OLD DOMINION UNIVERSITY

CS 495: Introduction to Web Science
Instructor: Micheal L. Nelson, Ph.D
Fall 2014 Thursdays 4:20pm – 7:10pm ECSB 2120

Assignment # 6
Joseph Elder UIN: 00844802

Honor Pledge
I pledge to support the Honor System of Old Dominion University. I will refrain from any form of academic dishonesty or deception, such as cheating or plagiarism. I am aware that as a member of the academic community it is my responsibility to turn in all suspected violations of the Honor Code. I will report to a hearing if summoned

October 28, 2014

Contents

1 Assignment 6

1.1 Question 1

1.1.1 The Problem

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

1.1.2 The Solution

The original study “An Information Flow Model for Conflict and Fission in Small Groups”, conducted by Wayne W. Zachary documents and interprets the social interactions of a karate class which faced a division due to an ideological dispute over lesson prices and leadership. This study appears to be a somewhat classic problem to be used as a dataset. The study includes several mathematical models representative of the existing connections in social settings outside of the class.

The models show connections or the paths which communication takes place in the group. Connections represent the flow of information between individuals. The models allow for a quantitative anthropological study of the group. By interpreting and applying some theory to the dataset represented by the models it is possible to predict the outcome of the fission in the group.

By representing the data from the model graphically the connections of the karate class are much more easily visualized. Using python version 2.7.6 as well as the “networkx” and “matplotlib.pyplot” python modules, a graph is produced by taking the data from the karate class study and plotting it with mathplotlib. Running “src/karate_club.py” will produce the graph as a png image called “graph.png” this code is based directly on the example included on the networkx website.

Though the class had nearly 60 members, the graph will only include the 34 members of the club who regularly participated in events. This measure is to prevent the display of nodes with no connections or

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0
2	1	0	1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0
3	1	1	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0
4	1	1	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1
10	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
11	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
17	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
20	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
22	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	1
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
28	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1
29	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	1
31	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
32	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	1
33	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0	1	1	0	0	0	0	0	1	1	1	0	1
34	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	0

Table 1: Matrix of Relationhips in the Club

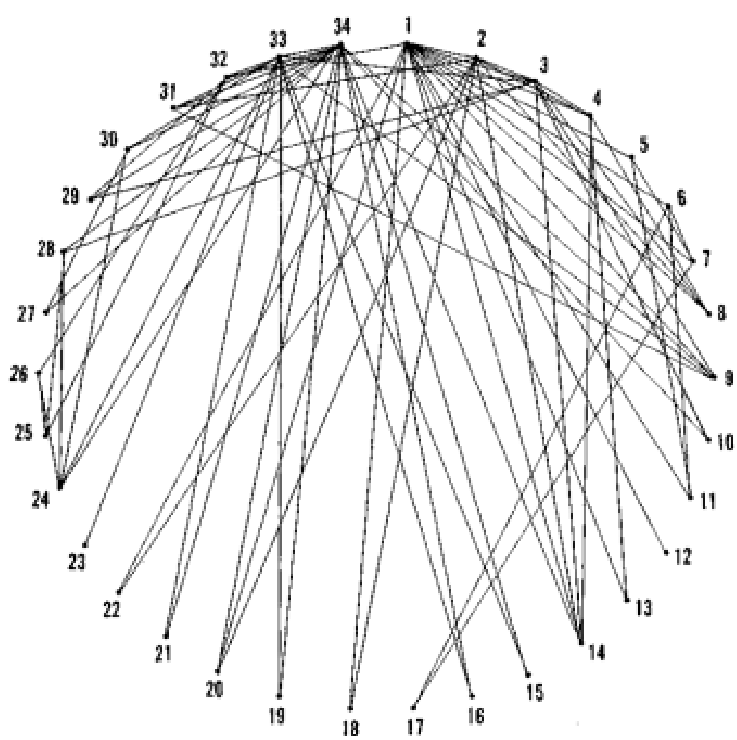

Figure 1: Visualization of Group Connections

vertices with no edges.

