

Assignment 5 Report

Professor Nelson

By Tyler Medina

3/16/17

Part 1

1.1 Using the Girvan-Newman Algorithm

To use the Girvan-Newman algorithm, I created an R program that determines the edge betweenness of the nodes and then calculates which nodes have the highest edge betweenness. R studio has a library that has the karate club data available called “igraph”. The program calculates the edge betweenness of the vertices and sorts the data into order. This makes it easy to pull and delete the edges with the highest edge betweenness.

```
library(igraph)
library(igraphdata)

data(karate)
k_network<-karate

while (clusters(k_network)$no < 2){

  edgeBetweenness <- edge.betweenness(k_network)
  sort <- order(-edgeBetweenness)
  highestEB <- sort[-1]
  deleteEdge <- get.edge(k_network, highestEB)
  k_network <- delete.edges(k_network,E(k_network,deleteEdge))
}

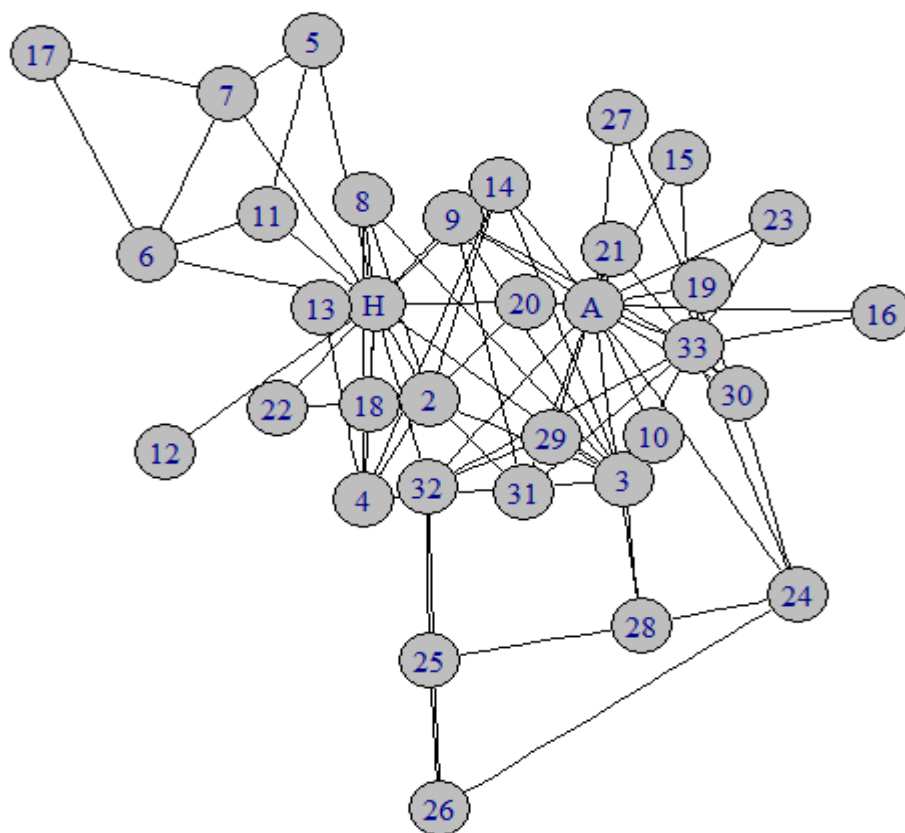
plot.igraph(k_network, vertex.color="purple",vertex.shape="circle",vertex.label=
  edge.color="black",main="Karate Club (2 group split)",
  layout=layout.kamada.kawai)
```

1.2 Graphing the Data

The following two graphs show the relationships between people in the karate club before the split and then after the split respectively.

