2017

Assignment 5

ROSS REELACHART

PROFESSOR MICHAEL L. NELSON

Question 1.)

Before I began trying to prove or disprove the result of the famous Karate Club (Zachary, 1977)[1], I wanted to at least plot the initial state of the Karate Club before the break up. Considering that the study was done 40 years ago, and was itself a very famous study, I assumed that the data was already available in full in some form. This would at least save me the time of re-recording the data by hand.

Fortunately, the program that I planned on using for this assignment, RStudio, already had the data available in one its installable libraries[2]. It required me to install the *igraphdata* library, but that was easily done using the "*install.packages("igraphdata")*" command. After that, it was simply a matter of loading the data up in RStudio and plotting it out.

```
library(igraph)
library(igraphdata)

data(karate)

club <- karate

plot.igraph(club,
  vertex.color="purple",
  vertex.frame.color="#000000",
  vertex.shape="circle",
  vertex.size=15,
  vertex.label.color="#ffffff",
  edge.color="black",
  main="Zachary's Karate Club Graph Before Breakup",
  vertex.label.font=2,
  layout=layout.kamada.kawai,
  vertex.label.cex=1.2
)
```

Figure 1 - The R script "plotPreKarateClub.r" that plots the Karate Club before the break up.

(Continued on next page)

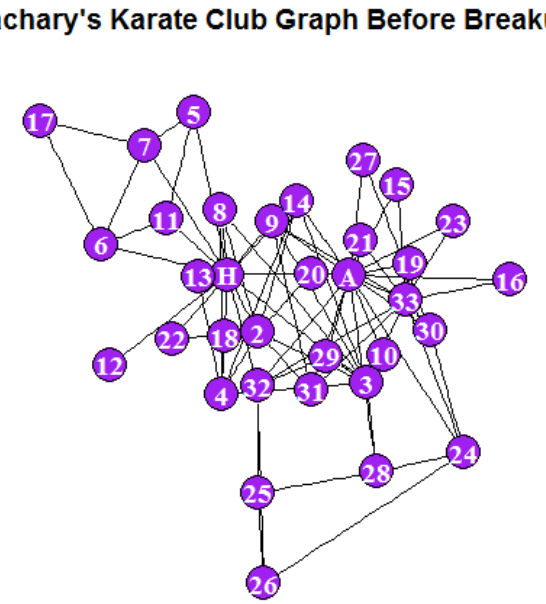

Figure 2 - The actual graph of the Karate Club before the break up. Node "A" is John A., and node "H" is Mr. Hi.

Now that I had an initial plotting of the Karate Club, I had something to compare the post-break up plot against. This would use some of the same code for the initial plotting, specifically the code that creates the visual plot itself. The new additions to the code would be two things: a variable that allows me to control how many post-break up groups I want, and a block of code that replicates the Girven-Newman algorithm. I mostly did this by following the summarized steps in the Wikipedia article[3], and using the "*betweenness"* and *"edge"* functions found in the *igraphdata* library[4]. After that, the Karate Club was plotted again using new colors and a new title.

(Continued on next page)

```r
library(igraph)
library(igraphdata)

data(karate)

club <- karate
#Used for creating the number of groups for the break up. Change to 3,4,5 for extra credit.
grouping <- 2

#Based algorithm off the description given in the Wikipedia article
https://en.wikipedia.org/wiki/Girvan%E2%80%93Newman_algorithm. This while loop handles the first two steps
mostly.
while( clusters(club)$no < grouping ) {
  #1) The betweenness of all existing edges in the network is calculated first.
  edgeBetweenness <- edge.betweenness(club)

  #1.1) Order the collected edges in decreasing order
  decreasingBetweenness <- order(edgeBetweenness, decreasing = TRUE)

  #Credit to: http://stackoverflow.com/questions/652136/how-can-i-remove-an-element-from-a-list for how to
  delete one element from a list.
  #1.2) Get the edge with the highest betweenness from ordered list.
  highestBetweenness <- decreasingBetweenness[-1]

  #1.3) Get the highest betweenness edge from the club.
  edgeToDelete <- get.edge(club, highestBetweenness)

  #2.) Delete the edge from the club.
  club <- delete.edges(club, E(club, P = edgeToDelete))
}

plot.igraph(club,
  vertex.color="green",
  #vertex.color="blue", for 3 grouping
  #vertex.color="red", for 4 grouping
  #vertex.color="pink", for 5 grouping
  vertex.frame.color="#000000",
  vertex.shape="circle",
  vertex.size=15,
  vertex.label.color="#ffffff",
  edge.color="black",
  main="Zachary's Karate Club After Breakup (5 Groups)",
  vertex.label.font=2,
  layout=layout.kamada.kawai,
  vertex.label.cex=1.2
  )
```

