# Assignment 6

CS 595: Introduction to Web Science

Fall 2013

Shawn M. Jones

Finished on October 31, 2013

# 1

## Question

```
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.

Useful sources include:

* Original paper

http://aris.ss.uci.edu/~lin/76.pdf

* Slides

http://www-personal.umich.edu/~ladamic/courses/networks/si614w06/ppt/lecture18.ppt

http://clair.si.umich.edu/si767/papers/Week03/Community/CommunityDetection.pptx

* Code and data

http://networkx.github.io/documentation/latest/examples/graph/karate_club.html

http://nbviewer.ipython.org/url/courses.cit.cornell.edu/info6010/resources/11notes.ipynb

http://stackoverflow.com/questions/9471906/what-are-the-differences-between-community-detection-algorithms-in-igraph/9478989#9478989

http://stackoverflow.com/questions/5822265/are-there-implementations-of-algorithms-for-community-detection-in-graphs

http://konect.uni-koblenz.de/networks/ucidata-zachary

http://vlado.fmf.uni-lj.si/pub/networks/data/ucinet/ucidata.htm#zachary
```

## Answer

The results of the Karate Club split could have been predicted using the weighted graph of social interactions. The level of accuracy depends on the algorithm used.

The weighted graph of Karate Club data was loaded into R and run through the program shown in Listing 1. Lines 24-44 use the hierarchical divisive Girvan-Newman Betweenness Clustering algorithm to split the graph, stopping when two clusters are created. From this code, the graphs in Figures 1 and 2 were produced.

Using this implementation, we can create a split in the group, and even predict that there will be a split. The algorithm goes through 18 iterations before two groups are created.

Figure 1 shows the graph of the Karate Club's relationships prior to the split. As with Zachary's original paper, the node labeled 34 represents "John A." and the node labeled 1 represents "Mr. Hi".

Figure 2 shows the graph of the Karate Club's relationships after running the Girvan-Newman Betweenness Clustering Algorithm from Listing 1.

Table 1 shows the individual identifiers in column 1. Column 2 shows the actual recorded group membership after the split (*Officers'* is John A's faction). Column 3 shows Zachary's modeled group membership after the split. Column 4 shows my Girvan-Newman modeled group membership after the split. Column 5 shows whether my Girvan-Newman implementation resulted in a *Hit* (correctly calculated membership) or *Miss* (incorrectly calculated membership).

Zachary's Ford and Fulkerson procedure had a 97% success rate. My Girvan-Newman implementation has a 94% success rate, making it inferior in this case but still effective at predicting almost all of the group memberships. My implementation also predicted that individual 9 would stay with Mr. Hi, which is the one membership that Zachary missed.

Either way, it can be shown that this split could have been predicted to greater than 90% accuracy using this data.

One item that Zachary's paper did not take into account was when this split could have been predicted. A possibility for future research is to build snapshot graphs for a given group *over time*, and then using the Ford and Fulkerson procedure or the Girvan-Newman algorithm to find the earliest point at which a split can be predicted. Is it possible that the split could have been accurately predicted when one or more of the edge weights was lower (e.g. an individual hadn't attended a karate tournament *yet*, but will gain an edge point for doing so in a subsequent snapshot graph)?

| Individual | Actual Group Membership From Split | Zachary's Ford and Fulkerson Procedure Modeled Group Membership From Split | Girvan-Newman Modeled Group Membership From Split | Hit/Miss For Girvan-Newman |
|---|---|---|---|---|
| 1 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 2 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 3 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 4 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 5 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 6 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 7 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 8 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 9 | Mr. Hi | Officers' | Mr. Hi | Hit |
| 10 | Officers' | Officers' | Mr. Hi | Miss |
| 11 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 12 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 13 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 14 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 15 | Officers' | Officers' | Officers' | Hit |
| 16 | Officers' | Officers' | Officers' | Hit |
| 17 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 18 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 19 | Officers' | Officers' | Officers' | Hit |
| 20 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 21 | Officers' | Officers' | Officers' | Hit |
| 22 | Mr. Hi | Mr. Hi | Mr. Hi | Hit |
| 23 | Officers' | Officers' | Officers' | Hit |
| 24 | Officers' | Officers' | Officers' | Hit |
| 25 | Officers' | Officers' | Officers' | Hit |
| 26 | Officers' | Officers' | Officers' | Hit |
| 27 | Officers' | Officers' | Officers' | Hit |
| 28 | Officers' | Officers' | Officers' | Hit |
| 29 | Officers' | Officers' | Officers' | Hit |
| 30 | Officers' | Officers' | Officers' | Hit |
| 31 | Officers' | Officers' | Officers' | Hit |
| 32 | Officers' | Officers' | Mr. Hi | Miss |
| 33 | Officers' | Officers' | Officers' | Hit |
| 34 | Officers' | Officers' | Officers' | Hit |

