the most extremely conservative right-wing senators according to various analysis of the voting patterns of the Senate, such as Inhofe, Barrasso, and Coburn (Lewis, 2019 & GovTrack, 2019).

Lastly, Class 4 is composed of Republicans without a specific leaning. I believe that their behavior is best described by the heatmap that shows a long red rectangle on the left that crosses the whole graph top to bottom without having drastic changes in color. This shows a group of senators, mostly part of Class 4, that are relatively close to both Republican and Democratic alike. Their position could be defined as true centrist and is probable that they tend to cosponsor senators of both parties without a specific preference.

Question 5

Table 4 – Odds ratios of ERGM coefficients

Factor	Lower	Odds Ratio	Upper
edges	3e-04	0.0012	0.0048
nodefactor.party.Republican	0.2801	0.3	0.3213
nodematch.party	5.5056	6.3028	7.2156
nodefactor.gender.Male	0.2564	0.3296	0.4235
nodematch.gender	1.4233	1.9098	2.5625
nodematch.region	1.5581	1.9406	2.417
absdiff.joined	0.9853	0.993	1.0008
gwesp.fixed.0.6	19.8658	41.4298	86.4011

Nodefactor.party.Republican shows that Republicans are 70% less likely to have a strong cosponsorship tie than Democrats. This is probably due to the fact that during the 110th Congress the Republicans were in minority. Hence, they probably cosponsored fewer bills overall, and as a consequence created less strong cosponsorship ties than the Democrats.

Nodematch.party shows that members of the same party are 6.3 times more likely to have a strong cosponsorship relationship than members of different parties. This is hardly surprising considering the results of the previous question, and it is further proof that there is strong party homophily in the simple network. Even the lowest estimate in the odds ratio confidence intervals is still double the next highest (nodematch.region). This is also the second strongest term after gwesp.fixed.0.6. which shows how important party ideology is in the creation of strong cosponsorship relations.

Nodefactor.gender.Male shows that male senators male senator are about 67% less likely to have a strong cosponsorship tie than female senators. This is further proof that female senators are very active and tend to cosponsor more bills than their male counterpart (Klein, 2015).

Nodematch.gender shows that same-sex senators are 1.91 times more likely to have a strong cosponsorship bill than between one male and one female senators, which shows the existence of some gender homophily in the simple network. The homophily for female senators might be even stronger the finding from Klein (2015) which also states “Women in the Senate traditionally meet for periodic dinners (...) Commentators have in part attributed this clear bipartisan effort in which women in the Senate can form personal relationships to the willingness of women to work on legislation with each other"

Nodematch.region shows that senators of the same Census region are 1.94 times more likely to have a strong cosponsorship bill senators from different regions, which shows the existence of some regional homophily in the simple network. This is probably due to shared interest from senators in cosponsoring bills that might benefit more their own census region.

Absdiff.joined is the only non-significant term, which suggests that the difference between the year that senators joined the Senate does not significantly change the probability of a strong cosponsorship relation forming.

Gwesp.fixed.0.6 shows that the existence of common neighbors is the most important factor in determining the probability that two senators will have a strong cosponsorship relation between them. This is supported also by the results of Question 1 that showed the simple network as having strong transitivity, as expected by a real-life social network. The odds ratio shows that the existence of one common neighbor between two senators increases the probability of the two senators having a strong cosponsorship relation (i.e. triadic closure) by a staggering 41 times. This clearly shows that having strong cosponsorship relations with key members of the Senate (such as the ones identified by Question 2) provides a pathway for new strong relations with members of their ego network.

Table 5 – Probability of Ties between two women from the same region, who joined in the same year, and who have both cosponsored a bill with one other senator

Democrat – Democrat	**Republican – Republican**
0.5332011	0.09321472

Bibliography

Bourdieu, P. (1986). The forms of capital. In J. Richardson, *Handbook of theory and research for the sociology of education year* (1st ed., pp. 239-258). Westport: Greenwood Press.

GovTrack. Norm Coleman, former Senator for Minnesota. Retrieved from https://www.govtrack.us/congress/members/norm_coleman/300024

Granovetter, M. (1973). The Strength of Weak Ties. *American Journal Of Sociology*, *78*(6).

Klein, M. (2015). Working Together and Across the Aisle. Retrieved from https://www.quorum.us/data-driven-insights/working-together-and-across-the-aisle-female-senators-pass-more-legislation-than-male-colleagues/311/

Lewis, J. Voteview. Retrieved from https://www.voteview.com

Scott, J. (2017). *Social network analysis* (4th ed.). London: SAGE.

