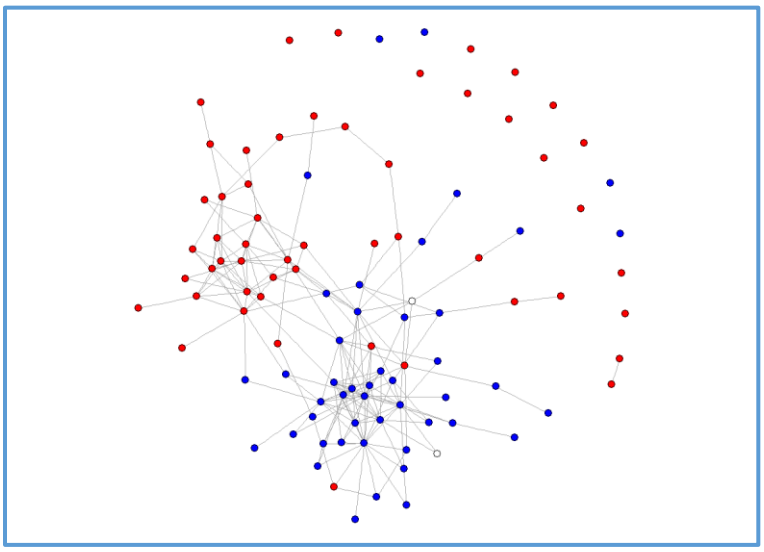
15.071 Analytics Edge - Homework Assignment # 5 Social Network Analysis

Problem-1.) Social Networks of the United States Senate

1.a.) Screenshot of the network graph:

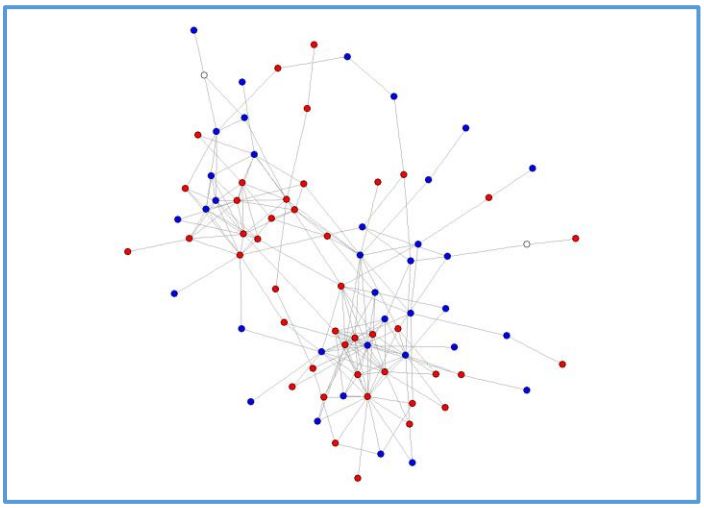


Observations:

- There are 100 senators observed.
- There is 1 large connected component with 81 members, there is 1 another connected component with 2 senators. The rest of the network is all **17** individual senators with no connections.
- There are **19** senators disconnected from the largest connected component.
- There are 15 Republican senators and 4 Democrat senators disconnected from the largest component.
- There are 2 Independent senators. One of them is connected to both Republicans and Democrats. The other Independent senator is connected to only Democrat senators.

1.b.) Social Senators

There are 81 senators in the largest connected component of this social network. The plot is below:



