Assignment Five

Jacob Berlin

CS432 – Spring 2016

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

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

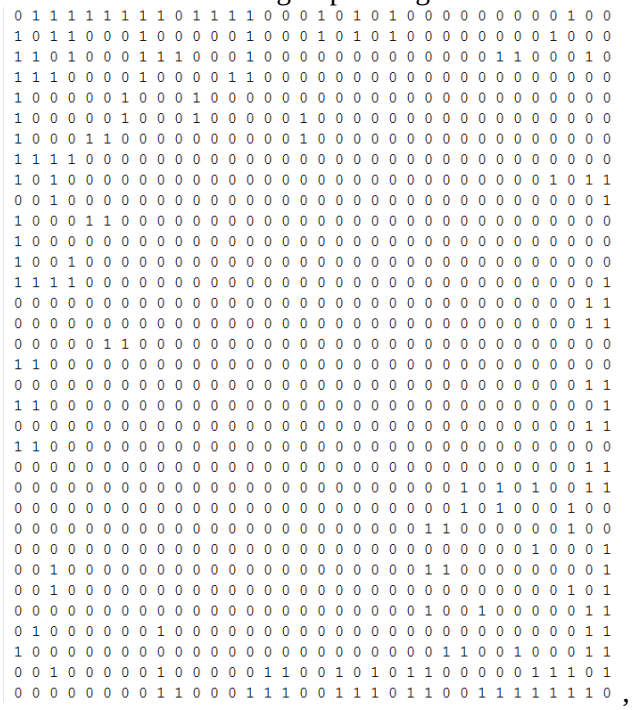
http://igraph.org/python/doc/igraph-pysrc.html#Graph.community_edge_betweenness

First, to start this problem we need to look at what the Karate Club (Zachary, 1977) split was initially comprised of. The story of the Karate Club Split is that there was a club comprised of 34 different members that split apart due to disagreements between Mr. Hi, the instructor, and the president of the club, John A. According to the paper written by Zachary, there were eight different situations that lead to the students have relations to each other.

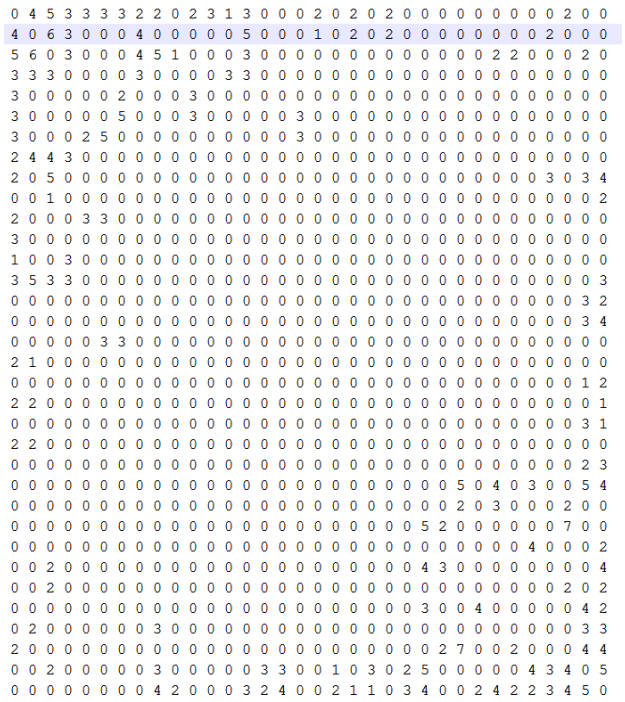
1. Association in and between academic classes at the university.
2. Membership in Mr. Hi's private karate studio on the east side of the city where Mr. Hi taught nights as a part-time instructor.
3. Membership in Mr. Hi's private karate studio on the east side of the city, where many of his supporters worked out on weekends.
4. Student teaching at the east-side karate studio referred to in (2). This is different from (2) in that student teachers interacted with each other, but were prohibited from interacting with their students.
5. Interaction at the university rathskeller, located in the same basement as the karate club's workout area.
6. Interaction at a student-oriented bar located across the street from the university campus.
7. Attendance at open karate tournaments held through the area at private karate studios.
8. Attendance at intercollegiate karate tournaments held at local universities. Since both open and intercollegiate tournaments were held on Saturdays, attendance at both was impossible.

