Assignment 5

David Bayard

March 24, 2019

QUESTION 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?

Solution:

This paper will demonstrate that a mathematical model or algorith, Girvan Newman in this case, could have been used to predict the results of the Karate Club split.

To start off, a description of the Girven Newman alrorithm must be provided. In simplest terms, the Girven Newman alrogith works by finding the edge with the highest central betweenness. This edge is then removed, and the central betweenness of each edge is recalculated. This process is repeated until two separate components are recognized. There is a library in Python called networkx, which provides functions that find the highest central edge betweenness, and remove specific edges, among other things. This library was used to graph the Karate Club before the split, and to iteratively remove edges from the Karate Club graph.

In more formal terms, the formula to calculate the central edge betweenness for the vertex is:

$$C_B(v) = \sum_{(s \neq v \neq \ t \epsilon V)} \frac{\sigma_{st}(v)}{\sigma st}$$

Where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v.

This formula provides access to the edge with the higest betweenness, which will be removed repeatedly to disconnect the communities in the graph. The functions used to calculate the central edge betweenness and remove the edge are displayed below, as well as a sorting method. This is used to obtain the element with the highest central betweenness value, rather then the one with the lowest edge betweenness.

```python
eb = nx.edge_betweenness_centrality(G0)

    sortedDict = (sorted(eb.items(), key=lambda key: key[1], reverse=True))


    return sortedDict[0][0]
```
Listing 1: Finding central edge betweenness

Once the edge with the highest betweenness value found, a function is called to remove the edge, and this is repeated until the graph is split into two communities. This mimics the real life Karate Club split because an edge between any two nodes represents an association between two members of the Karate Club, whom interact outside of the club. As a result, the graph contains nodes and edges which represent club members, and the people whom they directly interact with. Thus, removing the edge with the highest betweenness is representative of removing the individual whom was most involved in both Mr.Hi's group and John. A's group. The code to split the communities is shown below:

```python
while sum(1 for x in components) == 1:

```

```
3    # Remove Original Node
4    formattedGraph.remove_edge(*find_best_edge(formattedGraph))
5
6    components = nx.connected_component_subgraphs(formattedGraph)
7
8  plt.show()
9
10  result = [c.nodes() for c in components]
11
12  for c in components:
13    result.extend(girvan_newman(c))
```

Listing 2: Extracting Friend Count

On pages 4 and 5, images displaying the removal of edges with the higest central betweenness values are displayed. The first image represents the orginal Karate Club Graph, where each node represents a person in the Karate Club. For the case of this example, node 0 will represent Mr. Hi, and node 33 represents the president of the Karate Club, John A. Every node is connected by an edge, which, as stated earlier, represent an association between two nodes, or people, outside of the Karate Club.
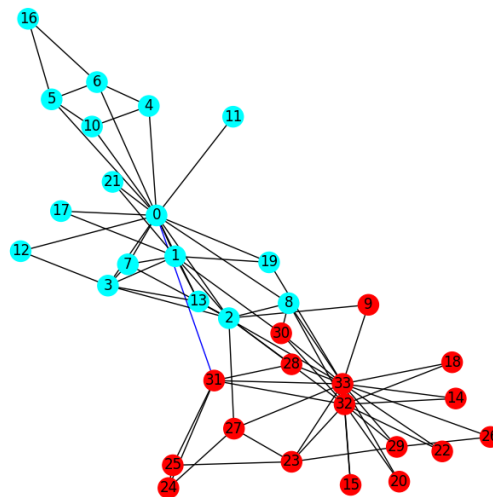
(a) Initial Graph