1.c.) Degree

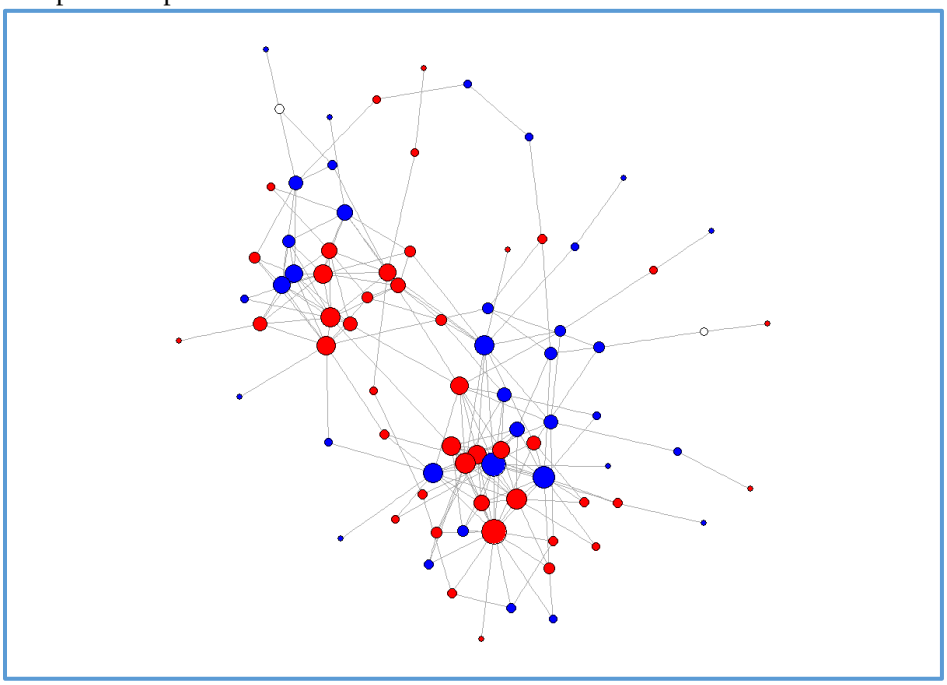
Sherrod Brown, Democrat from OH, has the highest degree of centrality = **19**.
The top 12 senators in terms of degree are listed below:

#/Name / Degree / Part Affiliation

1. Sherrod Brown / 19 /D
2. Richard Blumenthal / 18 / D
3. Jeff Merkley / 15 / D
4. Al Franken / 13 / D
5. Edward Markey / 13 / D
6. Kirsten Gillibrand / 12 / D
7. Orrin Hatch / 12 / R
8. Jon Tester / 12 / D
9. Tammy Baldwin / 11 / D
10. Michael Crapo / 11 / R
11. Marco Rubio / 11 / R
12. Sheldon Whitehouse / 11

There are 9 Democrats and 3 Republicans in the top 12.

The plot is copied below:

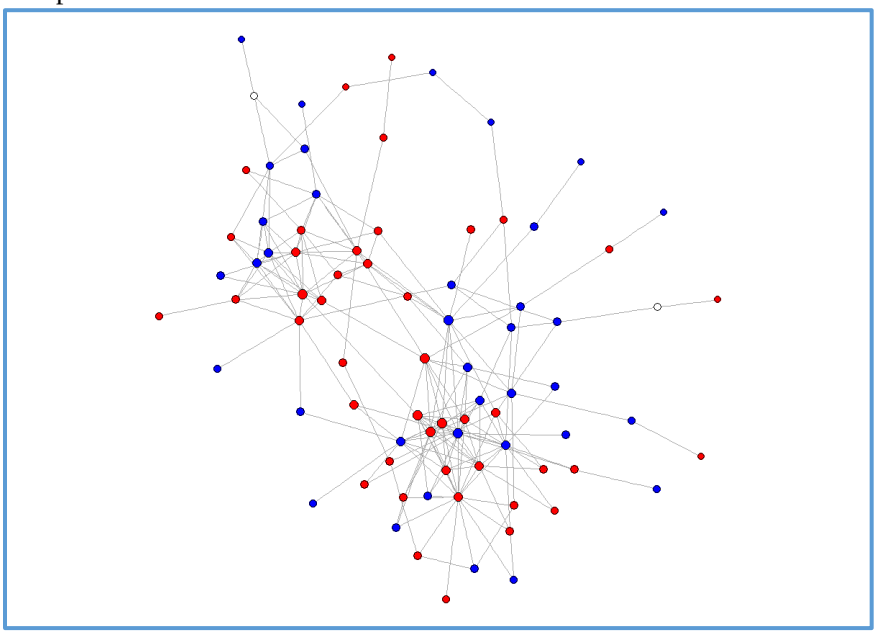


1.d.) Closeness

I observe less variability in closeness compared to degree. The closeness factor varies between **0.005128205** (the highest value) for Richard Blumenthal, D, CT and **0.002409639** (the lowest value) for Dan Sullivan, R, AK.