Under the assumption of the Girvan & Newman algorithm, we can assume that we can predict the outcome of the group's split.

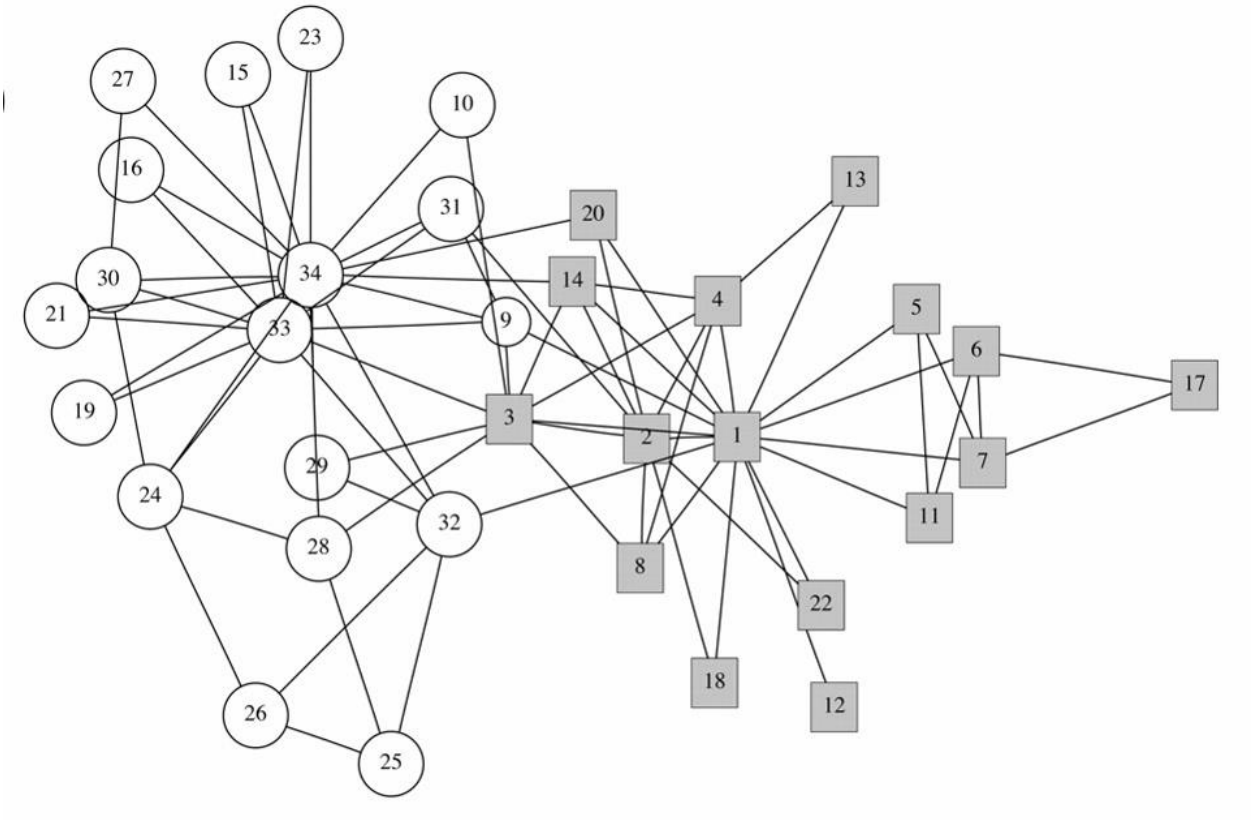
Shown the data for the group's weight and interactions with each other:



,



Initially, looking at the relations of each of the members inside of the starting club we can see that each of the members is linked together by having some sort of interaction with each other outside of the club.