Travers, J., & Milgram, S. (1969). An Experimental Study of the Small World Problem. *Sociometry*, *32*(4).

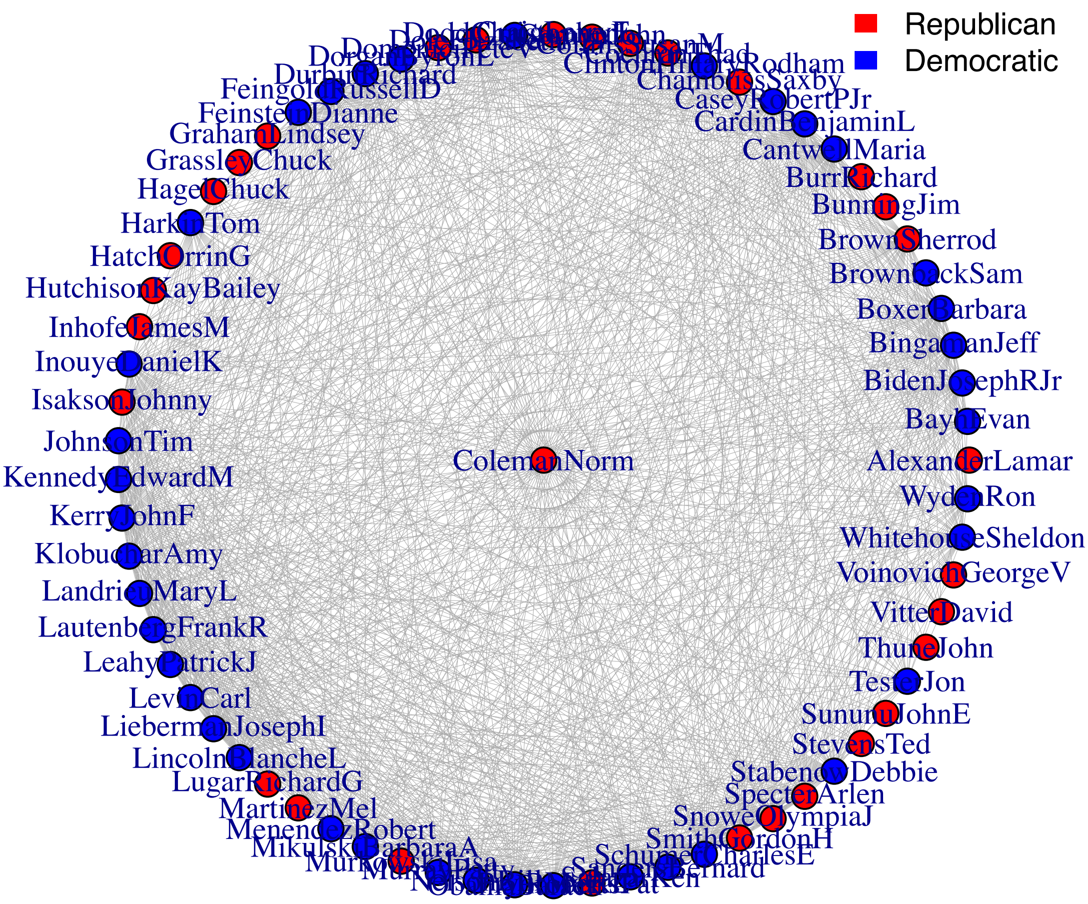
Wellman, B. (1979). The Community Question: The Intimate Networks of East Yorkers. *American Journal Of Sociology*, *84*(5).

Zengerle, J. (2013). Get Mitch. Retrieved from https://www.politico.com/magazine/story/2013/11/get-mitch-mcconnell-99376_Page2.html