The generated graph is somewhat similar to figure 1 in the original report by Zachary W., both figures show the friendships or essentially the paths which information would flow throughout the group. Upon initial inspection of both figure 1 and figure 2 several conclusions can be drawn. The first item to note is that it is clear that there are two hubs shown in figures one as 1 and 34, and in figure 2 as 0 and 33.

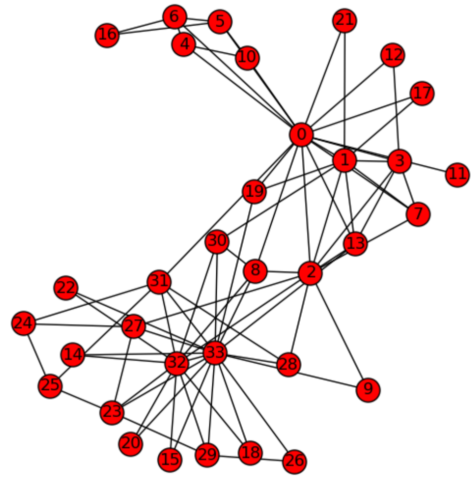

Figure 2: Karate Class Network Graph

The hubs represent the complex social environment which take place around essentially the leaders of the group, the lowest number in both figures represents Mr Hi. the karate instructor, and the highest number in both cases represents the highest officer of the club. Given this context it becomes easier to understand what is going on in the graphs. This is the first point of evidence toward the model accurately predicting the split.

The large number of connections for the highest officer and Mr. Hi in both graphs show their connections to the members of the group and form a subgroup where information is passed quickly since these members are likely to be connected with each other as well. Social connections to members in other subgroups or factions do exist so information will eventually reach everyone but it must travel through several members to reach the other side. This concept is similar to the “source” and “sink” concepts studied by Ford and Fulkerson who were interested in network flow theory.

The source of information originates with the leader of a faction. That information then flows throughout the group until it reaches the sink which would be the other faction leader. By qunatifing this information and appling it across the original matrix another matrix is formed. The resulting matrix is weighted according to the scale of friendships in the group based on social interactions.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	0	4	5	3	3	3	3	2	2	0	2	3	2	3	0	0	0	2	0	2	0	2	0	0	0	0	0	0	0	0	0	2	0	0
2	4	0	6	3	0	0	0	4	0	0	0	0	0	5	0	0	0	1	0	2	0	2	0	0	0	0	0	0	0	0	2	0	0	0
3	5	6	0	3	0	0	0	4	5	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	3	0
4	3	3	3	0	0	0	0	3	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	3	0	0	0	0	0	2	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	3	0	0	0	0	0	5	0	0	0	3	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	3	0	0	0	2	5	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	2	4	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	2	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	4	3
10	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
11	2	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	3	5	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	4
17	0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
20	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1
22	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	4	0	2	0	0	5	4
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	3	0	0	0	2	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	2	0	0	0	0	0	0	7	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	2
28	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	3	0	0	0	0	0	0	0	0	4
29	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	4	0	0	0	0	0	3	2
31	0	2	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3
32	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	7	0	0	2	0	0	0	4	4
33	0	0	2	0	0	0	0	0	3	0	0	0	0	0	3	3	0	0	1	0	3	0	2	5	0	0	0	0	0	4	3	4	0	5
34	0	0	0	0	0	0	0	0	4	2	0	0	0	3	2	4	0	0	2	1	1	0	3	4	0	0	2	4	2	2	3	4	5	0

Table 2: Quantified Matrix of Relative Strengths of the Relationships in the Karate Club

The dataset is explored further using python 2.7.6 and the python module “igraph” the source code for all code pertaining to igraph for question 1 is contained in the file “src/karate_chop.py”. Vertices are set for the graph and edges are added approiariatly in python. Each vertex represents one of the 34 group members, each non-directional connection between two group members is called an edge; there are 78 total edges in the graph. Once the edges and vertices are set in igraph the problem can be more easily manipulated in python.

