CS 595: Assignment 6

Mallika Kogatam

Fall 2014

Contents

1 Problem 1

Question

```
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-algorith

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

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

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

Answer

1. I had hard time understanding the concept of Karate Club. But after lot of research and reading the paper again and again I had a clear picture.
2. I figured that using R is a best when compared to python. There are different data sets available for Karate Club online. I downloaded "karate.gml" file and started working with that.
3. I loaded the "karate.gml" file into R and run through the program shown in Listing 1 at line 5. Lines 16-29 use the Girvan-Newman Betweenness Clustering Algorithm to slit the graph into various clusters.
4. Using this implementation, I can split the group into as many clusters as I need by changing the threshold value. This algorithm goes through 16 iterations before two groups are created. From this code Figures 1 and 2 were produced.
5. Figure 1 shows the graph for the Karate Club's relationship prior to the split. According to Zachary's original paper, the node labeled 1 represents "Mr.Hi" and the node labeled 34 represent "John A".
6. Figure 2 shows the graph of the Karate Club's relations after running the Girvan-Newman Betweenness Clustering Algorithm for Listing1
7. After the graph is produced I wanted to verify it so I did a little search online and realized that the weighted graph of karate Club data is available as predefined from the package "igraphdata".
8. So I loaded this data into R and run through the program shown in Listing 1 instead of loading it form the "karate.gml" file.
9. To my surprise the algorithm I used goes through 18 iterations before two groups are created. From this data set Figure 3 and 4 were produced.
10. I checked with the code I wrote for the algorithm again and did a little more research and failed to find the reason behind different solutions.
11. Figure3 shows the graph of the Karate Club's relation prior to the split and Figure4 shows the graph after the split.
12. Table 1 shows the results compared with Zachary's original predictions and the actual data. Column 5 shows whether my Girvan-Newman Algorithm implementation resulted in a *Hit* or *Miss* for the data set loaded from the "igraphdata", where as column 7 shows for the data set that is loaded from "karate.gml" file.
13. Zachary's Ford and Fulkerson procedure had a $\frac{33}{34} = 97\%$ success rate.
14. My Girvan-Newman implementation has a $\frac{32}{34} = 94\%$ success rate for the data which loaded directly form the "igraphdata" package in R.
15. On the other had My Girvan-Newman implementation has a $\frac{31}{34} = 91\%$ success rate for the data which is loaded from the "karate.gml" file.
16. So I preferred using the data that is loaded from the "igraphdata" for my future research.
17. My Girvan-Newman is 94% success where as Zachary's Ford and Fulkerson is 97%, but my Girvan-Newman is still effective at predicting almost all of the group memberships.
18. My implementation also predicted that individual 9 would stay with Mr.Hi, which is missed by Zachary.
19. Either way, it can be proved that this split could have been predicted to greater the 90% accuracy using this data and these algorithms.
20. In Zachary's paper he did not consider when the split could have been predicted.
21. But by using the Girvan-Newman Algorithm we can find the earliest point at which a split can be predicted. And that can be done by looking into the number of iterations.
22. So by this we can say that the mathematical model (Girvan-Newman Algorithm) is 94% success, so it's pretty much close to the reality.

```
1  library(igraph) # for graph functions
2  library(igraphdata) # for karate club data
3
4  data(karate)# loading the weighted graph for karate club from the packages.
5  #karate <- read.graph("Z:/mallika/cs595/A6/karate/karate.gml",format = "gml")
6  #Used the above line to load karate.gml file
7  club <- karate
8  threshold <- 2
9
10
11 plot.igraph(club,
12             main="Karate Club Graph Prior to Breakup",
13             )
14 i <- 0
15
16 while( clusters(club)$no < threshold ) {
17   #calculate the betweenness between edges by using the function edge.betweenness
18   club.edge.betweenness <- edge.betweenness(club)
19   #Ordering the edges in descresing order basing on their betweenness values.
20   decreasing.betweenness <- order(club.edge.betweenness, decreasing = TRUE)
21   #collect the node which has the highest betweenness and stored it in a variable.