Figure 3 - The R script "plotPostKarateClub.r" that performs the Givan-Newman algorithm to replicate the results of the actual Karate Club study, then plots the graph. Includes code that can be used for more than 2 clusters.

(Continued on next page)

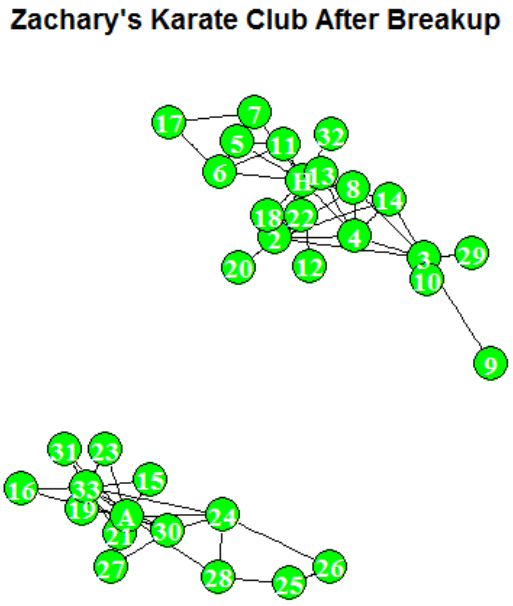

Figure 4 - The actual plot of the Karate Club break up given the Girvan-Newman algorithm.

Now, according to the original paper, Zachary's break up prediction was nearly 100 percent accurate to the actual break up. The only mistake was that individual 9 would end up joining Mr. Hi's group, as opposed to the predicted membership to John A's group. With the Girvan-Newman algorithm in my code, I also had a nearly accurate prediction of the break up. The algorithm correctly placed individual 9, but it incorrectly placed individuals 10 and 32.

So I can confidently say that the result of Zachary's Karate Club has been proven using the Girvan-Newman algorithm with an approximate 94 percent accuracy.

(Continued on next page)

Question 2.)

The ability to create more than 2 clusters was built into the original R script that created the 2 clusters. By changing the *grouping* variable to 3, 4 or 5 I can make the while loop algorithm create 3, 4 or 5 clusters. I then also changed the colors of the nodes for variety's sake.

In every case, John A. and Mr. Hi held onto the largest clusters, while the other clusters tended to be groups of four individuals whom I suppose had slightly stronger connections with each other than the rest of the group. This also seems more realistic because personal friend groups, in my experience, tend to max out around four people.

