Lab 4

Naomi Liftman

2023-04-04

Lab 4

I will be using the `civil_rights.csv` data, as the data we are using for our project is two datasets each with around 35,000 edges.

Part 1

Question 1

Load your data, convert it into a network, and plot it. For now, only load rio , igraph, and tidyverse.

```
library(rio)
library(igraph)
library(tidyverse)
```

```
#import data
civ <- import("/Users/naomiliftman/Desktop/GOV338/lab4/civil_rights.csv")
#create a network
civ_net <- graph.data.frame(civ, directed=T)
#plot
plot(civ_net, edge.width=.3,  edge.arrow.size=.3,
     vertex.color="#11876f", vertex.label.color = "black")
```
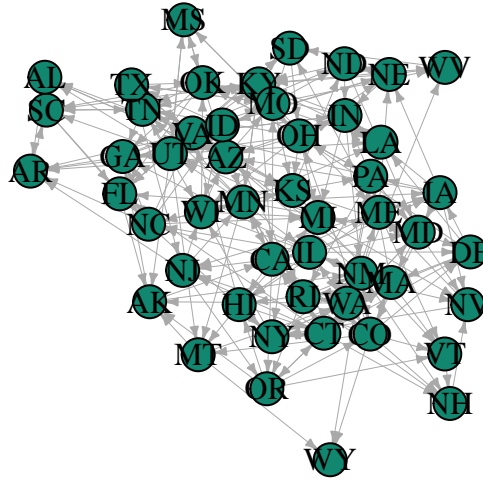
Question 2

```
#providing descriptive statistics
gsize(civ_net)
```

```
## [1] 263
```

```
vcount(civ_net)
```

```
## [1] 50
```

```
summary(degree(civ_net))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3.00    8.00   10.00   10.52   13.00   21.00
```

```
edge_density(civ_net)
```

```
## [1] 0.1073469
```

There are 50 nodes and 263 edges. The minimum is 3, maximum is 21, and mean is 10.52. The density of the network is .107.

Question 3

```
degree(civ_net)
```

```
## VA SC UT AZ FL MT WA OR TN OH WI NM IA OK AL NJ NC GA RI AK CT KY IL MN CA HI
## 14  9 17 14  7  7 13  9 14 15 10 13 12 14  6  9 12  9 11  8 16 12 13 21 16  9
## NY MA MD CO ME NE MI ND PA DE MO IN NV ID TX SD NH VT KS MS LA AR WV WY
##  9 12 12 14 13 10  9  8 12  9 12 12  7  8  9  6  6  6 13  6  7  8  5  3
```

The three nodes with the highest degrees in the data are MN (Minnesota) with 21, UT (Utah) with 17, and then CT (Connecticut) and CA (California) are tied for the third highest degree with 16 each.

Part 2

Question 4

```
library(concoR)
#turn into a matrix
civ_mat <- as.matrix(get.adjacency(civ_net))
#correlation of ties
civ_cor <- data.frame(cor(civ_mat))
civ_cor <- civ_cor[-c(1:1), ]
civ_cor %>%
  summarize(max = max(VA))
```

```
##         max
## 1 0.3831233
```

The node that is most strongly correlated with the first node (VA) is Virginia, as they are the same node and thus have a correlation of 1. Other than that the other node that is most strongly correlated with the first is MI (Minnesota).

Question 5

```
blocks <- concor_hca(list(civ_mat))
blocks %>%
  summarize(m = max(block))
```
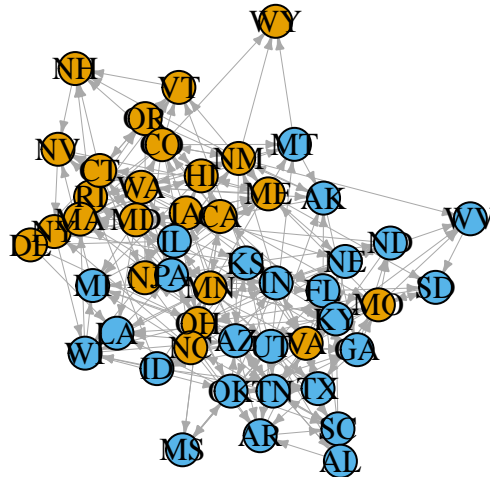
```
##   m
## 1 2
```

Using Concor's hierarchical clustering algorithm we produced two blocks.

Question 6

```
V(civ_net)$blocks <- blocks$block

plot.igraph(civ_net, vertex.color=V(civ_net)$blocks, edge.width=.5, edge.arrow.size=.3,        vertex.lab
```
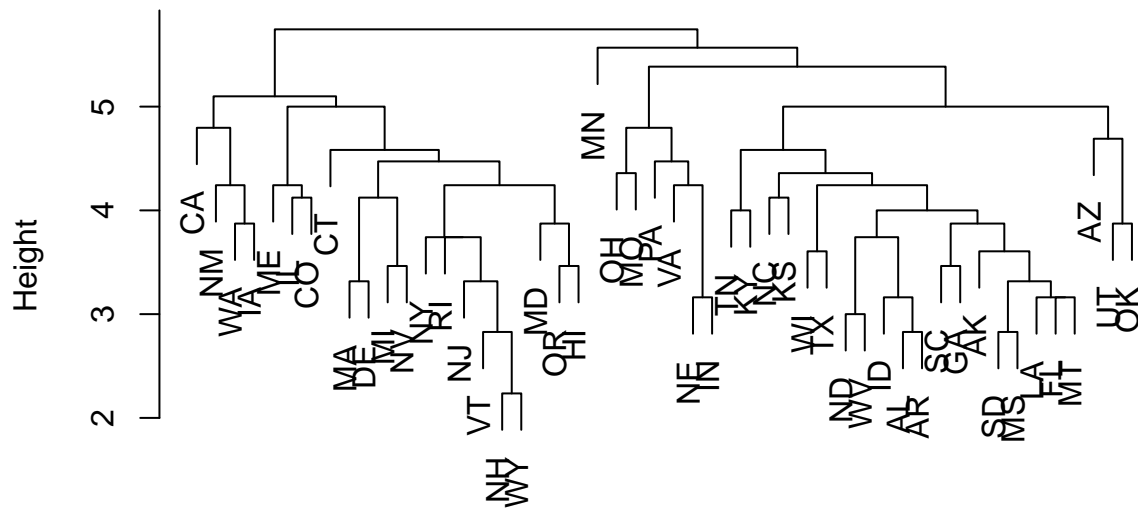