22   highest.betweenness <- decreasing.betweenness[-1]
23
24   edge.to.delete <- get.edge(club, highest.betweenness)
25   cat(i)# check the number of itteration it took to slipt it into 2 clusters
26   club <- delete.edges(club, E(club, P = edge.to.delete))
27   #Delete the edge which had the highest betweenness
28   ++i
29 }
30
31
32 plot.igraph(club,
33              main="Karate Club Split, Predicted by Girvan-Newman Betweenness",
34              )
```

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

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	Girvan-Newman Modeled Group Membership From Split ("karate.gml")	Hit/Miss For Girvan-Newman
1	Mr. Hi	Mr. Hi	Mr. Hi	Hit	Mr. Hi	Hit
2	Mr. Hi	Mr. Hi	Mr. Hi	Hit	Mr. Hi	Hit
3	Mr. Hi	Mr. Hi	Mr. Hi	Hit	John A	Miss
4	Mr. Hi	Mr. Hi	Mr. Hi	Hit	Mr. Hi	Hit
5	Mr. Hi	Mr. Hi	Mr. Hi	Hit	Mr. Hi	Hit
6	Mr. Hi	Mr. Hi	Mr. Hi	Hit	Mr. Hi	Hit
7	Mr. Hi	Mr. Hi	Mr. Hi	Hit	Mr. Hi	Hit
8	Mr. Hi	Mr. Hi	Mr. Hi	Hit	Mr. Hi	Hit
9	Mr. Hi	John A	Mr. Hi	Hit	John A	Miss