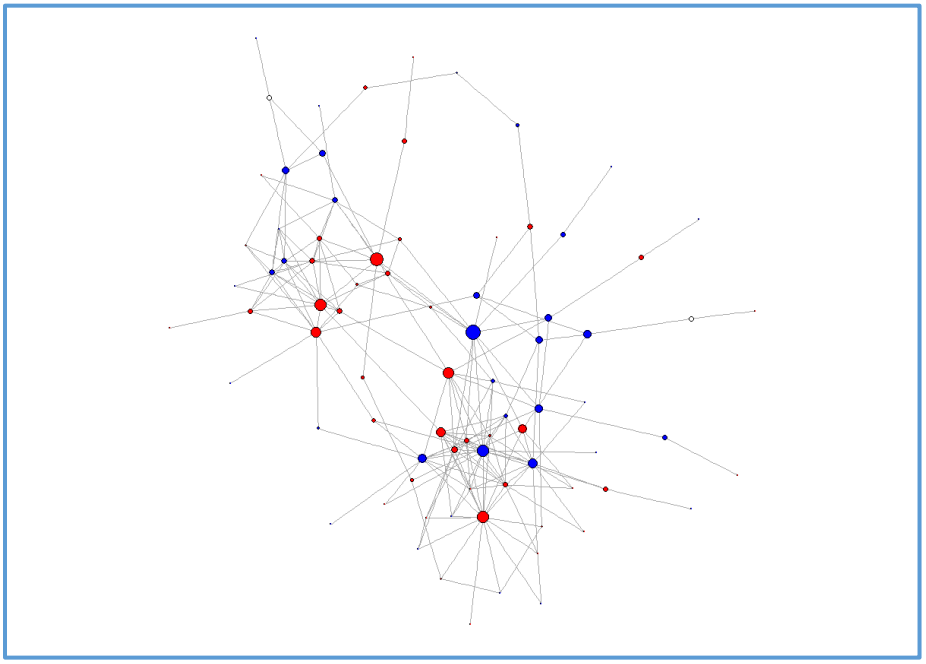
Also Richard Blumenthal has the lowest Average Shortest Path value of **2.4375** whereas Dan Sullivan has the largest Average Shortest Path value of **5.1875.**

The plot is below:



1.e.) Betweenness

The senator with the highest betweenness is Jon Tester whose betweennness score is **672.2660451**



1.f.) Glue

These senators below stand out with high betweenness and relatively low degree with low degrees. Their party affiliation is indicated in the list below:

id	name	state	party	degree	closeness	betweenness
68	Patty Murray	WA	D	6	0.0045662	252.2952979
18	Susan Collins	ME	R	6	0.0042553	207.4566682
11	Maria Cantwell	WA	D	4	0.003861	186.0087851
8	John Boozman	AR	R	6	0.0034247	178.6145529
81	Brian Schatz	HI	D	5	0.0038314	162.802845
53	Angus King	ME	I	4	0.004	157.5913059
6	Roy Blunt	MO	R	3	0.0036765	150.8723033
69	Clarence Nelson	FL	D	4	0.0038462	144.057017

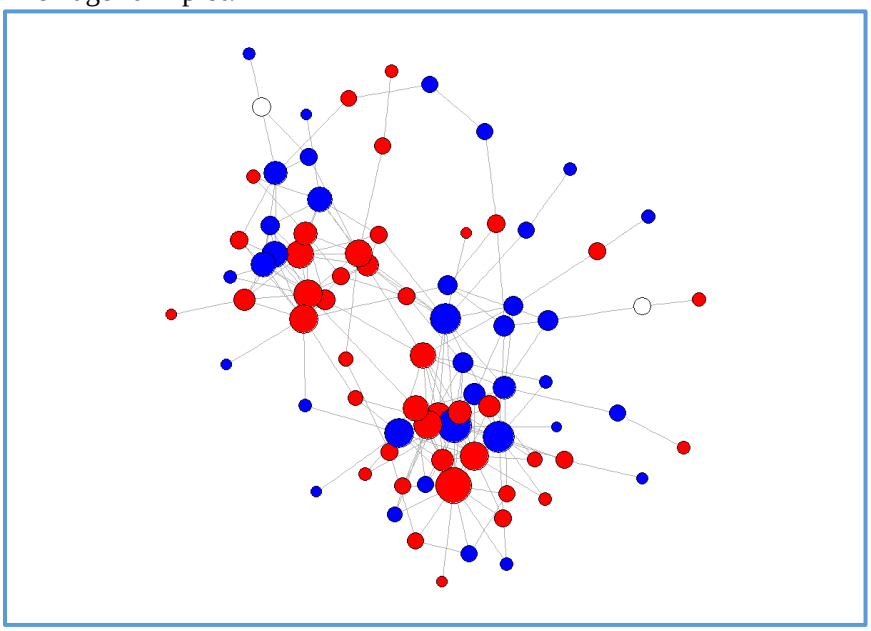
1.g.) PageRank

The senator with the highest PageRank is Sherrod Brown with PageRank score of **0.039258809**
We observe that there is a large positive correlation between PageRank score and degree, as displayed in the list below. The correlation between PageRank and Degree column is **0.9818692.**