We can immediately see that some people, more than others, are connected to specific groupings around themselves. This makes it easy to see that both node 1 and node 34 are centralized, giving the probability that they are John A. and Mr. Hi. To start the process of splitting the groups we need to see how many people that each of the nodes are linked to. To do this, I ran the python program 'KarateClubNodeDegree.py'.

```
#!/usr/bin/env python
"""
Zachary's Karate Club graph

Data file from:
http://vlado.fmf.uni-lj.si/pub/networks/data/Ucinet/UciData.htm

Reference:
Zachary W. (1977).
An information flow model for conflict and fission in small groups.
Journal of Anthropological Research, 33, 452-473.
"""
import networkx as nx
G=nx.karate_club_graph()
print("Node Degree")
for v in G:
    print('%s %s' % (v,G.degree(v)))
```

Which gave the output:

```
Node Degree
0 16
1 9
2 10
3 6
4 3
5 4
6 4
7 4
8 5
9 2
10 3
11 1
12 2
13 5
14 2
15 2
16 2
17 2
18 2
19 3
20 2
21 2
22 2
23 5
24 3
25 3
26 2
27 4
28 3
29 4
30 4
31 6
32 12
33 17
```

Now, using the generalized Girvan & Newman algorithm, we follow the steps provided in the 'Community Detection' powerpoint.

- Calculate edge-betweenness for all edges
- Remove the edge with highest betweenness
- Recalculate betweenness
- Repeat until all edges are removed, or modularity function is optimized (depending on variation)

To start, we have to understand what 'betweenness' actually is. Girvan & Newman altered a variant of 'Vertex Betweenness Centrality', which originally is defined as the total number of shortest paths that pass through each vertex on the network. They created the variant, called 'Edge Betweenness Centrality' which is the number of shortest paths that pass through a given edge.

The purpose of calculating and taking out betweenness in essence shows the ability to have complete grouping and predict 'clusters' of data…precisely what we are looking for with the split, two separate groups.

"If there is more than one shortest path between a pair of vertices, each path is given equal weight such that the total weight of all the paths is unity"

Formally, the equation is:

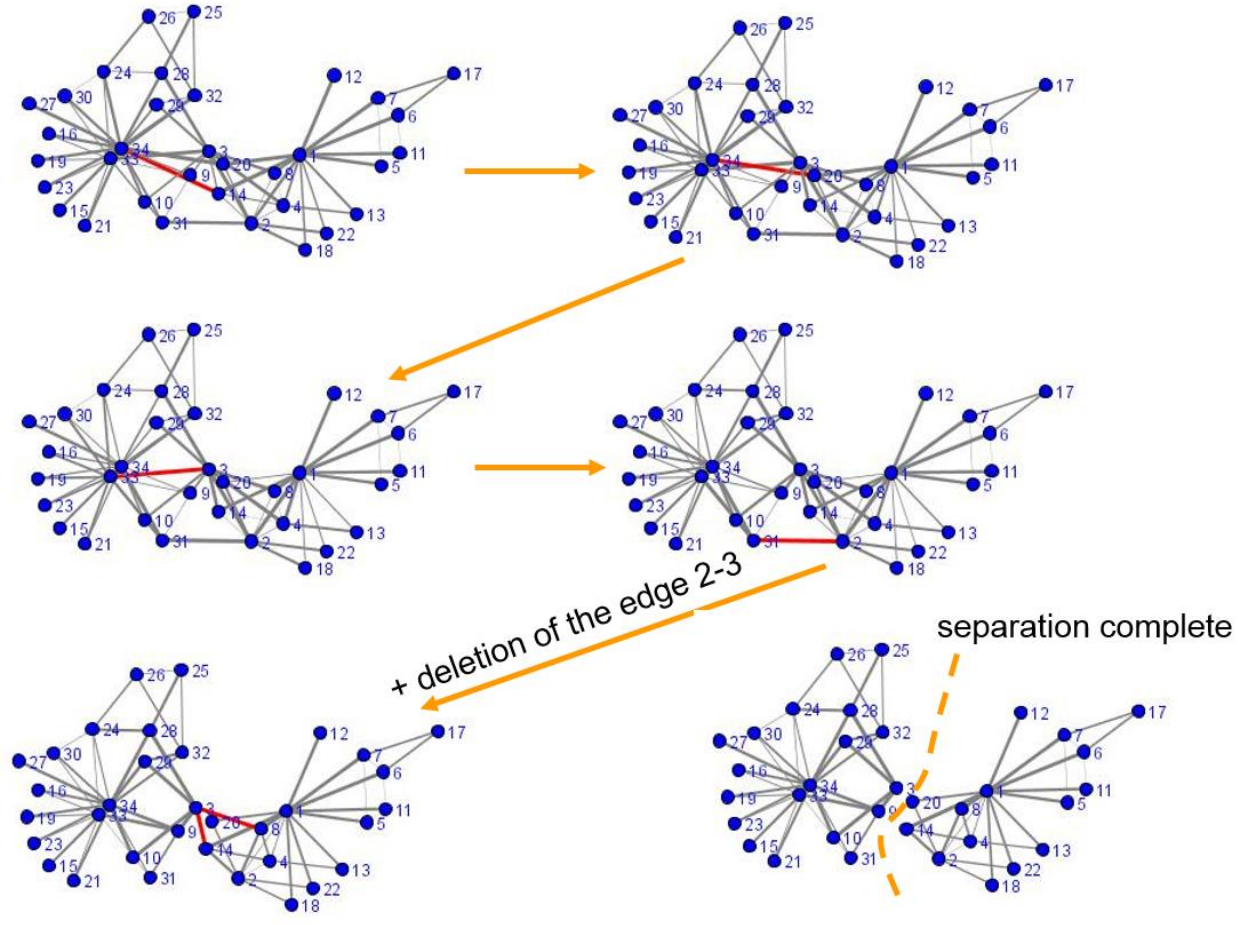
$$c_B(e) = \sum_{s,t \in V} \frac{\sigma(s,t|e)}{\sigma(s,t)}$$