10	John A	John A	Mr. Hi	Miss	John A	Hit
11	Mr. Hi	Mr. Hi	Mr. Hi	Hit	Mr. Hi	Hit
12	Mr. Hi	Mr. Hi	Mr. Hi	Hit	Mr. Hi	Hit
13	Mr. Hi	Mr. Hi	Mr. Hi	Hit	Mr. Hi	Hit
14	Mr. Hi	Mr. Hi	Mr. Hi	Hit	John A	Miss
15	John A	John A	John A	Hit	John A	Hit
16	John A	John A	John A	Hit	John A	Hit
17	Mr. Hi	Mr. Hi	Mr. Hi	Hit	Mr. Hi	Hit
18	Mr. Hi	Mr. Hi	Mr. Hi	Hit	Mr. Hi	Hit
19	John A	John A	John A	Hit	John A	Hit
20	Mr. Hi	Mr. Hi	Mr. Hi	Hit	Mr. Hi	Hit
21	John A	John A	John A	Hit	John A	Hit
22	Mr. Hi	Mr. Hi	Mr. Hi	Hit	Mr. Hi	Hit
23	John A	John A	John A	Hit	John A	Hit
24	John A	John A	John A	Hit	John A	Hit
25	John A	John A	John A	Hit	John A	Hit
26	John A	John A	John A	Hit	John A	Hit
27	John A	John A	John A	Hit	John A	Hit
28	John A	John A	John A	Hit	John A	Hit
29	John A	John A	John A	Hit	John A	Hit
30	John A	John A	John A	Hit	John A	Hit
31	John A	John A	John A	Hit	John A	Hit
32	John A	John A	Mr. Hi	Miss	John A	Hit
33	John A	John A	John A	Hit	John A	Hit
34	John A	John A	John A	Hit	John A	Hit