id	name	state	party	degree	closeness	betweenness	pagerank
9	Sherrod Brown	OH	D	19	0.0046948	434.2739259	0.039258809
5	Richard Blumenthal	CT	D	18	0.0051282	405.2955275	0.036033578
64	Jeff Merkley	OR	D	15	0.0046083	287.4174125	0.030650711
89	Jon Tester	MT	D	12	0.0051282	672.2660451	0.02879278
37	Kirsten Gillibrand	NY	D	12	0.0046729	234.8079204	0.025809421
78	Marco Rubio	FL	R	11	0.0043478	351.8496189	0.025501048
59	Edward Markey	MA	D	13	0.0044643	87.0417751	0.024969451
35	Al Franken	MN	D	13	0.004902	143.1669515	0.024858574
42	Orrin Hatch	UT	R	12	0.004902	415.3974167	0.024831223
24	Michael Crapo	ID	R	11	0.0041841	97.1395214	0.023496598
45	Dean Heller	NV	R	9	0.0047847	548.9794148	0.022345409
25	Ted Cruz	TX	R	10	0.0042194	119.7816926	0.021361158

The PageRank plot is copied below.

The PageRank plot:

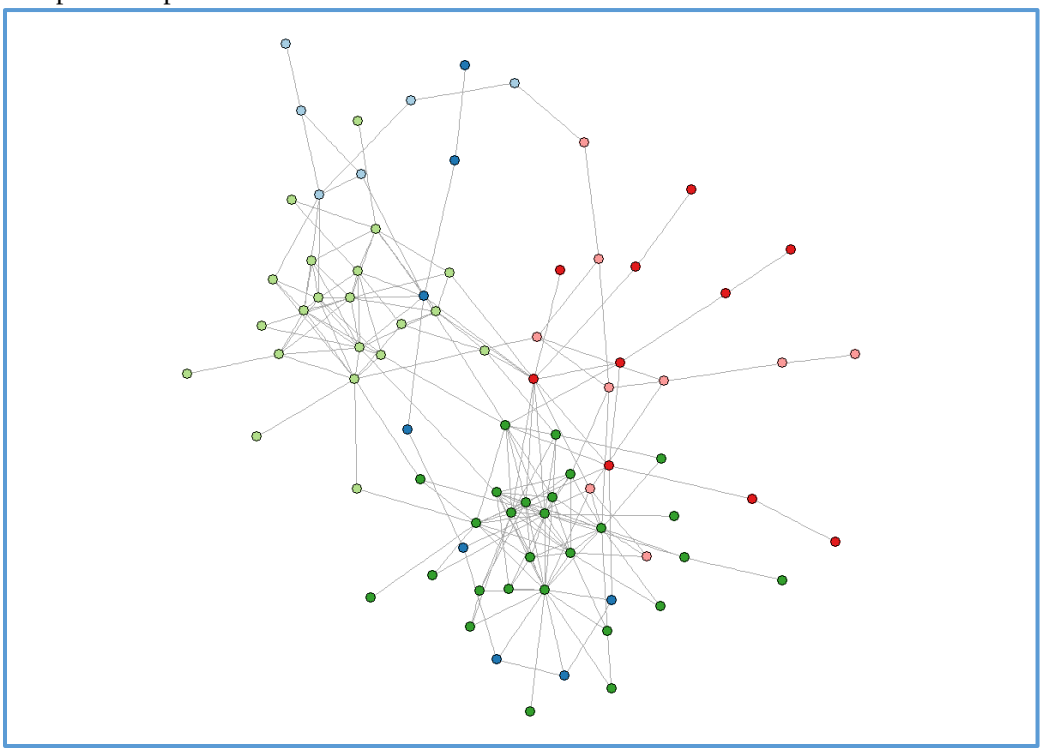


1.h.) Community Detection