Table 1: Results of Split, as predicted by Girvan-Newman Implementation

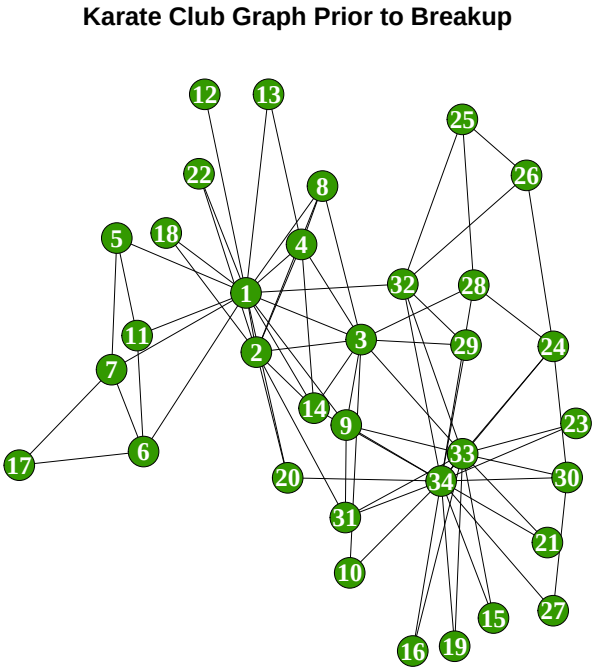


Figure 1: Karate Club Graph Before Split

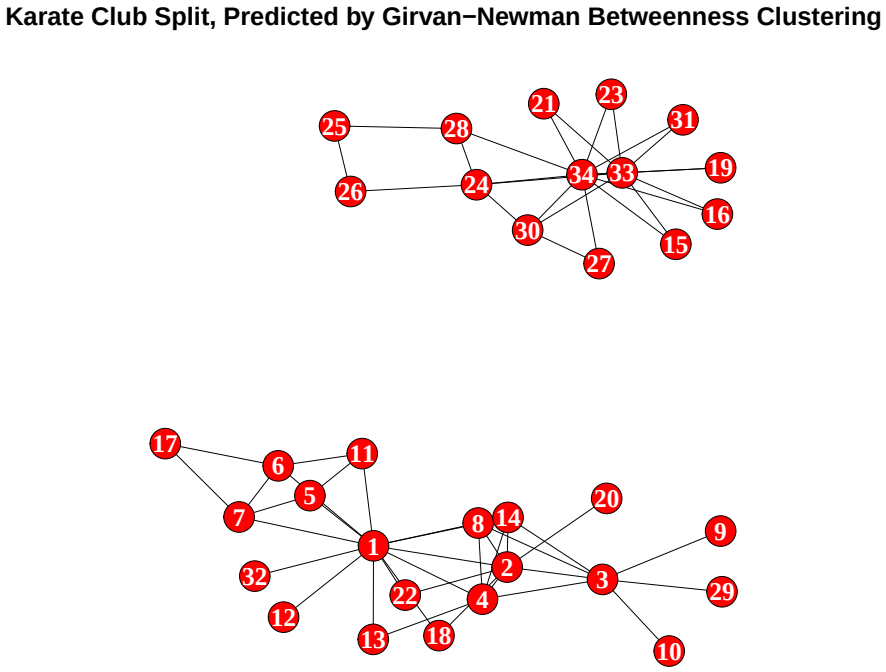


Figure 2: Karate Club Graph Split Predicted by Girvan & Newman Betweenness Clustering

```
1  library(igraph) # for graph functions
2  library(igraphdata) # for karate club data
3
4  data(karate)
5
6  club <- karate
7  threshold <- 2
8
9  # graph prior to the breakup
10 plot.igraph(club,
11   vertex.color="#339900",
12   vertex.frame.color="#000000",
13   vertex.shape="circle",
14   vertex.size=11,
15   vertex.label.color="#ffffff",
16   edge.color="black",
17   main="Karate Club Graph Prior to Breakup",
18   vertex.label.font=2,
19   layout=layout.kamada.kawai,
20   vertex.label.cex=1.2
21 )
22
23 # Algorithm below graciously provided by Corren McCoy
24 # also described by University of Michigan
25 while( clusters(club)$no < threshold ) {
26   # calculate betweenness of all edges
27   club.edge.betweenness <- edge.betweenness(club)
28
29   # remove edge with highest betweenness
30   decreasing.betweenness <- order(club.edge.betweenness,
         decreasing = TRUE)
31
32   # for fun, pick off the ones with the lowest betweenness,
33   # and watch the club kick out one member at a time
34   #decreasing.betweenness <- order(decreasing.betweenness)
35
36   # get the one with the highest edge betweenness