Table 1: Results of Split, as predicted by my Girvan-Newman Implementation and also compared to Zachary's predictions and the actual data

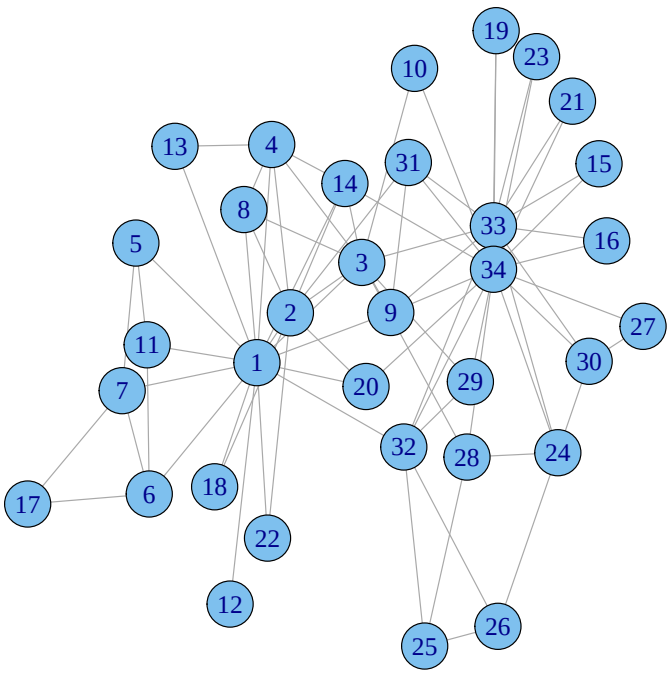

Figure 1: Karate Club Graph prior to split for "karate.gml"

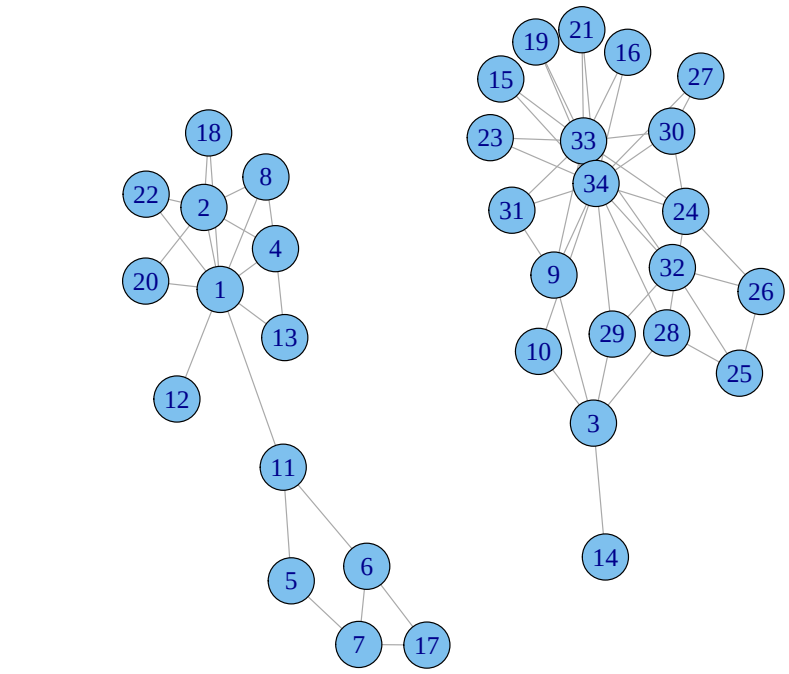

Figure 2: Karate Club Graph Split Into 2 Groups Predicted by Girvan & Newman Betweenness Clustering for "karate.gml"

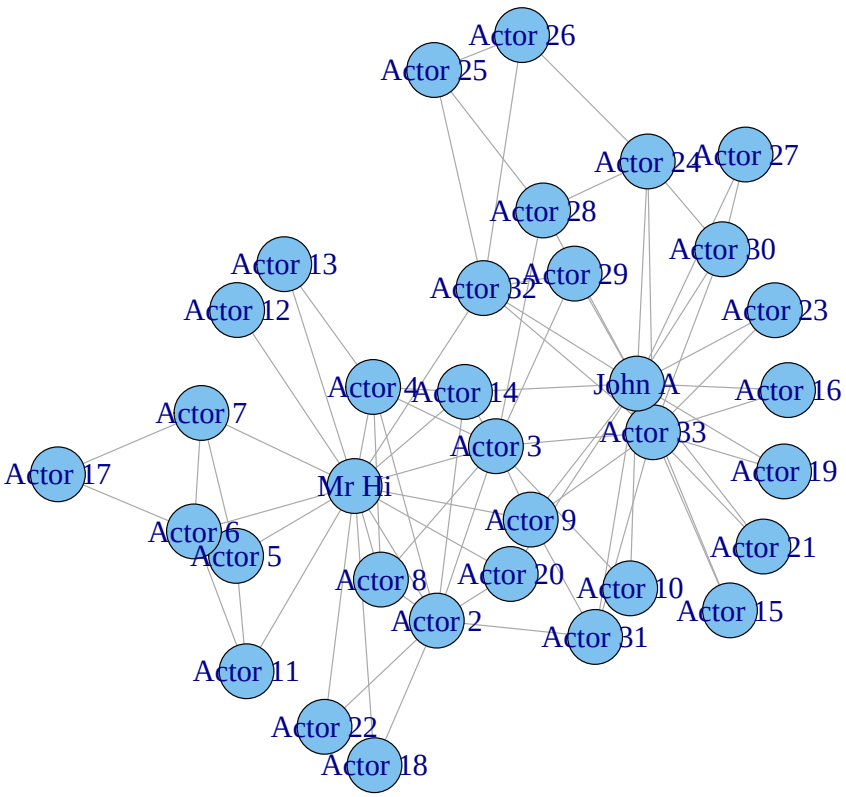

Figure 3: Karate Club Graph prior to split

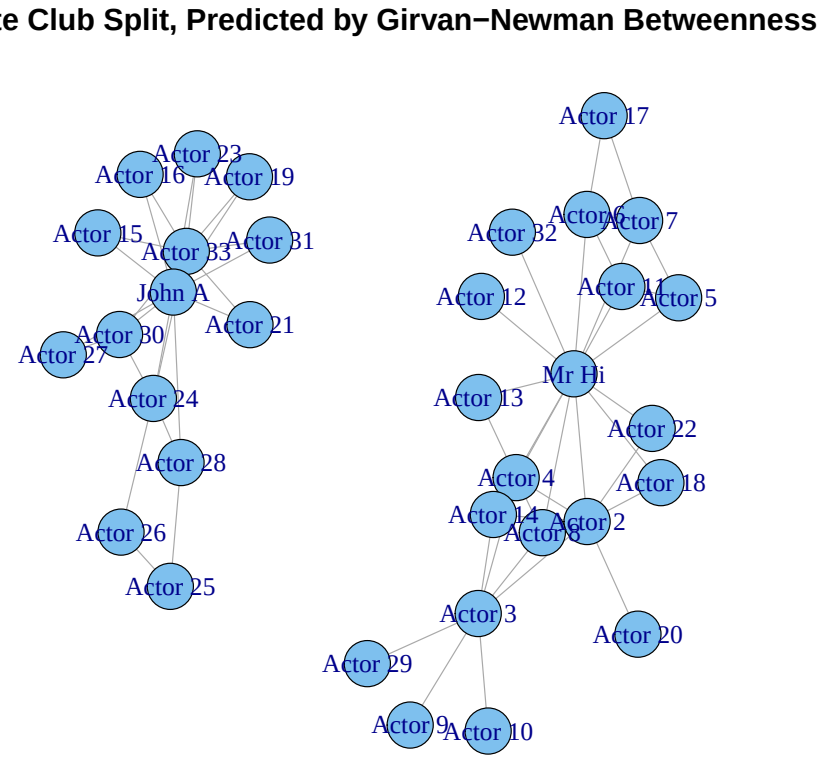

Figure 4: Karate Club Graph Split Into 2 Groups Predicted by Girvan & Newman Betweenness Clustering

2 Problem 2

Question

```
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?
```

1. I used the same R script shown in Listing1 to produce Figure5, 6 and 7. The value of the *threshold* is changed to 3, 4 or 5 as required to create the split. Even after spliting the karate club into 3, 4 or 5 groups Mr.Hi and John A hold the biggest groups.

