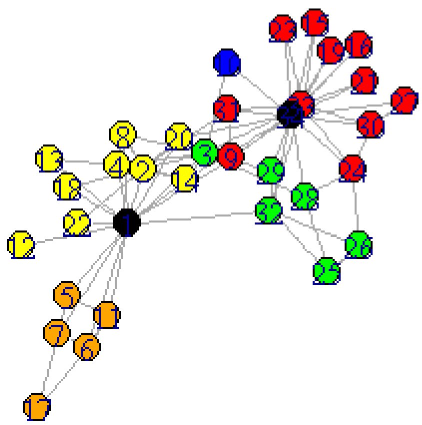
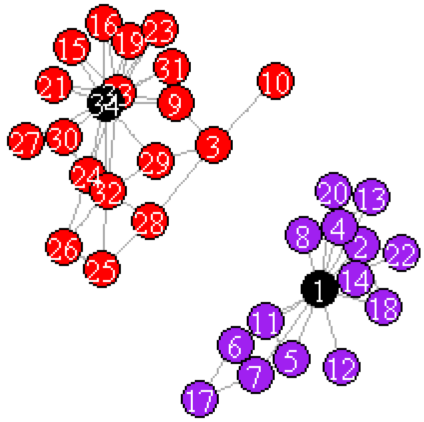
1.We know the result of the Karate Club (Zachary, 1977) split.
Prove or disprove that the result of split could have been predicted
by the weighted graph of social interactions. How well does the
mathematical model represent reality
Generously document your answer with all supporting equations, code,graphs, arguments, etc.

For the initial processing of the karate club graph Rstudio has a helpful function g = graph.famous("Zachary") with the karate club graph built in,
initial graph plot below: 1 and 34 represent the main cluster splits

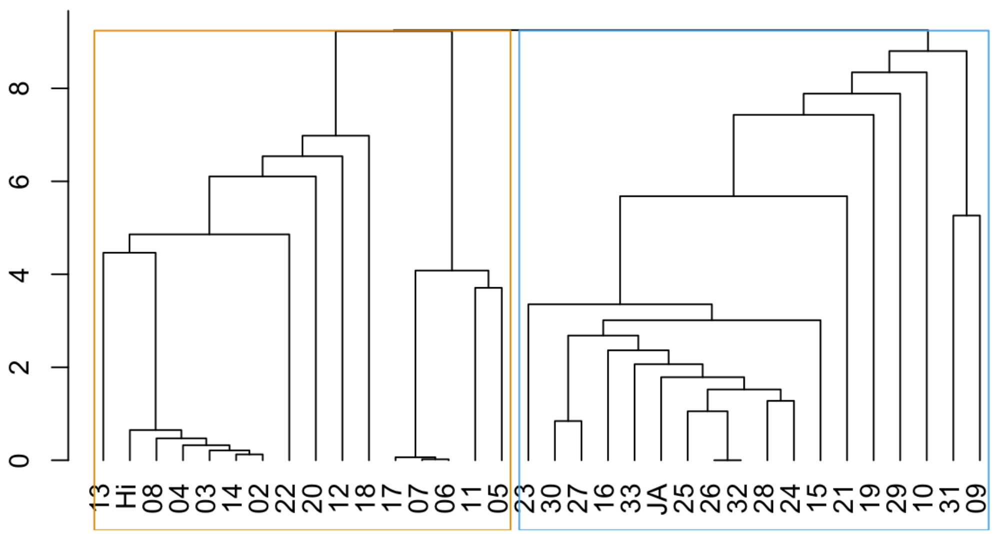


From the powerpoint slides on week 7 numbers 66-71 ,71 in particular I was given a clue on how I believe would need to process the data, the Girvan Newman algorithm, using the removal of edge betweens i could plot a hypothetical splitting of the karate club
And compare it to the actual split to determine if mathematical model represented reality correctly and could therefore be proven predictable
my predicted karate club plot:centers still black