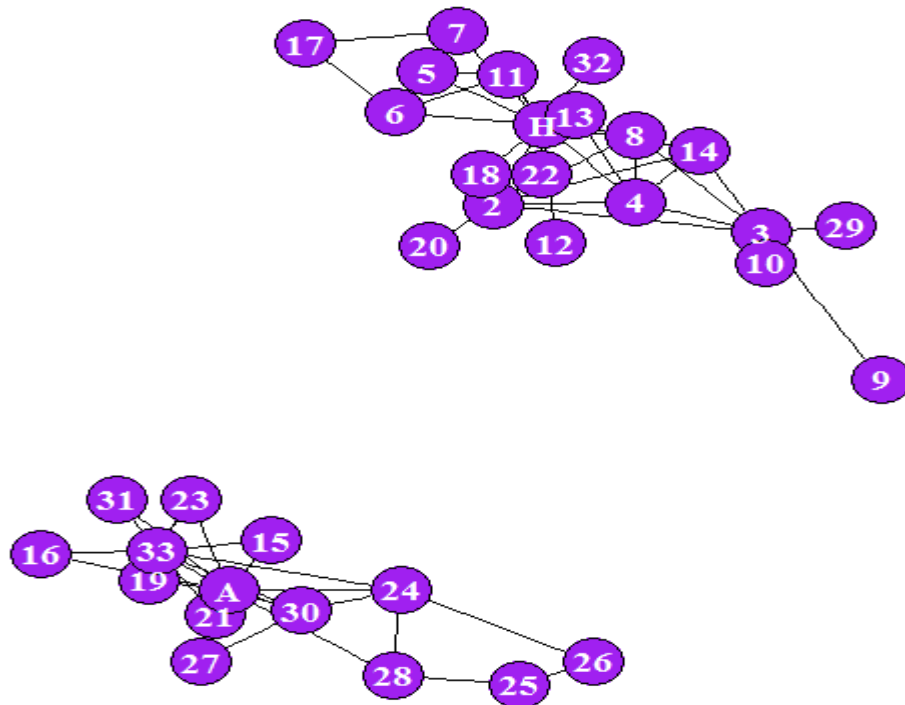
This space was intentionally left blank.

Zachary's Karate Club Network



This space was intentionally left blank.

Karate Club (2 group split)



1.3 Interpreting the Data

With the data in the graphs, we can compare with the results of Zachary's prediction to assess the accuracy of the Girvan-Newman algorithm. It's to be noted that individual 1 and 34 are Mr. Hi and John A. respectively.

This space was intentionally left blank.

Individual	Zachary's Model Faction	Girvan-Newman Model Faction
1	Mr. Hi	Mr. Hi
2	Mr. Hi	Mr. Hi
3	Mr. Hi	Mr. Hi
4	Mr. Hi	Mr. Hi
5	Mr. Hi	Mr. Hi
6	Mr. Hi	Mr. Hi
7	Mr. Hi	Mr. Hi
8	Mr. Hi	Mr. Hi
9	Mr. Hi	Mr. Hi
10	John	Mr. Hi
11	Mr. Hi	Mr. Hi
12	Mr. Hi	Mr. Hi
13	Mr. Hi	Mr. Hi
14	Mr. Hi	Mr. Hi
15	John	John
16	John	John
17	Mr. Hi	Mr. Hi
18	Mr. Hi	Mr. Hi
19	John	John
20	Mr. Hi	Mr. Hi
21	John	John
22	Mr. Hi	Mr. Hi
23	John	John
24	John	John
25	John	John
26	John	John
27	John	John
28	John	John
29	John	Mr. Hi
30	John	John
31	John	John
32	John	Mr. Hi
33	John	John
34	John	John

Based on the table above, it can be concluded that the Girvan-Newman algorithm is accurate in predicting the split of the Karate club. Individuals 10, 29, and 32 were placed in the wrong factions. This is a hit rate of $31/34$ which is about 91 percent. This is a very favorable hit percentage, but with a small data set of only 34 individuals, we may want to aim for a higher hit percentage. It is important to note that all the misses were classified into Mr. Hi's faction. The similarity between individuals 10 and 29 are that they are fringe members with no direct edge to Mr. Hi. They are connected to other individuals, who themselves are distant from Mr. Hi's node. Individual 32 has a direct link to Mr. Hi, but is supposed to be in John's faction. This seems to likely be an outlier. They may potentially have a relationship to Mr. Hi and John that is significantly different from the other members, and that could possibly make it hard to map them into the weighted graph.

References

Gabor, C. Package 'igraphdata'. July 13, 2015. Version 1.01. Available at:

<https://cran.r-project.org/web/packages/igraphdata/igraphdata.pdf>

Zachary, W. An information Flow Model for Conflict and Fission in Small Groups. 1977.

Vol 3. Pgs. 452 – 477. Available at: <http://aris.ss.uci.edu/~lin/76.pdf>

Kabacoff, R. Sorting Data. Quick R. 2017. Available at:

<http://www.statmethods.net/management/sorting.html>

Connected Components of a Graph. Igraph. 2015. Available at:

<http://igraph.org/r/doc/components.html>