The degree of connectivity can be calculated very quickly with the “.degree()” function. Calling this function on the dataset will return the number connections each vertex has. The igraph module includes automated functions for community detection given vertices and edges.

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

Figure 3: Command Prompt output from igraph displaying all vertices and edges

The output of the Several community detection methods used in “karate_chop.py” for predictions can be found in the file “out.txt”. The first method of community detection is based off of edge betweenness. Edge betweeness is simply the number of shortest paths that pass through a given edge. This solution produces a dendrogram which resembles a tree diagram which splits up the vertices into subgroups at varying levels. The bottom level of the graph is the linking of two largest subgroups, each of these subgroups shown is the predicted outcome of the group as a whole splitting into two parts.

Another dendrogram is generated when using the Walktrap method. The walktrap method is “an approach based on random walks. The general idea is that if you perform random walks on the graph, then the walks are more likely to stay within the same community because there are only a few edges that lead outside a given community.” This graph represents the predicted outcome of the split in a similar manner to how it is calculated for edge betweeness. The bottom level will represent the two largest subgroups of vertices or karate club members. This same solution will be used additionally to evaluate the dendrogram generated by the FastGreedy community detection method. The FastGreedy community detection method is similar to the edge betweenness method however it works from the bottom up instead of from the top down.

Additional predictions are made by the Leading Eigenvectors and Label Propagation community detection methods. Both use clustering to form two clusters out of the vertices. Each resulting cluster is a prediction of how the group might divide.

Though the models and methods used in community detection allow for some degree of accuracy, they are not perfect and will have some unpredictable exceptions. One notable case which would not be predictable is the occurrence of a club member who was loyal to the side of the officer and was predicted to join the officer's faction. Due to external circumstances the group memeber ended up continuing his karate lessons with Mr. Hi instead. The models could not predict a situation like this and are not 100% accurate.

1.2 Question 2

1.2.1 The Problem

(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?

1.2.2 The Solution

Using python and igraph, a similar method of community detection can be utilized to make predictions about what would happen if the group was to divide into smaller groups composed of 3, 4, and 5 members. The same club members will be used so the vertices will be the same. Since the same member are being used the edges will be the same as well.

Predictions can be made using the community detection methods from question one which produced dendrograms as output when given the dataset . By Examining these graphs it is possible to determine the prediction of what groups would look like with 3,4, or 5 members. The method used for prediction make several assumptions, the first is that groups generally strive to have as many members as possible so 5 would be ideal however some groups will have only 3 or 4. The second assumption is that groups are composed by the lowest possible level of the dendrogram without being larger than five. Given those assumptions and using the walktrap method from question one predictions can now be made.