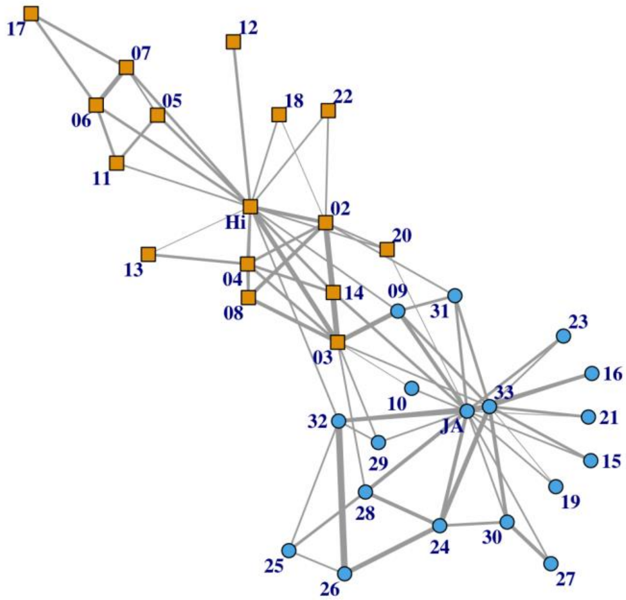


Analysis
Cluster A:1,2,4,5,6,7,8,11,12,13,14,17,18,20,22
15 split in A
Cluster B:3,9,10,15,16,19,21,23,24,25,26,27,28,29,30,31,32,33,34
19 split in B

true karate club split: data taken from “https://rstudio-pubs-static.s3.amazonaws.com/99456_35043d0a235e454787b2665087141914.html”
dendogram:



Correct graph plot:



Analysis:Hi=1 and JA=34
Cluster A:1,2,3,4,5, 6,7,8,9,11, 12,13,14,17,18 ,20,22
17 split in cluster A
Cluster B:10,15,16,19,21 ,23,24,25,26,27, 28,29,30,31,32, 33,34
17 split in Cluster B:

Comparison:
Cluster A:1,2,4,5,6 ,7,8,11,12,13 ,14,17,18,20,22
15 split in A
Cluster A:1,2,3,4,5, 6,7,8,9,11, 12,13,14,17,18 ,20,22
17 split in cluster A

Cluster B:3,9,10,15,16,19,21,23,24,25,26,27,28,29,30,31,32,33,34
19 split in B
Cluster B:10,15,16,19,21 ,23,24,25,26,27, 28,29,30,31,32, 33,34
17 split in Cluster B:

off by 2 values{3,9} in A and 3,9 in B

2 incorrectly placed out of 34 or 94.11% correct

Conclusion
as we can see the major differences are only off by less then 6 giving credence to the claim that the social split could be predicted at least with reliable certainty in this case