Modularity of the computed community assignment is **0.5071247**

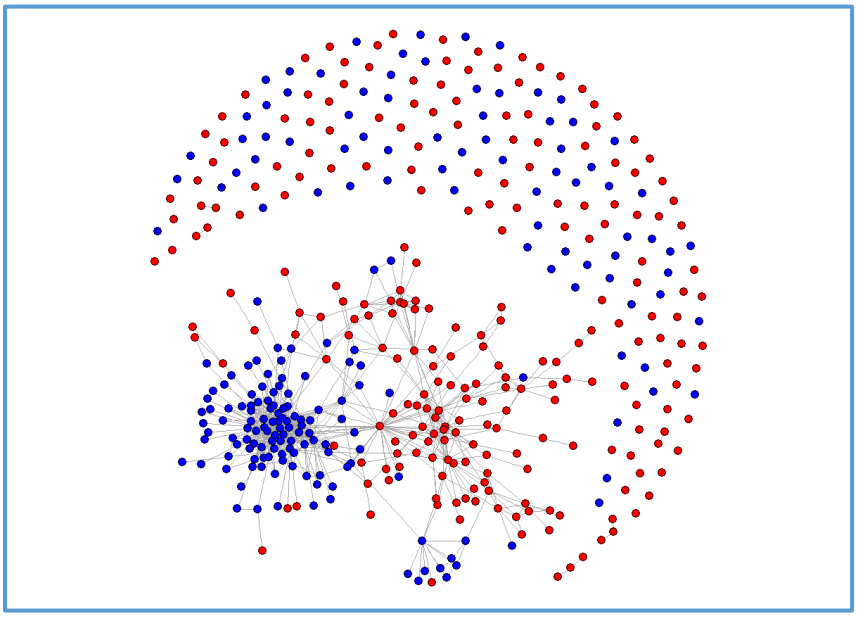
There are 6 communities selected. Only one of the six communities (community-1) is entirely Republican. All other communities are a mix of Republicans and Democrats.

The plot is copied below:



Problem-2.) Social Networks of the United States House of Representatives

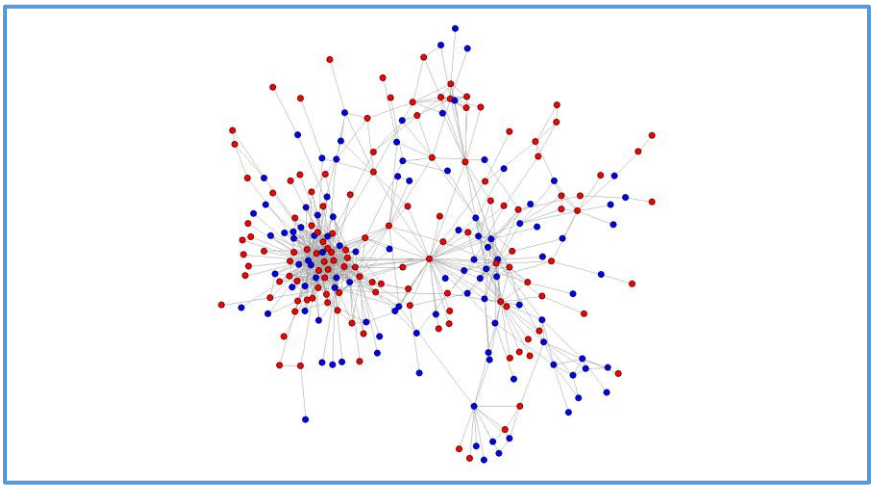
2.a.) The plot is copied below:



Observations:

- There are 438 house members observed.
- There is 1 large connected component with 240 members, there are 2 connected components with 3 members and 3 connected components with 2 members. The rest of the network is all 186 individual members with no connection.
- There are 198 house members disconnected from the largest connected component.
- There are 186 house members with no connections.