37   # Thanks Stack Overflow:
38   # http://stackoverflow.com/questions/652136/how-can-i-remove-
         an-element-from-a-list
39   highest.betweenness <- decreasing.betweenness[-1]
40
41   # acquire the edge to delete
42   edge.to.delete <- get.edge(club, highest.betweenness)
43
44   # get rid of it
45   club <- delete.edges(club, E(club, P = edge.to.delete))
46 }
```

```
47
48 plot.igraph(club,
49   vertex.color="#ff0000",
50   vertex.frame.color="#000000",
51   vertex.shape="circle",
52   vertex.size=11,
53   vertex.label.color="#ffffff",
54   edge.color="black",
55   main="Karate Club Split, Predicted by Girvan-Newman
         Betweenness Clustering",
56   vertex.label.font=2,
57   layout=layout.kamada.kawai,
58   vertex.label.cex=1.2
59   )
```

Listing 1: R program for Girvan & Newman Betweenness Clustering shown in Figures 1 and 2

## 2

### Question

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?

### Answer

The same R script shown in Listing 1 was run to produce Figures 3, 4, and 5. The value of the variable *threshold* on line 7 was changed to 3, 4, or 5 as needed to create the split. In all cases, "Mr. Hi" and "John A." held on to the largest groups even after the graph was split into 5 groups.

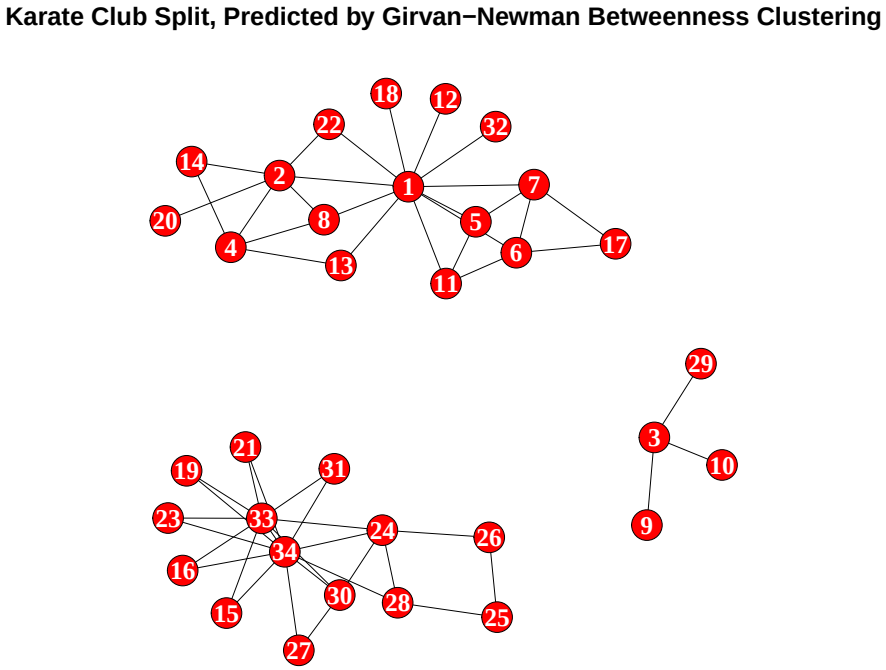


Figure 3: Karate Club Graph Split Into 3 Groups Predicted by Girvan & Newman Betweenness Clustering

