

Assignment 5

CS 532: Introduction to Web Science

Spring 2018

Chandrasekhar Reddy Muthyala

Finished on March 9, 2018

1

Question

CS 432/532 Web Science

Spring 2018

<http://anwala.github.io/lectures/cs532-s18/>

Assignment #5

Due: 11:59pm March 7

(10 points)

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.

Clues:

1. Draw original Karate club graph (two connected components) after split (Week 6 lecture, slide 98).
2. Run multiple iterations of graph partitioning algorithm (e.g., Girvan-Newman Algorithm) on experimental Karate club graph until the graph splits into two connected components.
3. Compare the connected components of the experimental graph (in 2.) with the original connected components of the split Karate club graph (in 1.). Are they similar?

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

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

<https://snap.stanford.edu/snappy/doc/reference/CommunityGirvanNewman.html>

http://igraph.org/python/doc/igraph-pysrc.html#Graph.community_edge_betweenness

Answer

To solve this problem initially I have gone through the class slides as well as URLs given in the question and got idea how to solve this problem and started coding using R and wrote script called **KarateClub.R** shown Listing 1, namely because using the igraph library and igraphdata, which provided the karate club dataset, was vastly easier and simpler to use with graphs.

To prove that the result of this graph split could have been predicted, I decided to use the edge betweenness algorithm provided by the igraph library, which actually used the Girvan-Newman algorithm [3]. The Girvan-Newman algorithm at each step in the graph checks if: an edge has the highest “betweenness” it would be removed from the graph and the modularity of the graph would then be recomputed. I decided to use the *cluster_edge_betweenness* function to cluster them nicely based on their strongly or weakly connections with each other actor.

The H and A nodes, as shown in Figure 1, represent Mr. Hi and John A respectively. It shows how they are at the center of each of their factions. Its assumed that after running the graph through the Girvan-Newman algorithm until a certain number of groups had been made with zero connections left to the other groups, that it would match fairly well like Figure 2. Luckily, the igraph library automatically calculated the betweenness for me and stored it in a variable that I could access and then manually delete the edges with the highest betweenness. I then checked if the number of components, or disconnected groups, was 2.

What I found was that my actual split was 94% accurate with with two actors being sent to the wrong group. Actors 3 and 14, making it 2/33 were grouped wrongly with John A’s officer faction as shown in Figure 3. This shows that this model does represent reality quite well and only missed a single actor compared to Zachary Wayne’s model prediction of 97% accuracy [4].