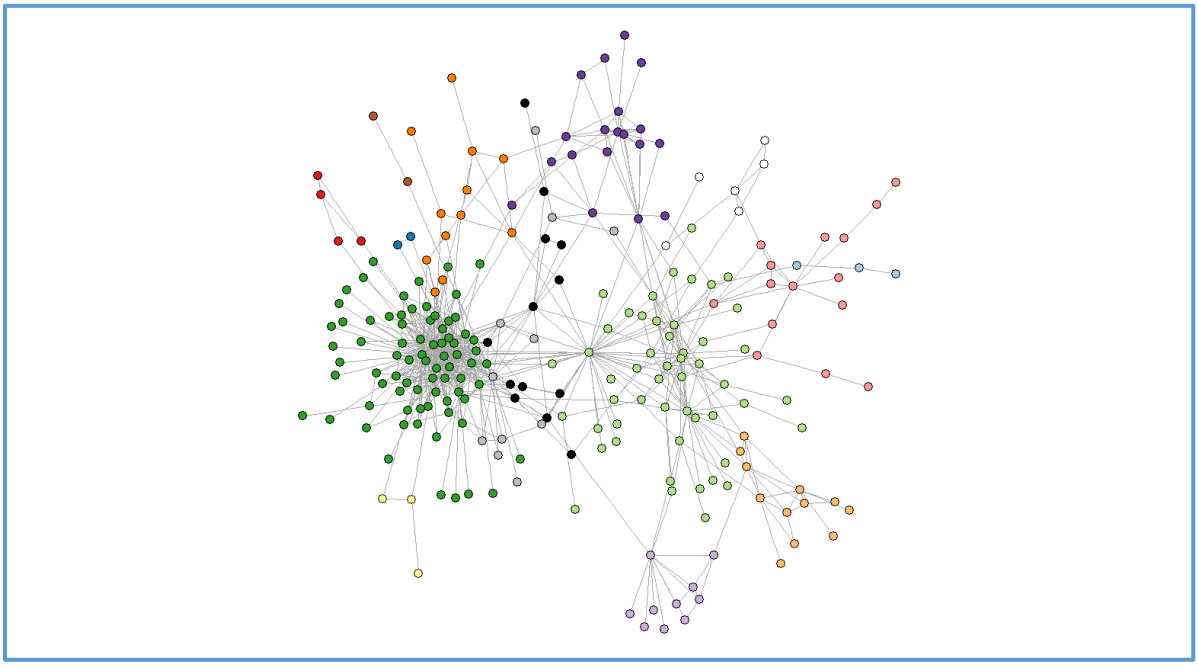
2.b.) Social House Members: There are 240 house members in the largest connected component of this social network. The plot is below:



2.h.) Community Detection

Modularity of the computed community assignment is **0.5220956**
There are 15 communities selected.
Community-2 entirely Democrat and composed of 2 members.
Community-5 is entirely Republican and composed 15 members.
Community-12 is entirely Republican and composed of 2 members.
Community-14 is entirely Republic and composed of 6 members.
All other communities are a mix of Republicans and Democrats.

The plot is copied below:



Problem-3.) Comparing the Senate and the House of Representatives

- The senators network has a larger percentage of its nodes in the social group: 81% (That's 81 senators out of 100 are connected to the largest connected component)
- The house members network has 54.8% percent of its nodes in the social group. That's 240 house members of 438 are in the largest connected component.
- The house members has higher modularity (=0.522) than the senators network which has modularity value of 0.507. This shows that house members social network has more dense connections between the nodes within modules but sparse connections between nodes in different modules. Modularity is a measure of the strength of division of a network into modules (communities).
- We observe that house members social network has been divided into a larger number of communities (15 communities) compared to senators social network which has 6 communities. This is somewhat expected since social house members have 240 nodes whereas social senators network has 81 nodes.

  If I could explore one thing in more depth in these communities, that would be looking at their tenures (i.e. how many times they have been re-elected by their constituencies) and analyzing if there is any meaning correlation between their connectedness and their tenures. For example, the members who are connected to the largest connected component and have higher degrees, have more successful re-election records or not? Or is it not a factor at all? I would like to explore these questions that looks at the reelection success under the light of connectedness/bipartisanship.