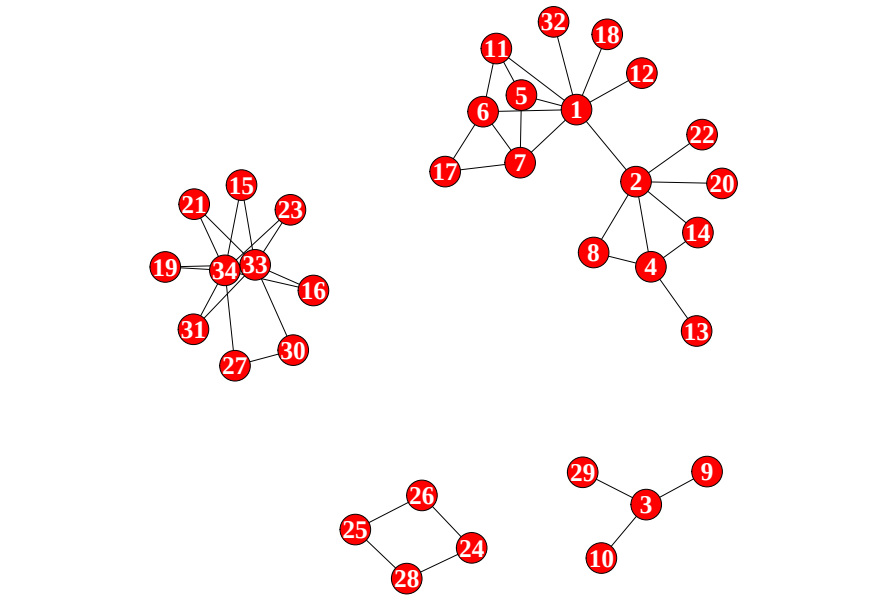


Figure 4: Karate Club Graph Split Into 4 Groups Predicted by Girvan & Newman Betweenness Clustering

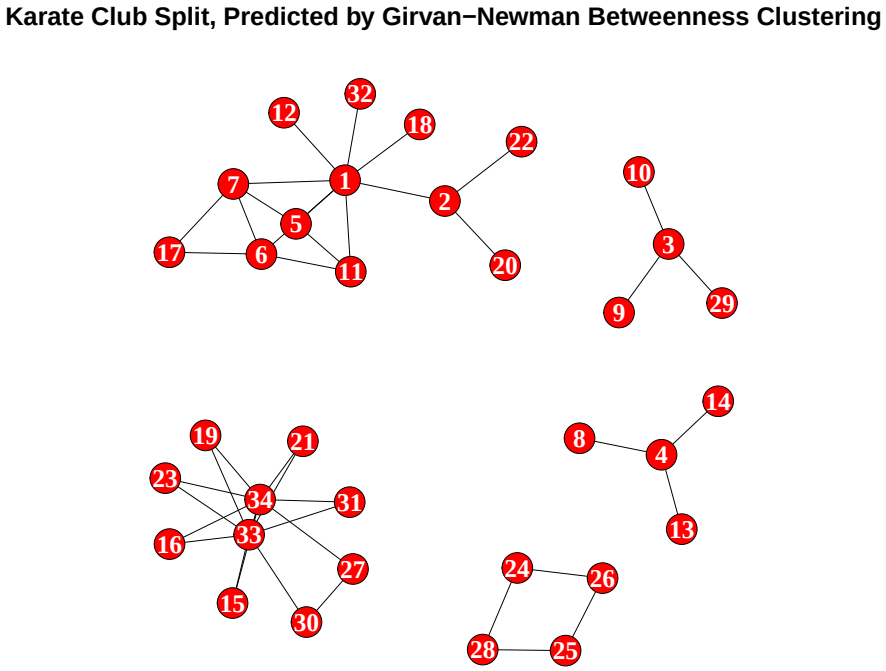


Figure 5: Karate Club Graph Split Into 5 Groups Predicted by Girvan & Newman Betweenness Clustering