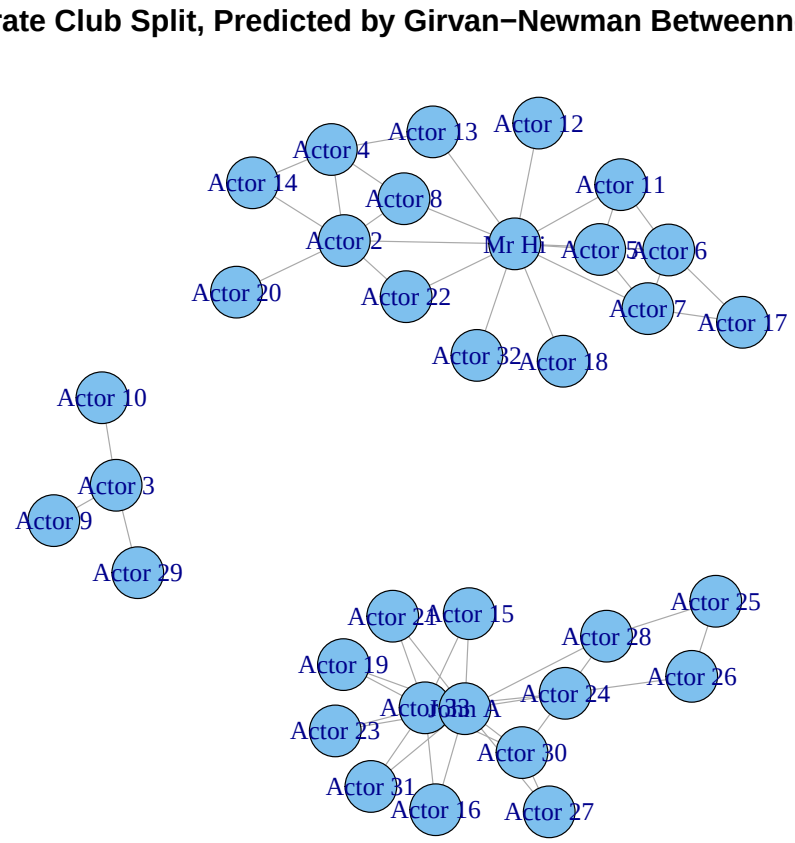

Figure 5: Karate Club Graph Split Into 3 Groups Predicted by Girvan & Newman Betweenness Clustering

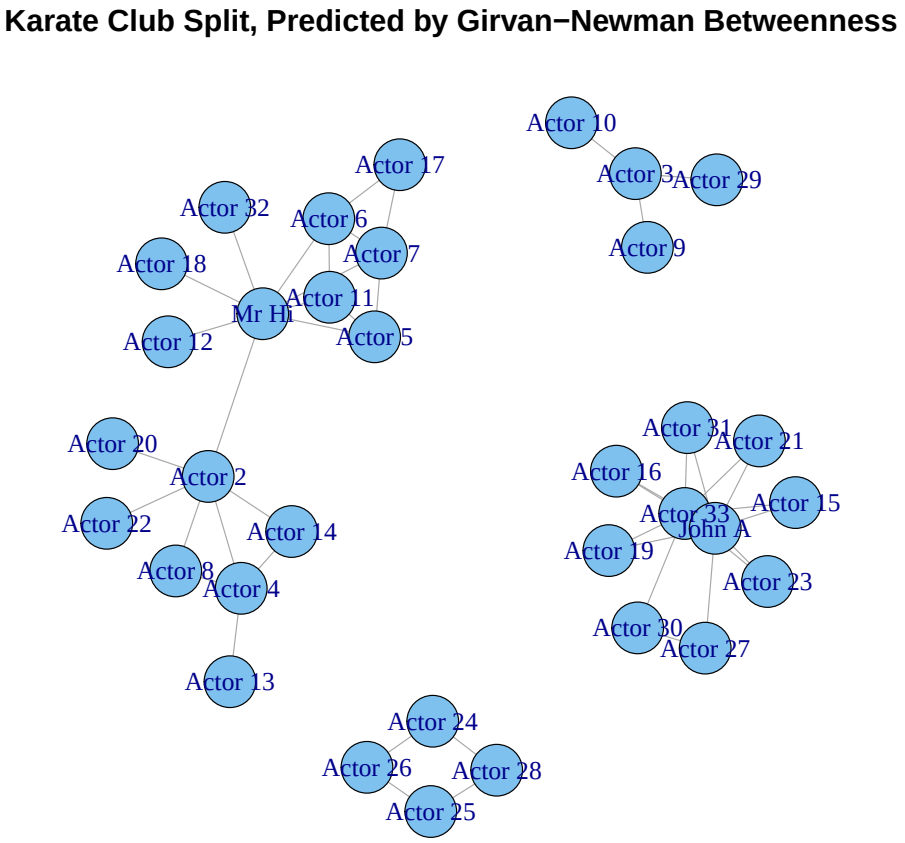

Figure 6: Karate Club Graph Split Into 4 Groups Predicted by Girvan & Newman Betweenness Clustering

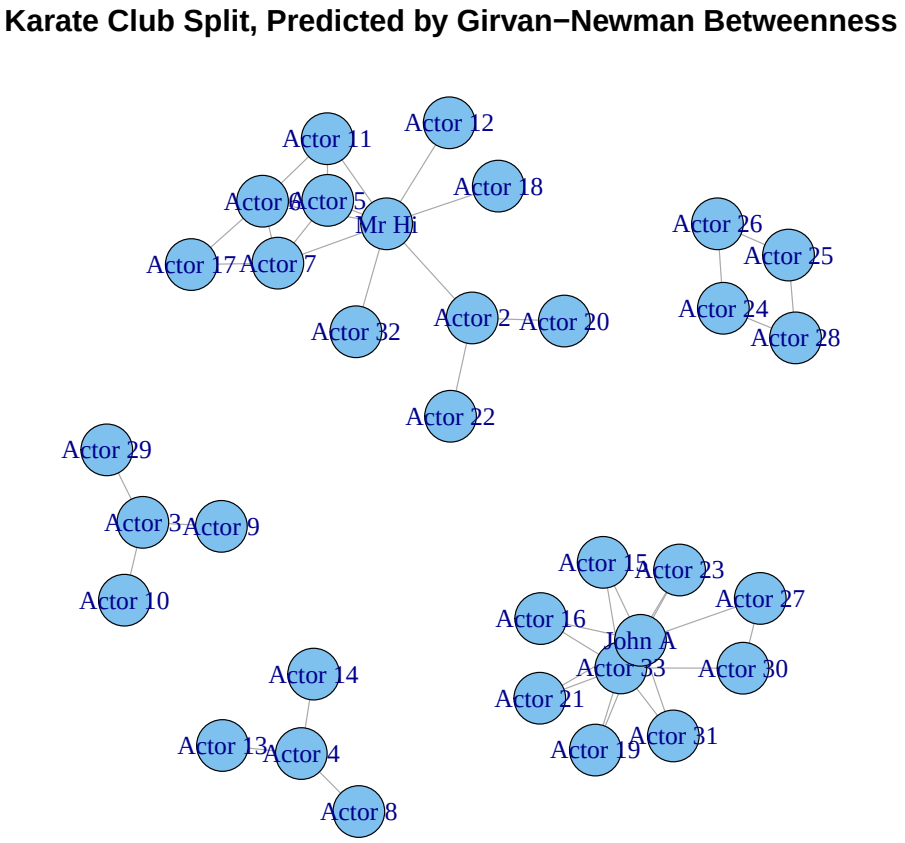

Figure 7: Karate Club Graph Split Into 5 Groups Predicted by Girvan & Newman Betweenness Clustering

Bibliography

[1] Algoriths fro community detection in graphs. http://stackoverflow.com/questions/5822265/are-there-implementations-of-algorithms-for-community-detection-in-graphs.

[2] Karate network. http://rstudio-pubs-static.s3.amazonaws.com/16513_2f2f84a794a64c19a086a4ec3123f889.html.

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

[4] Zachary dataset. http://www.casos.cs.cmu.edu/computational_tools/datasets/external/karate/index11.php.

[5] Gabor Csardi. R igraph manual pages, november 2013.

[6] Wayne W. Zachary. *An Information Flow Model for Conflict and Fission in Small Groups*, chapter 33, pages 452–473. Journal of Anthropological Research, 1977.