APPENDIX: R-Code

```
### HW-05 Social Network Analysis
### PROBLEM-1
### Part-1.a.)

install.packages("igraph")
install.packages("RColorBrewer")
library(igraph)
library(RColorBrewer)

senators = read.csv("senators.csv")
senatorLinks = read.csv("senateCosponsorship.csv")
head(senators)
head(senatorLinks)

G = graph.data.frame(senatorLinks, directed=F, senators)

set.seed(144)
mylayout = layout.auto(G)

par(mar=c(0, 0, 0, 0))
col <- ifelse(V(G)$party == "D", "blue",
              ifelse(V(G)$party == "R", "red", "white"))

plot(G, layout=mylayout, vertex.label=NA, vertex.size=3, vertex.color = col)

### Part-1.b.) Social Senators

comp = components(G)
comp$membership
table(comp$membership)

in.max.comp = (comp$membership == 2) # this is a vector of TRUE/FALSE
sg = induced_subgraph(G, in.max.comp) # this gives us a new graph object
sg_mylayout = mylayout[in.max.comp,]

socialSenators <- senators[in.max.comp,]
summary(socialSenators)
head(socialSenators)

plot(sg, layout=sg_mylayout, vertex.label=NA, vertex.size=3, vertex.color = col )
summary(socialSenators)

### Part-1.c.) Degree

degree(sg)
sort(degree(sg),decreasing=TRUE)

plot(sg, layout=sg_mylayout, vertex.label=NA, vertex.color = col,
     vertex.size=2*sqrt(degree(sg)))

socialSenators$degree = degree(sg)
head(socialSenators)
summary(socialSenators)
```

```
### Part-1.d.) Closeness

socialSenators_cl = closeness(sg)
summary(socialSenators_cl)

avgshortestpath=1/(closeness(sg)*(80))
summary(avgshortestpath)

plot(sg, layout=sg_mylayout, vertex.label=NA, vertex.color = col,
     vertex.size=50*sqrt(socialSenators_cl))

sort(closeness(sg),decreasing=TRUE)

socialSenators$closeness = closeness(sg)
socialSenators$AvgShortestPath = 1/(closeness(sg)*(80))

head(socialSenators)
summary(socialSenators)

### Part-1.e.) Betweenness

bn = betweenness(sg)
summary(bn)
socialSenators$betweenness = betweenness(sg)
head(socialSenators)
summary(socialSenators)

plot(sg, layout=sg_mylayout, vertex.label=NA, vertex.color =
col,vertex.size=0.2*sqrt(bn))

sort(betweenness(sg),decreasing=TRUE)

### Part-1.f.) Glue

head(socialSenators)
summary(socialSenators)

### Part-1.g.) PageRank

pr = page.rank(sg)$vector
summary(pr)

plot(sg, layout=sg_mylayout, vertex.label=NA, vertex.color = col,
vertex.size=70*sqrt(pr))

### Part-1.h.) Community Detection
set.seed(144)
community = cluster_spinglass(sg, spins = 100)

clust = community$membership
clust
table(clust)
community$modularity

color = c(brewer.pal(12, "Paired"), "black","white","gray")
plot(sg, layout=sg_mylayout, vertex.label=NA, vertex.size=3,vertex.color=color[clust])
socialSenators$CommunityAssignment = community$membership
```

```
### HW-05 Social Network Analysis
### PROBLEM-2
### Part-2.a.)

install.packages("igraph")
install.packages("RColorBrewer")
library(igraph)
library(RColorBrewer)

houseMembers = read.csv("houseMembers.csv")
houseMembersLinks = read.csv("houseCosponsorship.csv")

head(houseMembers)
head(houseMembersLinks)
G = graph.data.frame(houseMembersLinks, directed=F, houseMembers)

set.seed(144)
mylayout = layout.auto(G)
par(mar=c(0, 0, 0, 0))
col <- ifelse(V(G)$party == "D", "blue",
              ifelse(V(G)$party == "R", "red", "white"))

plot(G, layout=mylayout, vertex.label=NA, vertex.size=3, vertex.color = col)

### Part-2.b.)

comp = components(G)
comp$membership
table(comp$membership)

in.max.comp = (comp$membership == 1) # this is a vector of TRUE/FALSE
sg = induced_subgraph(G, in.max.comp) # this gives us a new graph object
socialhouseMembers <- houseMembers[in.max.comp,]

sg_mylayout = mylayout[in.max.comp,]
plot(sg, layout=sg_mylayout, vertex.label=NA, vertex.size=3, vertex.color = col )
summary(sg)

### Part-2.h.) Community Detection

set.seed(144)
community = cluster_spinglass(sg, spins = 100)
modularity(community)

clust = community$membership
clust
table(clust)

community$modularity

color = c(brewer.pal(12, "Paired"), "black","white","gray")
plot(sg, layout=sg_mylayout, vertex.label=NA, vertex.size=3,vertex.color=color[clust])
socialhouseMembers$CommunityAssignment = community$membership
```