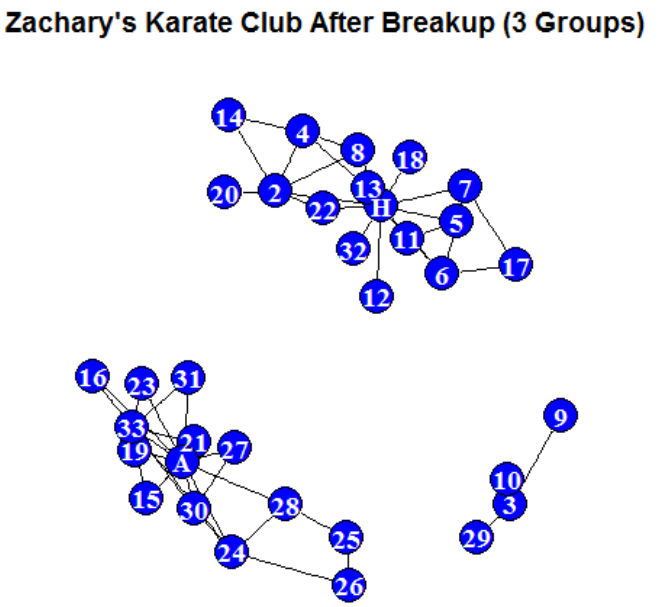

Figure 5 - The resultant groups of the Karate Club break up if there were 3 groups.

(Continued on next page)

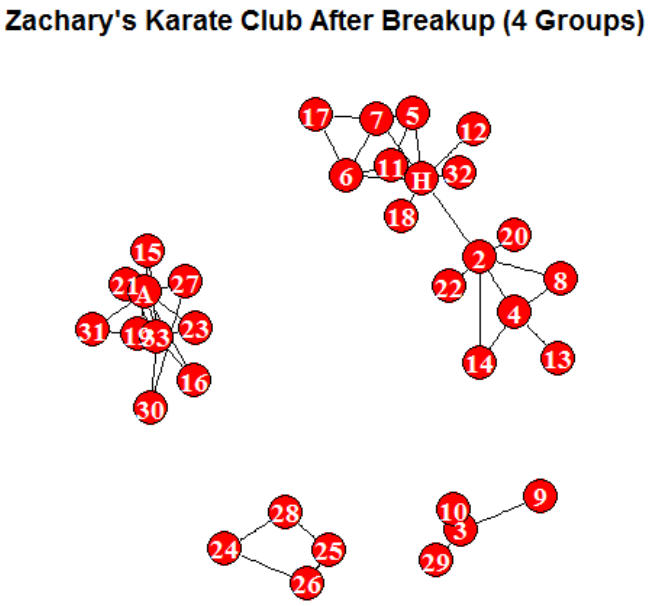

Figure 6 - The resultant groups of the Karate Club break up if there were 4 groups.

(Continued on next page)

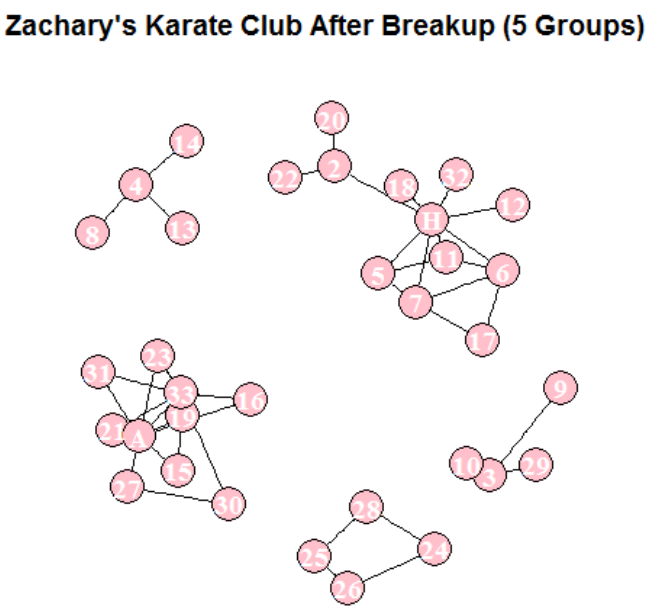

Figure 7 - The resultant groups of the Karate Club break up if there were 5 groups.

References

1.) http://aris.ss.uci.edu/~lin/76.pdf
2.) https://cran.r-project.org/web/packages/igraphdata/igraphdata.pdf
3.) https://en.wikipedia.org/wiki/Girvan%E2%80%93Newman_algorithm
4.) http://igraph.org/r/doc/betweenness.html