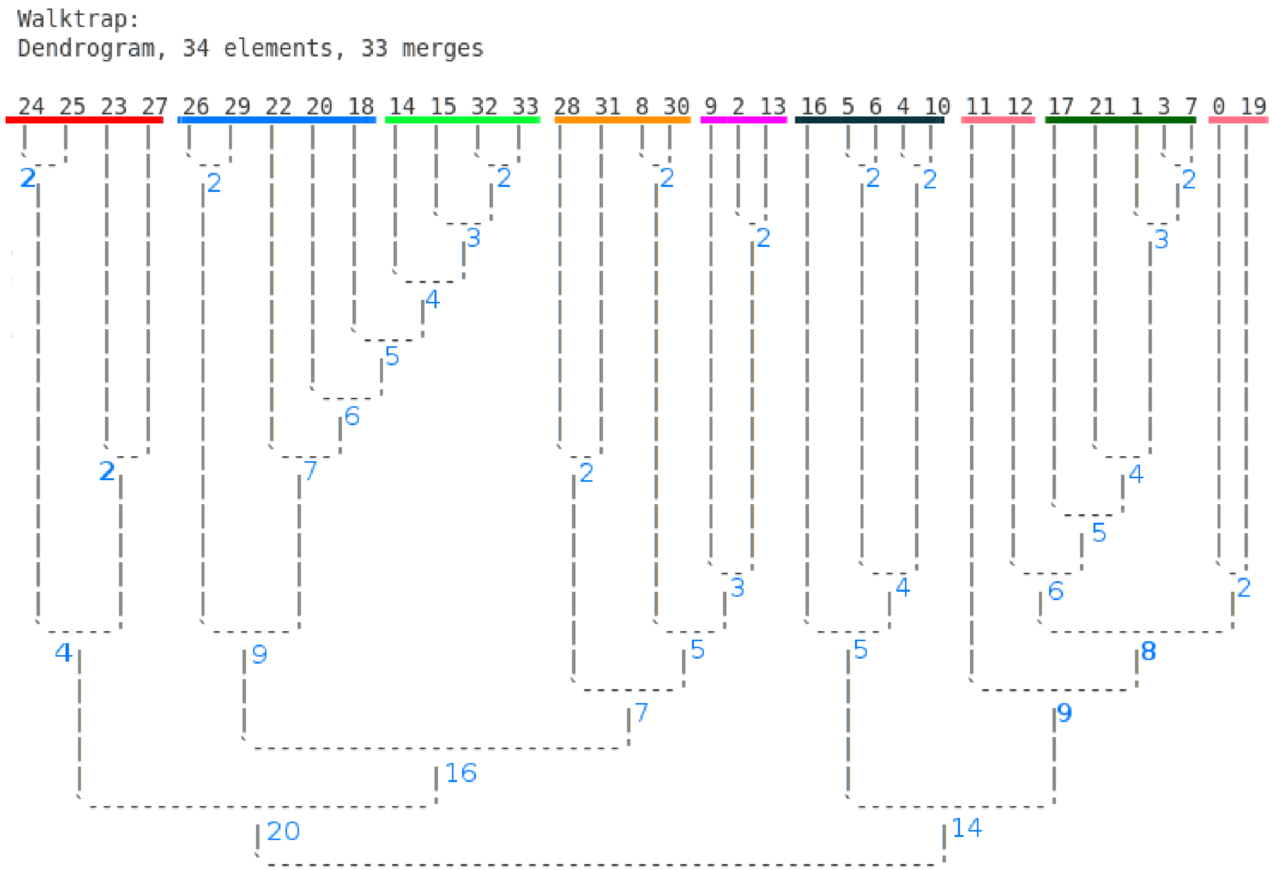

Figure 4: Grouping by Walktrap Dendrogram

Figure 4 shows the outcome of using the walktrap method in addition to the two assumptions made. The result is subgroups of the original group divided up by a method similar to how they were divided between two groups. The dendrograms do a nice job ordering by group communication by default. By counting the weight of each tree branch and investigating level by level you can find a likely prediction by this method as to how the groups might look. My prediction would be the 8 groups designated by a color bar in figure 4. Note that members 11, 12, 0, and 19 form a group of four. This method could be applied to the edge betweeness and fast greedy dendrograms as well.

References

1. https://docs.python.org/2/
2. http://aris.ss.uci.edu/~lin/76.pdf
3. http://www-personal.umich.edu/~ladamic/courses/networks/si614w06/ppt/lecture18.ppt
4. http://clair.si.umich.edu/si767/papers/Week03/Community/CommunityDetection.pptx
5. http://networkx.github.io/documentation/latest/examples/graph/karate_club.html
6. http://nbviewer.ipython.org/url/courses.cit.cornell.edu/info6010/resources/11notes.ipynb
7. http://stackoverflow.com/questions/9471906/what-are-the-differences-between-community-detection-algorithms-in-igraph/9478989#9478989
8. http://stackoverflow.com/questions/5822265/are-there-implementations-of-algorithms-for-community-detection-in-graphs
9. http://konect.uni-koblenz.de/networks/ucidata-zachary
10. http://stackoverflow.com/questions/5822265/are-there-implementations-of-algorithms-for-community-detection-in-graphs