Appendix

Figure 1

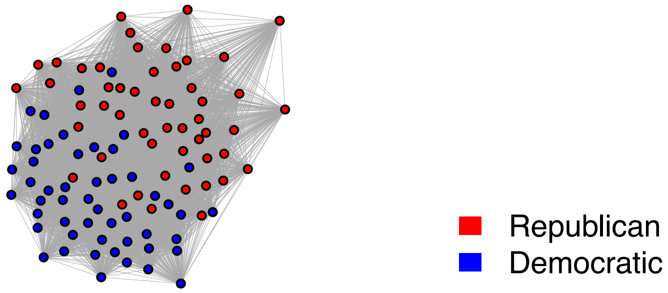
Norm Coleman Ego Graph



Larger nodes have higher eigenvector centrality

Figure 2

Vertex colour from Party memebership



Vertex colour from Louvain community detection algorithm

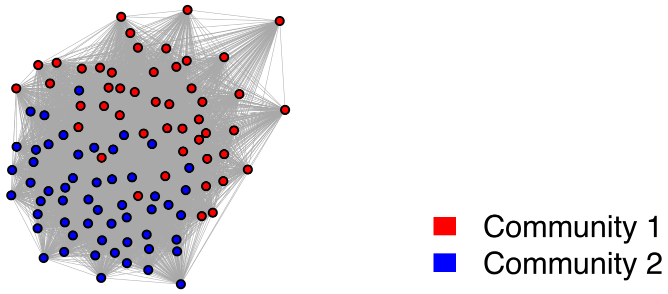


Figure 3

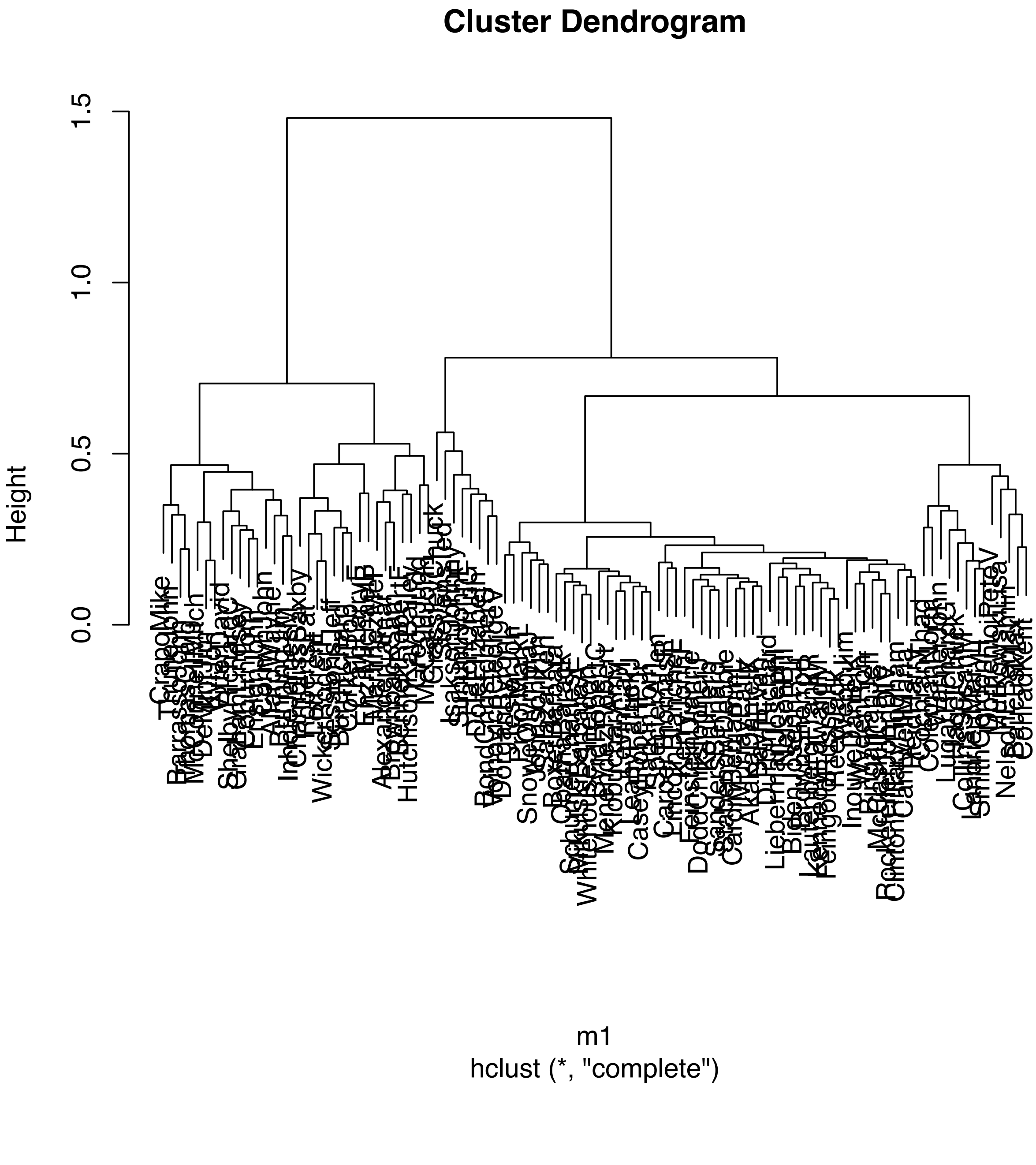