Now, to calculate the betweenness of the edges, using igraph in python:

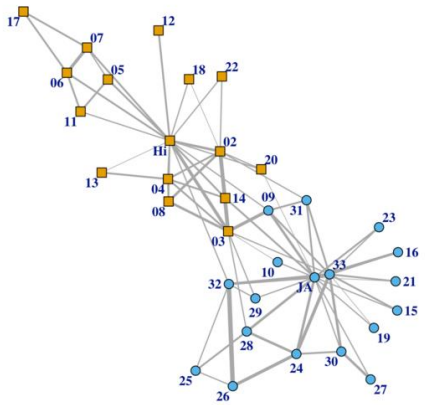
community_edge_betweenness(Karate, clusters=2, directed=False)

In this case, we need to make sure that for the fourth step of the Girvan &Newman algorithm we have to go only until there is a split in both of the groups instead of until every node is singled out.

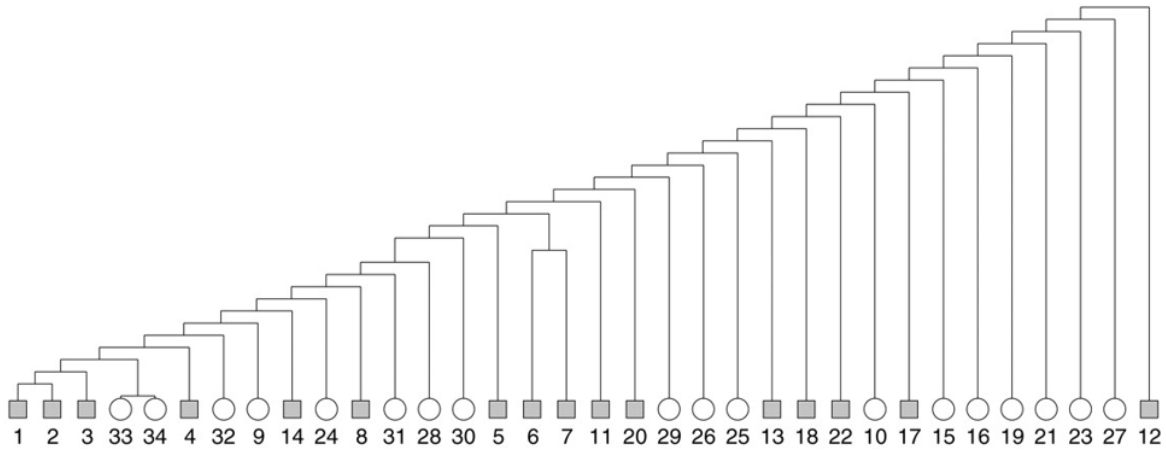
As shown in the ‘CommunityDetection’ PowerPoint, an illustration of the process is as follows:

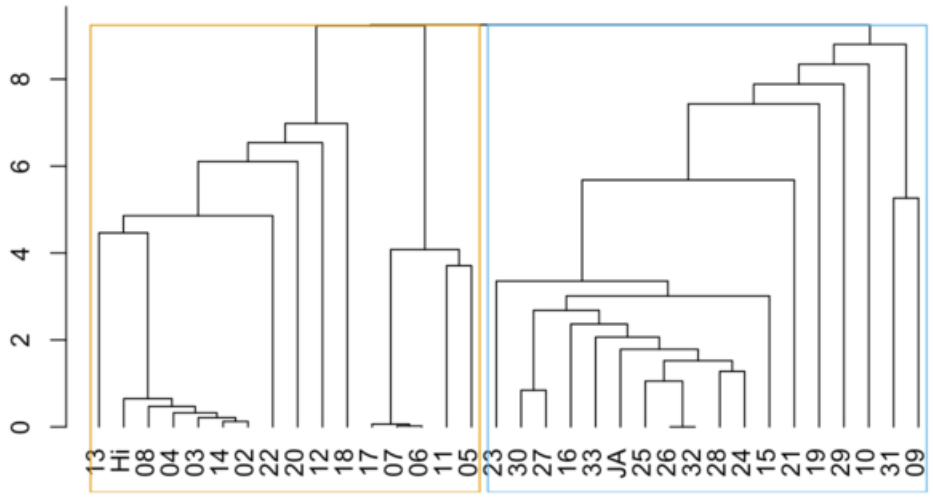


Finally, as you can see, the graph shows the distinct split in between the two groups, with John A. on the left and Mr. Hi on the right.



This can be shown otherwise by the hierarchal clustering tree and the result after the algorithm is added:





Given the original outcome that Zachary received in 1977 and the process shown above it is safe to assume that yes, prediction for the Karate Club split is definitely possible. (Zachary analysis shown below)

EVALUATION OF THE HYPOTHESES

INDIVIDUAL NUMBER IN MATRIX *C*	FACTION MEMBERSHIP FROM DATA	FACTION MEMBERSHIP AS MODELED	HIT/ MISS	CLUB AFTER SPLIT FROM DATA	CLUB AFTER SPLIT AS MODELED	HIT/ MISS
1	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
2	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
3	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
4	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
5	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
6	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
7	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
8	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
9	John	John	Hit	Mr. Hi's	Officers'	Miss
10	John	John	Hit	Officers'	Officers'	Hit
11	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
12	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
13	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
14	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
15	John	John	Hit	Officers'	Officers'	Hit
16	John	John	Hit	Officers'	Officers'	Hit
17	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
18	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
19	John	John	Hit	Officers'	Officers'	Hit
20	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
21	John	John	Hit	Officers'	Officers'	Hit
22	Mr. Hi	Mr. Hi	Hit	Mr. Hi's	Mr. Hi's	Hit
23	John	John	Hit	Officers'	Officers'	Hit
24	John	John	Hit	Officers'	Officers'	Hit
25	John	John	Hit	Officers'	Officers'	Hit
26	John	John	Hit	Officers'	Officers'	Hit
27	John	John	Hit	Officers'	Officers'	Hit
28	John	John	Hit	Officers'	Officers'	Hit
29	John	John	Hit	Officers'	Officers'	Hit
30	John	John	Hit	Officers'	Officers'	Hit
31	John	John	Hit	Officers'	Officers'	Hit
32	John	John	Hit	Officers'	Officers'	Hit
33	John	John	Hit	Officers'	Officers'	Hit
34	John	John	Hit	Officers'	Officers'	Hit
TOTALS		34 hits, 0 misses 100% hits, 0% misses			33 hits, 1 miss 97% hits, 3% misses	

References:

https://rstudio-pubs-static.s3.amazonaws.com/99456_35043d0a235e454787b2665087141914.html

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

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

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

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

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