Assignment 5

CS 532: Introduction to Web Science

Spring 2017

Grant Atkins

Finished on March 16, 2017

1

Question

```
CS 432/532 Web Science
Spring 2017
http://phonedude.github.io/cs532-s17/

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.

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

https://networkx.readthedocs.io/en/stable/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

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 **karateClub.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 **karateClub.R**, 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 3. 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 were grouped wrongly with John A as shown in Figure **??**
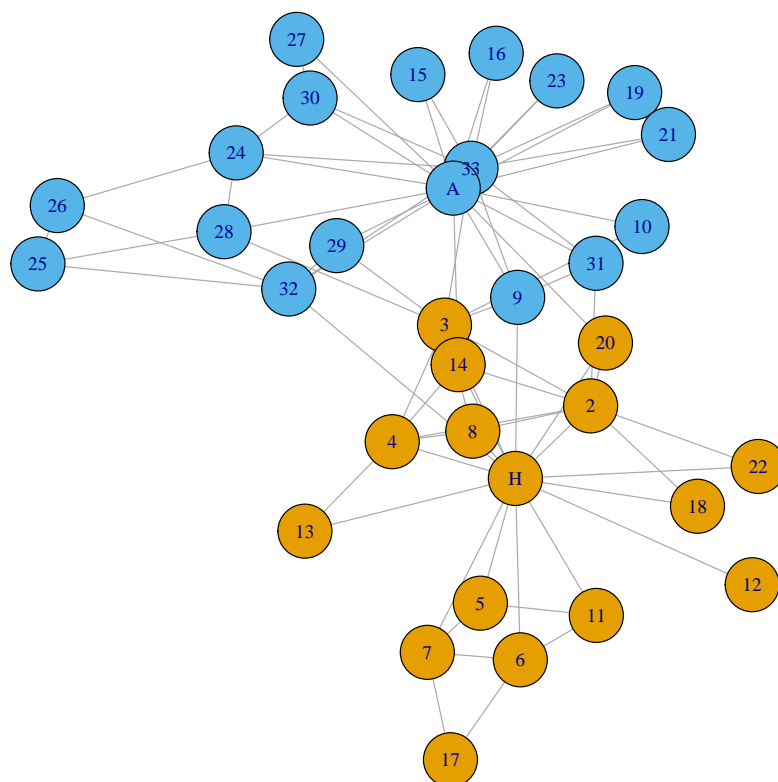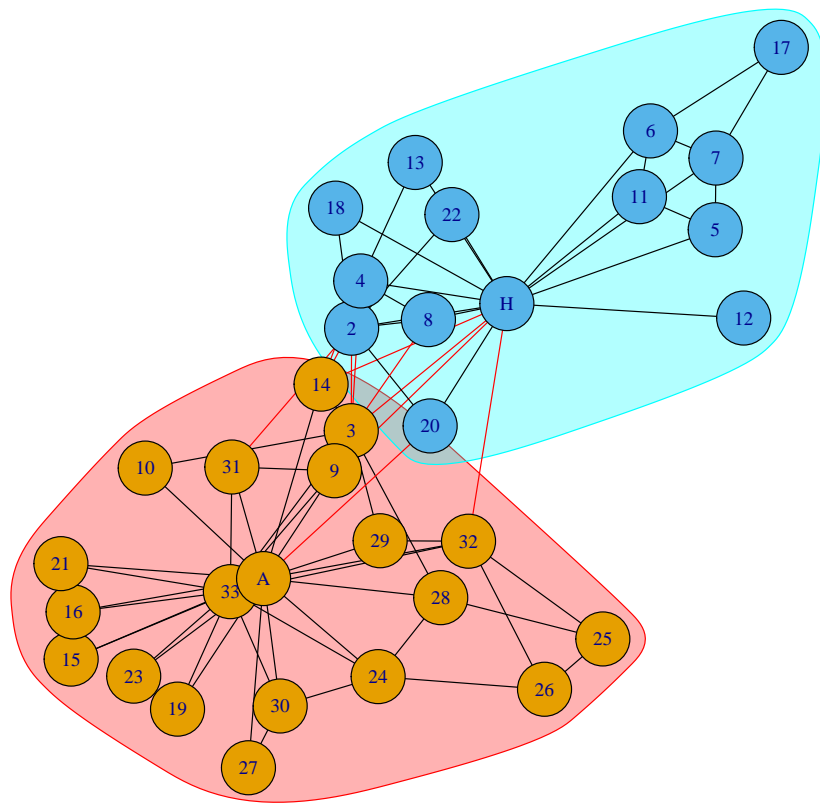
Figure 1: Original Karate Club Split
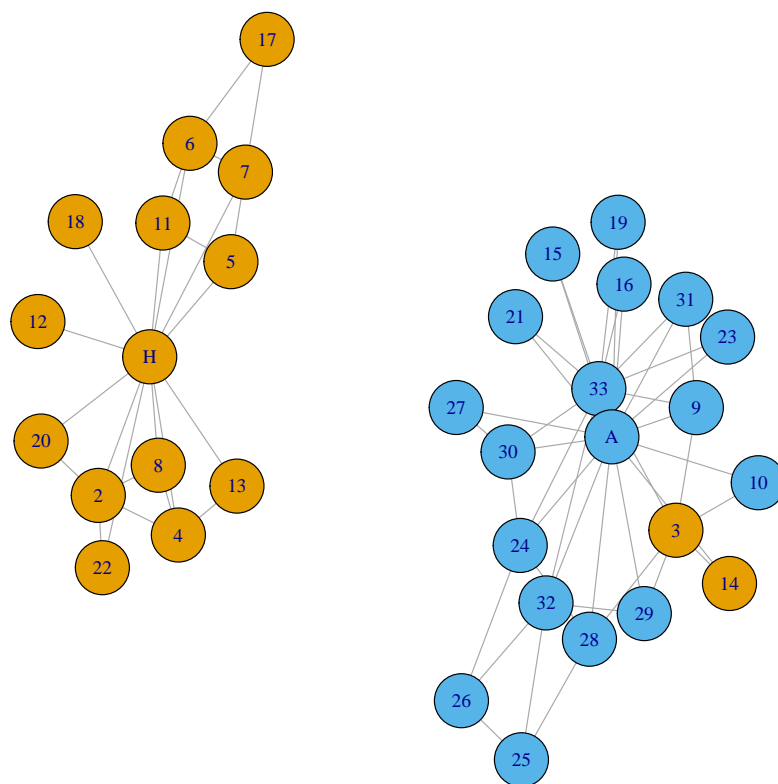
Figure 2: Desired groups when split

Figure 3: Desired groups when split

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

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`.