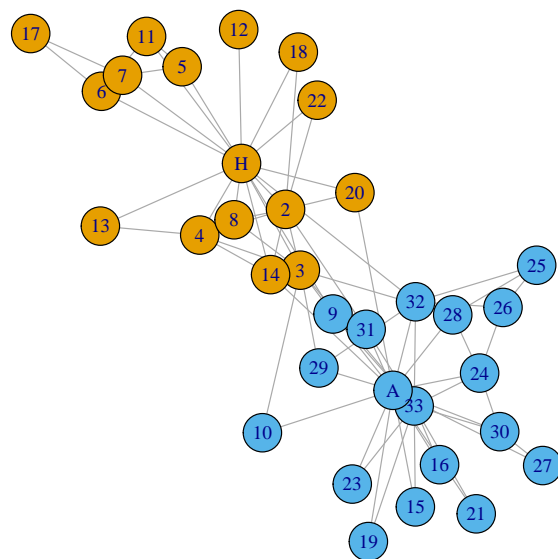


Figure 1: Original Karate Club Split

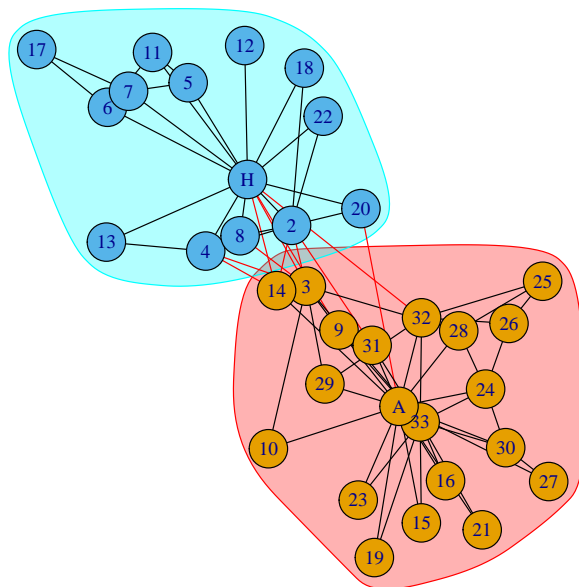


Figure 2: Desired groups when split

```

1  #!/usr/bin/env Rscript
2  ## Download and install the package
3  # install.packages("rjson", repos = "http://cran.us.r-project.
    org")
4  # https://cran.r-project.org/web/packages/igraphdata/README.html
5  library(igraph)
6  library(igraphdata) # provides karate data
7  setwd(getwd())
8  set.seed(20)
9  # convert to graph format
10 data(karate)
11 # print(karate)
12 kclub <- karate
13
14 clust <- cluster_edge_betweenness(kclub)
15
16 # Original split
17 # origCommunities <- plot.igraph(kclub)
18 # Second value is number of communities to create
19 groups <- cutat(clust, 2)
20 # Assumed should have split
21 # plot(structure(list(membership=groups), class="communities"),
    kclub)
22 connected_components_count <- count_components(kclub)
23
24 # index counter for node edges
25 edgeIndex <- 1
26 # split into groups, actual
27 while(connected_components_count != 3){
28   # starts at 1
29   edgesRemoved <- delete.edges(kclub, clust$removed.edges[seq(1,
    edgeIndex-1)])
30   connected_components_count <- count_components(edgesRemoved)
31   # set new Graph
32   origCommunities <- edgesRemoved
33   #get separate pdf according to the split
34   # pdf(paste("Iteration",edgeIndex,".pdf"))
35   # plot.igraph(origCommunities)
36   edgeIndex <- edgeIndex + 1
37 }
38
39 print(paste("Iteration Count",edgeIndex))
40 plot.igraph(origCommunities)

```

Listing 1: R script for splitting groups using Girvan-Newman algorithm

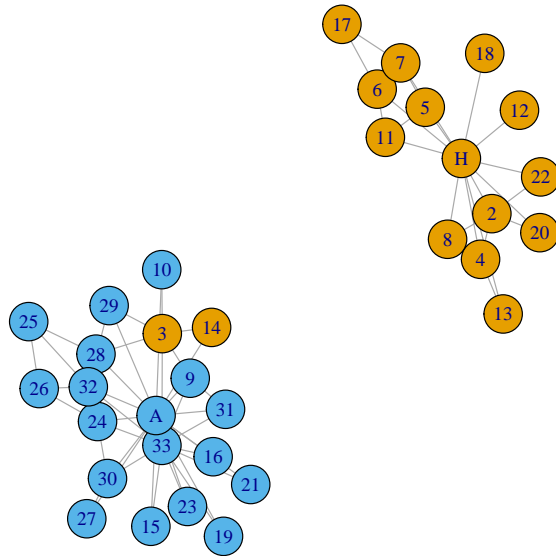


Figure 3: Actual predicted group split with Girvan-Newman from karate-Club.R

2

Question

(extra credit, 3 points)

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

For this question I approached it the exact same as the first question. I again used the R script as shown in Listing 1 to split the karate club into further groups. In the the script I simply changed the base case for the while loop up to 3, 4 and 5, which signified how many groups were required until it would cease deleting edges. For a 3 group split this is shown in Figure 4. For a 4 group split this is shown in Figure 5. For a 5 group split this is shown in Figure 6. I didn't attempt to verify the accuracy of these group splits as I couldn't find more information regarding a split to these degrees. All the iterations of the 2, 3, 4, 5 split graphs are stored in **twoGroupGraph.pdf**, **threeGroupGraph.pdf** , **fourGroupGraph.pdf** , **fiveGroupGraph.pdf**

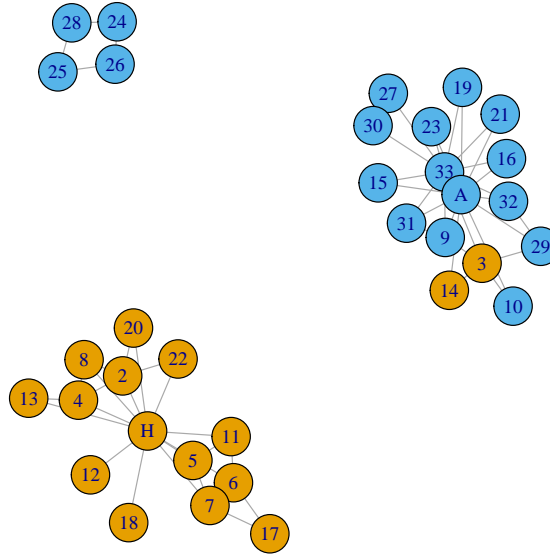


Figure 4: Group split of 3 with Girvan-Newman algorithm from karate-Club.R

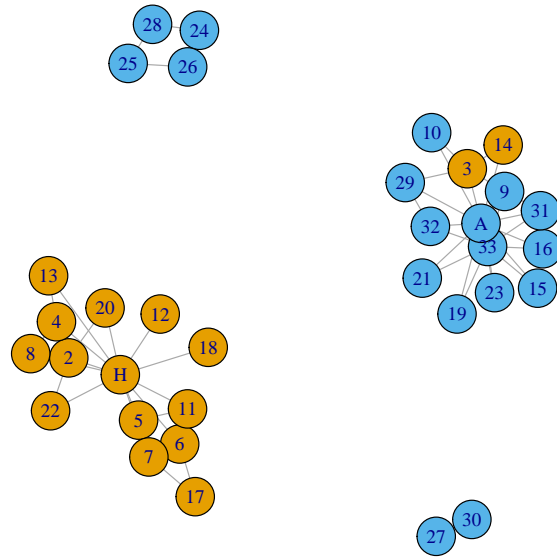


Figure 5: Group split of 4 with Girvan-Newman algorithm from karate-Club.R

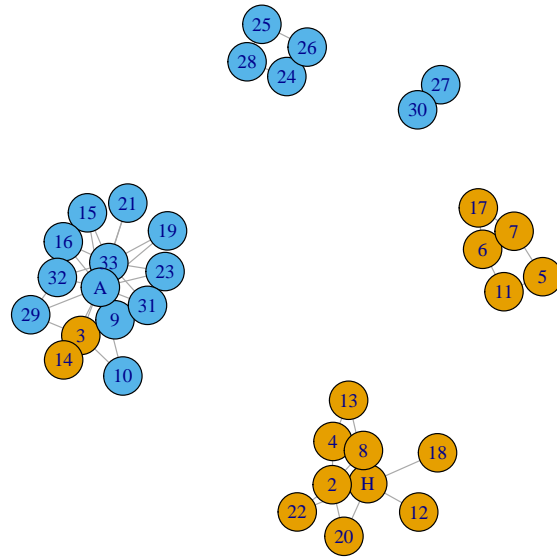


Figure 6: Group split of 5 with Girvan-Newman algorithm from karate-Club.R

References

- [1] Csardi, Gabor. “Package ‘graphdata’ ” iGraphData. Cran-R-Project, 13 July 2015. Web. 16 March 2017. <https://cran.r-project.org/web/packages/igraphdata/igraphdata.pdf>.
- [2] Csardi, Gabor, “Package ‘igraph’ ” iGraph. Cran-R-Project, 13 July 2015. Web. 16 March 2017. <http://igraph.org/r/doc/igraph.pdf>.
- [3] Rodrigues, David. “Finding Communities in networks with R and igraph” N.p., n.d. Web. 16 March 2017. <http://www.sixhat.net/finding-communities-in-networks-with-r-and-igraph.html>.
- [4] W. W. Zachary, An information flow model for conflict and fission in small groups, *Journal of Anthropological Research* 33, 452-473 (1977).