Question 7

Look at the plot, the correlation matrix, and your calculations for node degree. Were you surprised by which groups were put in the same block?

I was surprised by the seperation of the top four highest degrees in the network. Utah is the only one of the top four that is not in the same category with the other groups. I was also surpised that MI and VA are not in the same block even though they have the highest correlation.
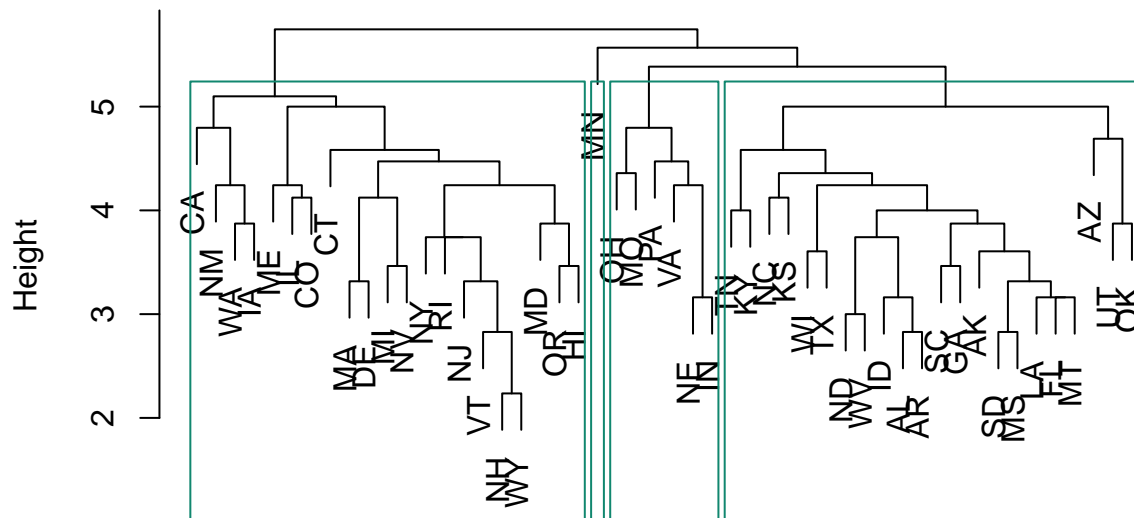
Part 3

Question 8

```
library(sna)
library(network)
ec<-equiv.clust(civ_mat, method="euclidean")
plot(ec, labels=ec$glabels, main="", xlab="", ylab="Height", sub="")
```
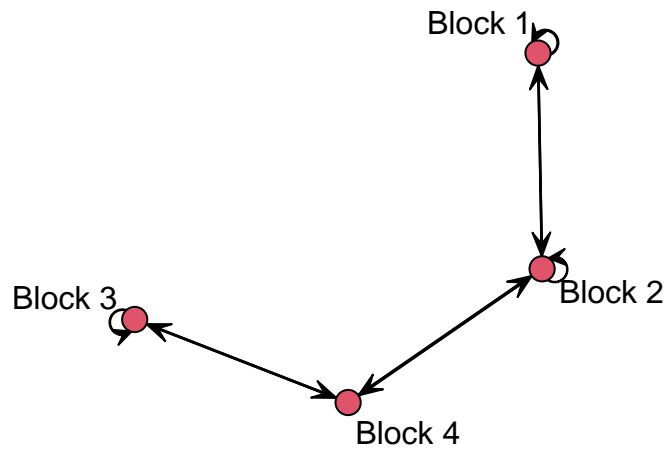
Question 9

I think that 4 blocks would best divide the data, so I will reproduce the plot with four blocks.

```
plot(ec, labels=ec$glabels, main="", xlab="", ylab="Height", sub="")
rect.hclust(ec$cluster,k=4,border="#11876f")
```

Question 10

```
bm<-blockmodel(civ_mat,ec, k=4)
gplot(bm$block.model > gden(civ_mat), diag=T,edge.lwd=bm$block.model*3,  main ="", displaylabels=TRUE)
```

Block 1

Block 3

Block 2

Block 4

Question 11

The first block appears to be many of the wealthier, liberal states such as California, Massachusetts, Oregon, ect. These states are grouped together clearly for ideological reasons, as they all have very similar politics. They do not fall into the same region. Minnesota is the only state in the second block. The third block appears to be mainly part of the Midwest, and a little bit of a hodge-podge of states. The last group is mostly southern right leaning states.