Figure 4

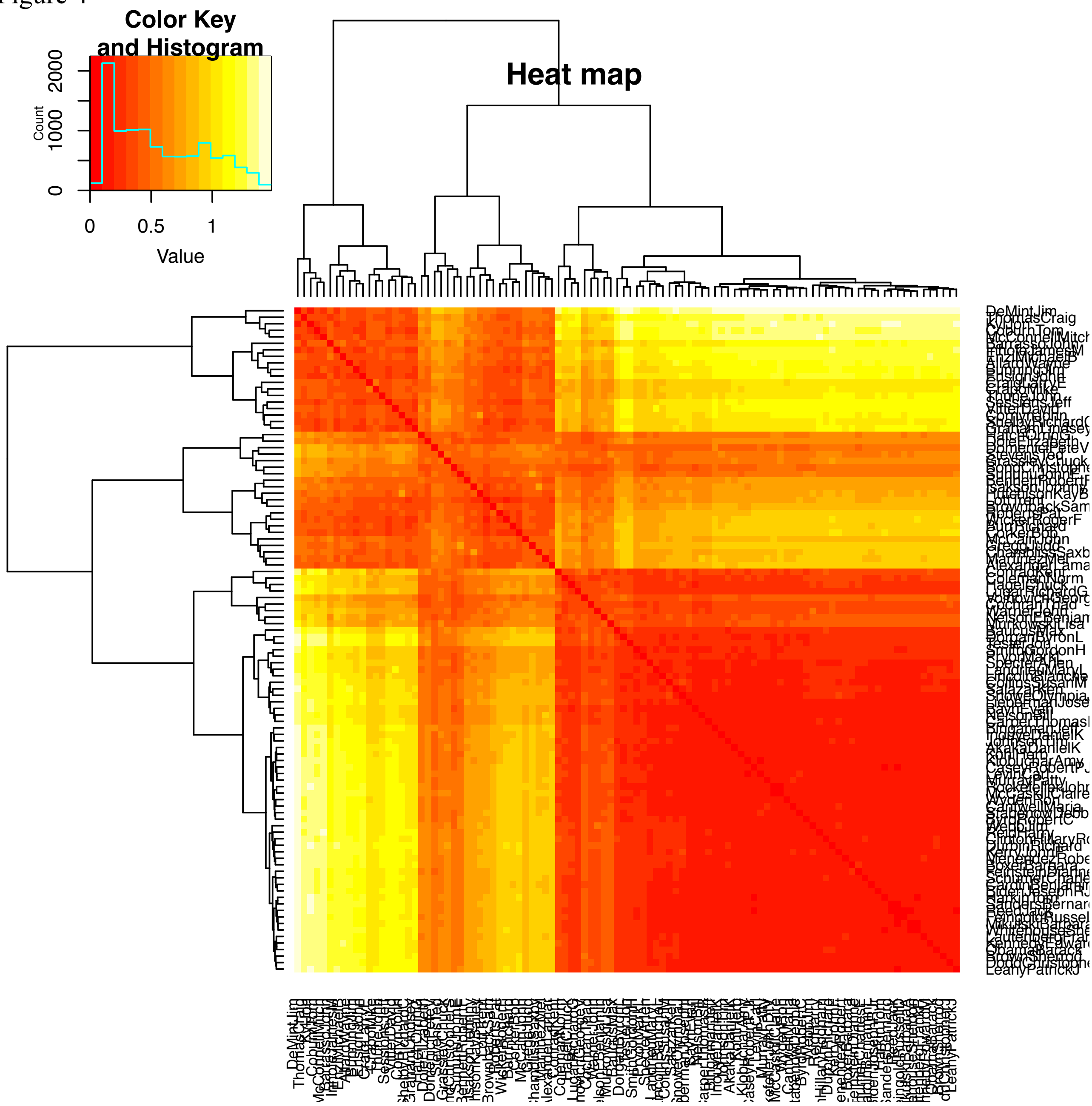

Color Key
and Histogram
Count
0
1000
2000
0
0.5
1
Value
Heat map

Figure 5

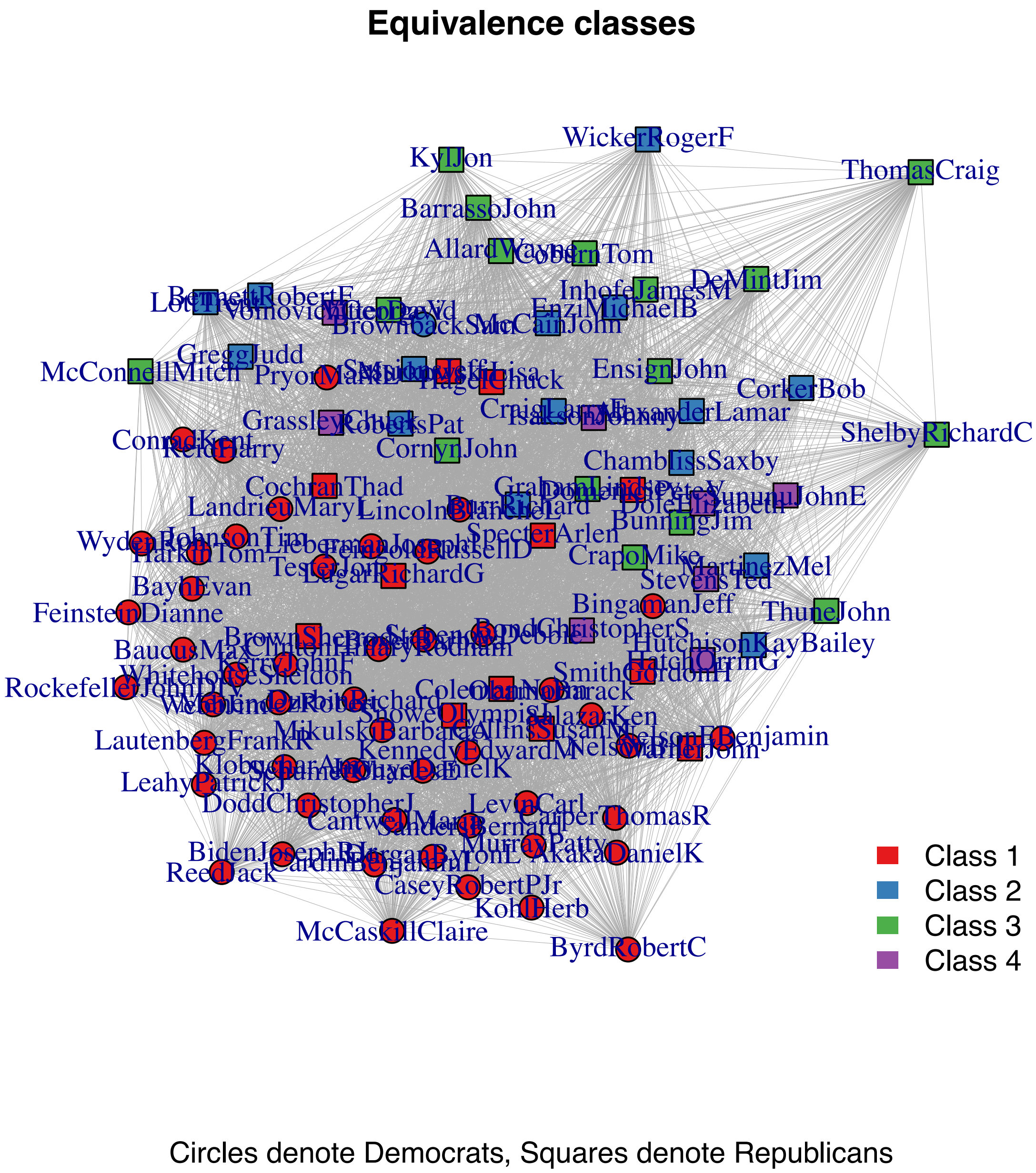

Equivalence classes
KylJon
WickerRogerF
ThomasCraig
BarrassoJohn
DeMintJim
McConnellMitch
EnsignJohn
CorkerBob
ShelbyRichardC
CornynJohn
ChamblissSaxby
CochranThad
BunningJim
ThuneJohn
BingamanJeff
FeinsteinDianne
BaynEvan
LautenbergFrankR
LeahyPatrickJ
ReedJack
LevinCarl
CaseyRobertPJr
KohlHerb
McCaskillClaire
ByrdRobertC
Class 1
Class 2
Class 3
Class 4
Circles denote Democrats, Squares denote Republicans

Figure 6 - 110th Congress Ideology (Lewis, 2019)