code:

```
library(igraph)
#information taken from https://rpubs.com/shestakoff/sna_lab5
g = graph.famous("Zachary")
maximal.cliques(g)
maximal.cliques(g, min = 1, max = 2)
numClusters=5


#now we calculate the edge betweenees,
#slides 93-95 give a good example of why this should work to show group seperation

ebc<-edge.betweenness.community (g, directed = TRUE,
                                 edge.betweenness = TRUE,
                                 merges = TRUE, bridges = TRUE)

#lpc<-label.propagation.community(g)
#fgc<-fastgreedy.community(g)
#lec<-leading.eigenvector.community(g)
#mlc<-multilevel.community(g)
#omc<-optimal.community(g)
#sgc<-spinglass.community(g)
#wtc<-walktrap.community(g)
#ifmc<-infomap.community(g)
#original no split
plot(g)
#3 part split
#plot(fgc,g)
#4 part split
#plot(lpc,g)
#5 part split
#plot(ebc,g)
if(numClusters==2){
V(g)$color="red"
V(g)$color[1]="black"
V(g)$color[34]="black"
}
 if(numClusters==3){
  V(g)$color="red"
  V(g)$color[1]="black"
```

```
  V(g)$color[10]="blue"
  V(g)$color[5:8]="yellow"
  V(g)$color[2]="yellow"
  V(g)$color[17:18]="yellow"
  V(g)$color[11:14]="yellow"
  V(g)$color[4]="yellow"
  V(g)$color[20]="yellow"
  V(g)$color[22]="yellow"
  V(g)$color[34]="black"

 }
if(numClusters==4){
  V(g)$color="red"
  V(g)$color[1]="black"
  V(g)$color[8]="yellow"
  V(g)$color[18]="yellow"
  V(g)$color[4]="yellow"
  V(g)$color[2]="yellow"
  V(g)$color[20]="yellow"
  V(g)$color[22]="yellow"
  V(g)$color[12:14]="yellow"
  V(g)$color[5:7]="orange"
  V(g)$color[11]="orange"
  V(g)$color[17]="orange"
  V(g)$color[10]="blue"
  V(g)$color[34]="black"

  }

if(numClusters==5){
  V(g)$color="red"
  V(g)$color[1]="black"
  V(g)$color[5:7]="orange"
  V(g)$color[10]="blue"
  V(g)$color[8]="yellow"
  V(g)$color[18]="yellow"
  V(g)$color[4]="yellow"
  V(g)$color[2]="yellow"
  V(g)$color[20]="yellow"
  V(g)$color[22]="yellow"
  V(g)$color[12:14]="yellow"
  V(g)$color[11]="orange"
  V(g)$color[17]="orange"
  V(g)$color[25:26]="green"
  V(g)$color[3]="green"
  V(g)$color[28:29]="green"
```

```
  V(g)$color[32]="green"
  V(g)$color[34]="black"

}
```

pulled in part from http://www.sixhat.net/finding-communities-in-networks-with-r-and-igraph.html and from assitance from other class mates

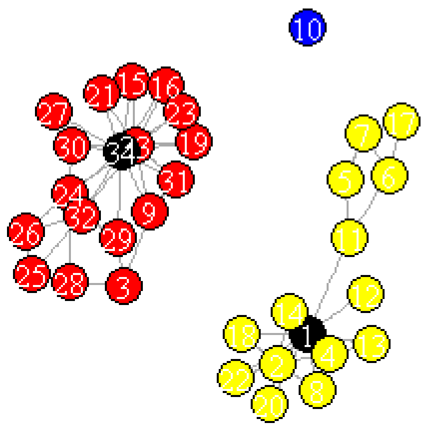
```
mods <- sapply(0:ecount(g), function(i){
  g2 <- delete.edges(g, ebc$removed.edges[seq(length=i)])
  cl <- clusters(g2)$membership
    if(no.clusters(g2)==numClusters){

  plot(g2,layout=layout.fruchterman.reingold,vertex.label.color="white",vertex.size=20)
    }
  }
)
```

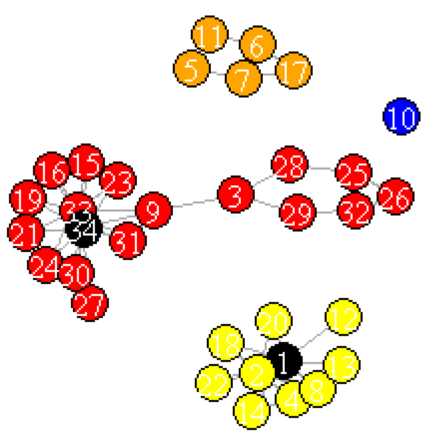
initially with the code I attempted to split the groups using the commented out community functions but that was getting me no where until a class mate recommended the six hat link

2. **We know the group split in two different groups. Suppose the disagreements in the group were more nuanced -- what would the clubs look like if they split into groups of 3, 4, and 5?**

3 cluster split:



4 cluster split:



5cluster split: