Assignment 5

CS 532: Introduction to Web Science

Spring 2018

Hrishikesh Gadkari

1

Question

```
Assignment #5
Due: 11:59pm March 16

(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 lec
2. Run multiple iterations of graph partioning algorithm (e.g., Girvan-Newman Algorit
3. Compare the connected components of the experimental graph (in 2.) with the origin

Useful sources include:

* Original paper

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

* Week 6 Slides:

https://docs.google.com/presentation/d/1ihf6N8bHgzM5VLAyHkmF_i5JGUBVpCSdsvYpk8XgHwo/e

* 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

https://networkx.github.io/documentation/networkx-1.10/reference/generated/networkx.g

https://networkx.github.io/documentation/networkx-1.9/examples/graph/karate_club.html

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

http://stackoverflow.com/questions/9471906/what-are-the-differences-between-community

http://stackoverflow.com/questions/5822265/are-there-implementations-of-algorithms-fo

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

I first attempted to solve this problem using python with networkx's library for graphing and created a script called **club.py**. I later stopped attempting to use this because I learned that it didn't provide visual graphs without installing mathplotlib. Therefore I switched to using R and wrote script called **club.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 an 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].

```
1  library(igraph)
2  library(igraphdata) # provides karate data
3  setwd(getwd())
4
5  # ensure same seed/graph sample each run
6  set.seed(20)
7  # convert to graph format
8  data(karate)
9  kclub <- karate
10
```

```
11 # remove edges
12 clust <- cluster_edge_betweenness(kclub)
13 # Original split
14 origCommunties <- plot.igraph(kclub)
15 # Second value is number of communities to create
16 groups <- cutat(clust, 2)
17 # Assumed should have split
18 plot(structure(list(membership=groups), class="communities"),
       kclub)
19 connected_components_count <- count_components(kclub)
20
21 # index counter for node edges
22 edgeIndex <- 1
23 # split into groups, actual
24 while(connected_components_count != 2){
25   # starts at 1
26   edgesRemoved <- delete.edges(kclub, clust$removed.edges[seq(1,
         edgeIndex-1)])
27   connected_components_count <- count_components(edgesRemoved)
28   # set new Graph
29   origCommunties <- edgesRemoved
30   edgeIndex <- edgeIndex + 1
31 }
32
33 print(paste("Iteration Count",edgeIndex))
34 plot.igraph(origCommunties)
```

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

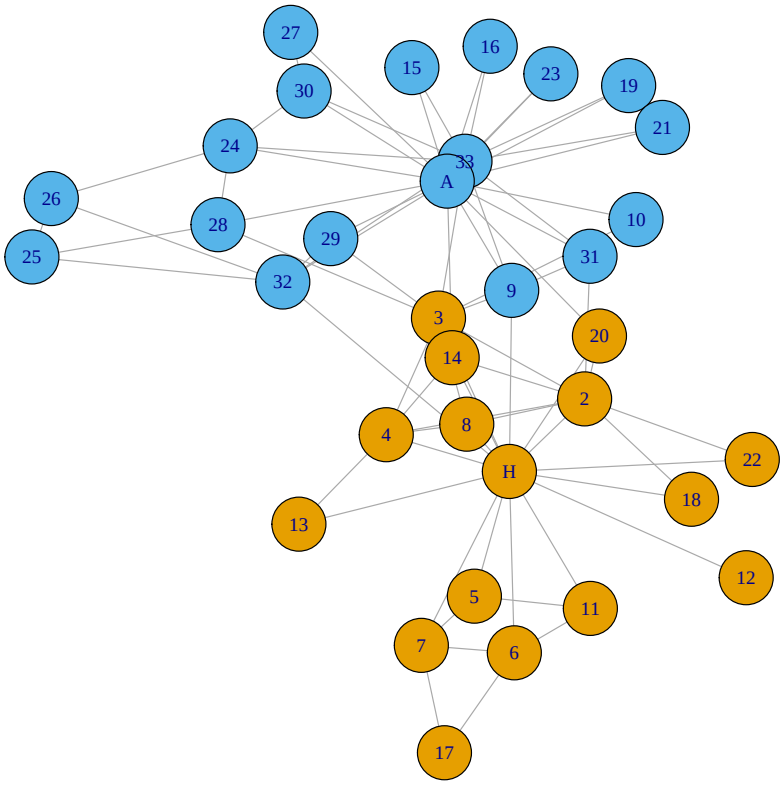

Figure 1: Original Karate Club Split

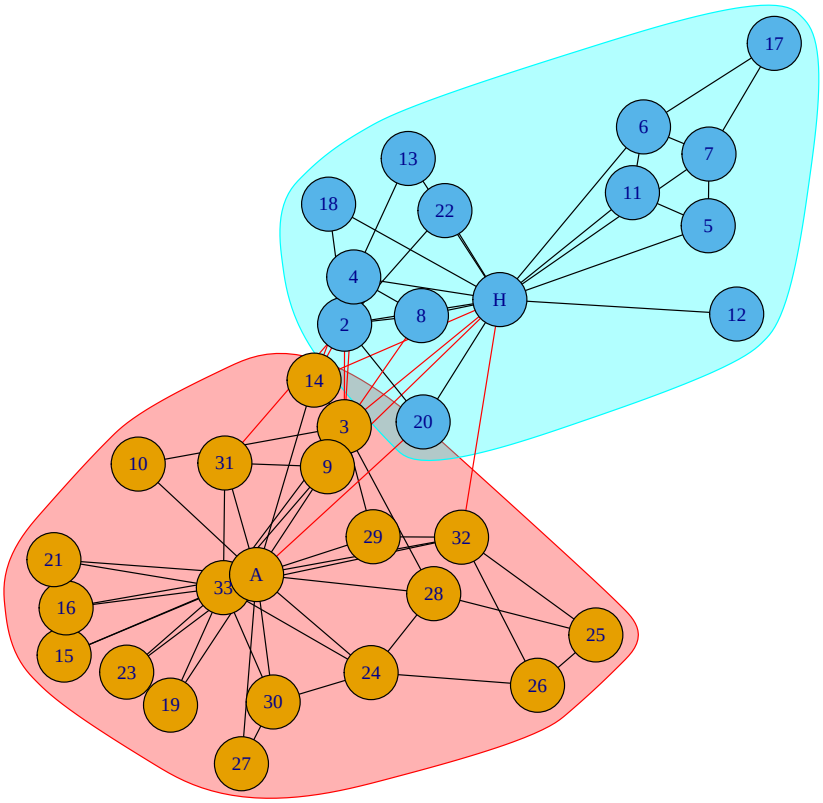

Figure 2: Desired groups when split

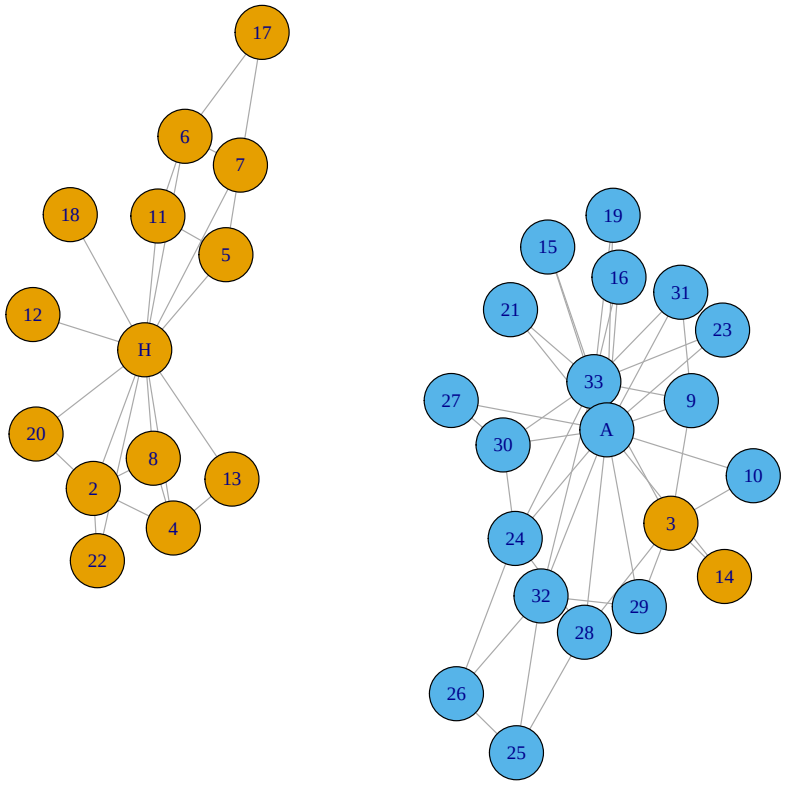

Figure 3: Actual predicted group split with Girvan-Newman from club.R

2

Question

(extra credit, 3 points)

3. 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.

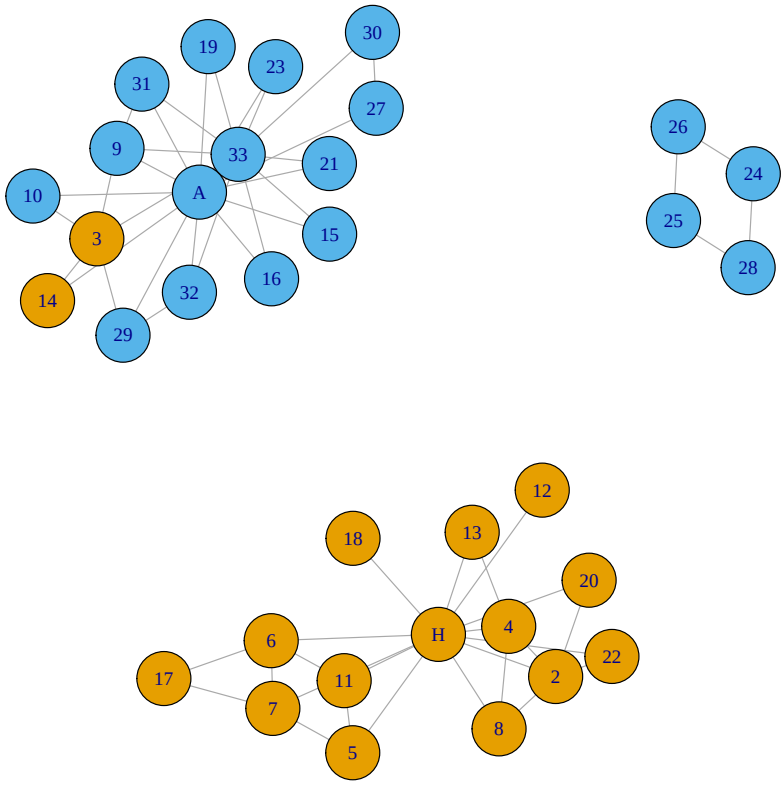

Figure 4: Group split of 3 with Girvan-Newman algorithm from club.R

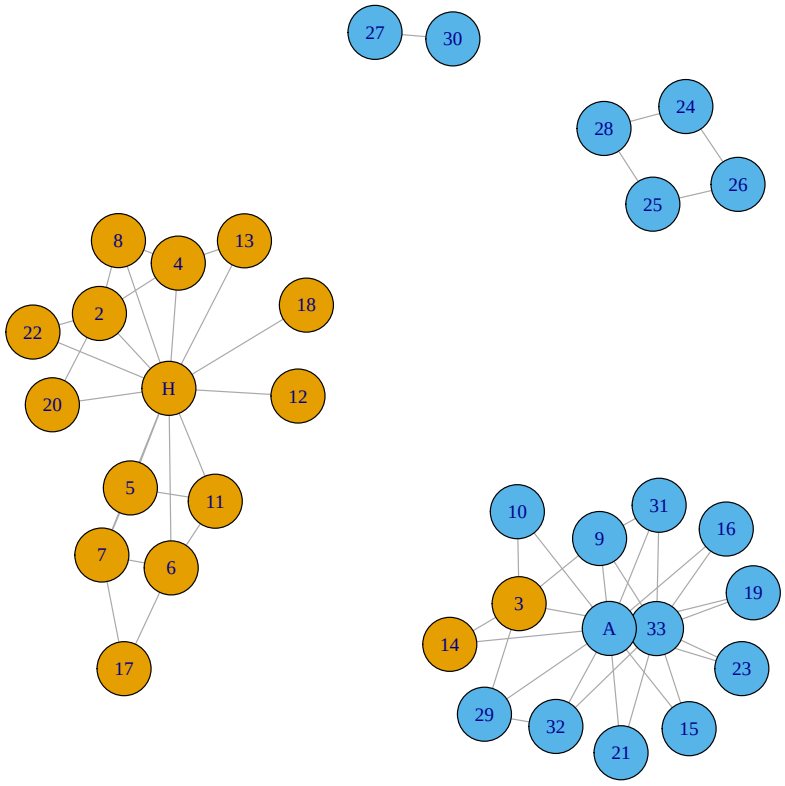

Figure 5: Group split of 4 with Girvan-Newman algorithm from club.R

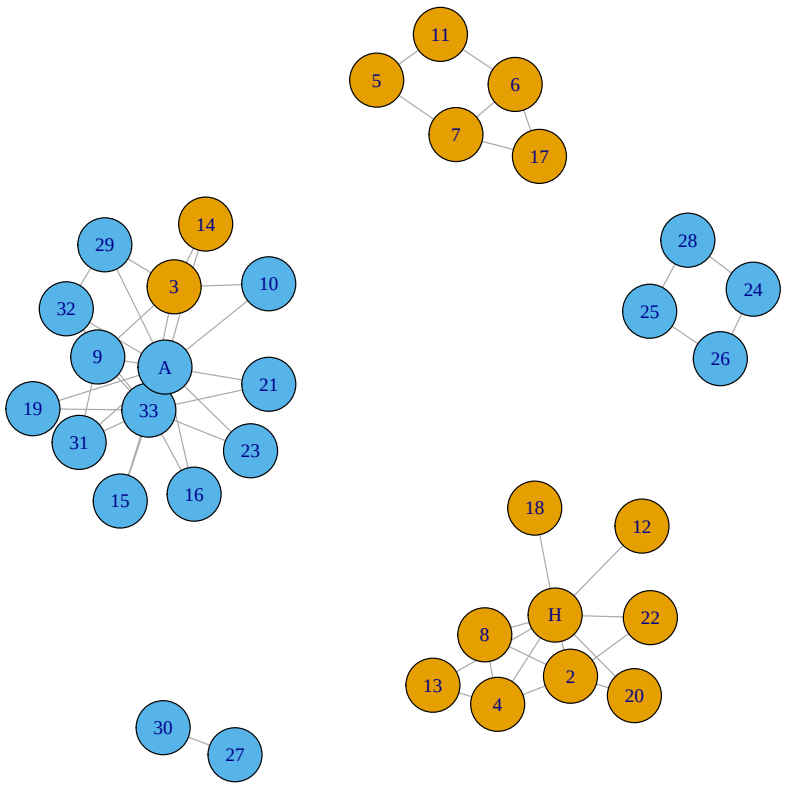

Figure 6: Group split of 5 with Girvan-Newman algorithm from club.R

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).