The graph above is the first representation of the Karate Club graph before the Girvan-Newman method is graphed. In this illustration, there is an edge colored in blue, which is the edge with the highest central betweenness value. This edge will be removed, and a new graph will be constructed, and this process will be repeated until two the two seperate communities are formed. This graph is produced using the networkx function below.
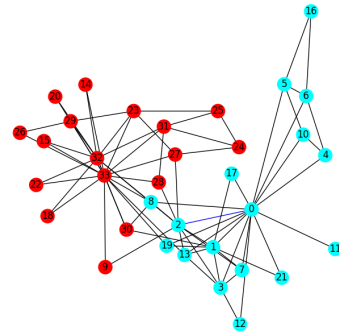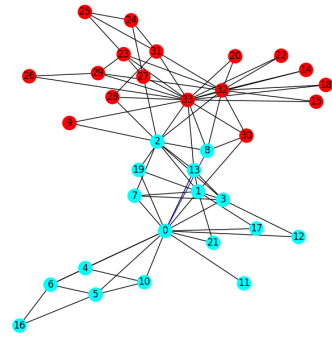
```
1  G=nx.karate_club_graph()
```
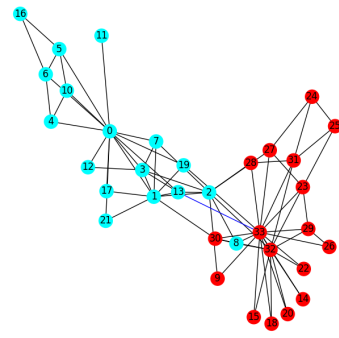
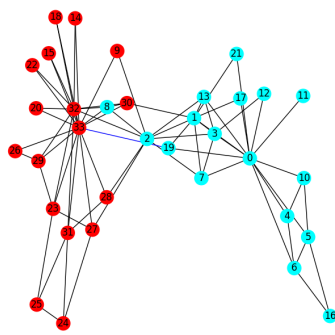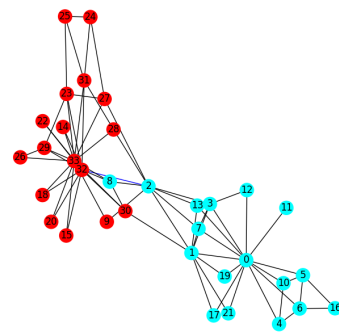Listing 3: Karate Club Graph
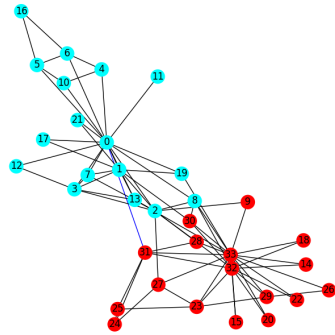
(a) Iteration 1

(b) Iteration 2

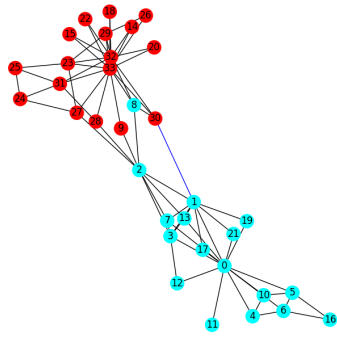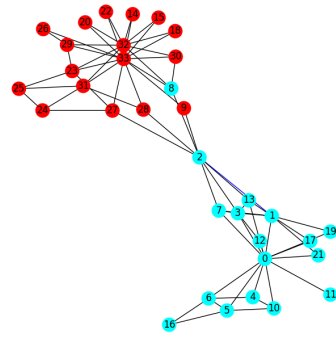(c) Iteration 3

(d) Iteration 4

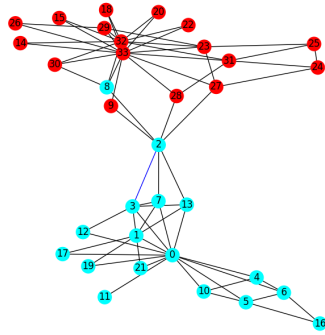(e) Iteration 5

(f) Iteration 6
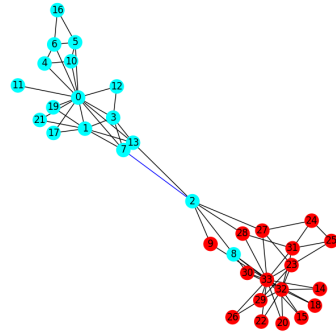
Figure 0.2: Splitting Graph Part1
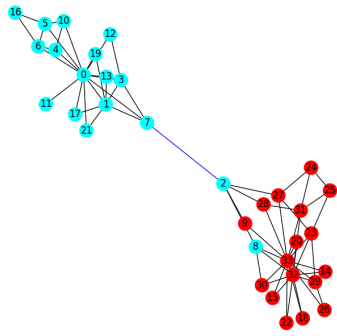
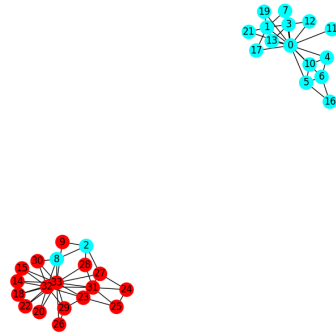(a) Iteration 7

(b) Iteration 8

(c) Iteration 9

(d) Iteration 10

(e) Iteration 11

(f) Fully Split

Figure 0.3: Splitting Graph Part 2

By examining the final product, which display the split communities, it is evident that the graph could have been used to predict the Karate Club split. This evident, because the final graph only differs from the real Karate Club split with regards to node 2. This node should belong to Mr.Hi's group, but instead exists within John A's group. With only this point of infliction, the Girvan-Newman algorithm provides 33 correct predictments out of 34. This results in 97% of the predictions being accurate, concluding that it is indeed possible to predict the Karate Club split with marginal error.

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?
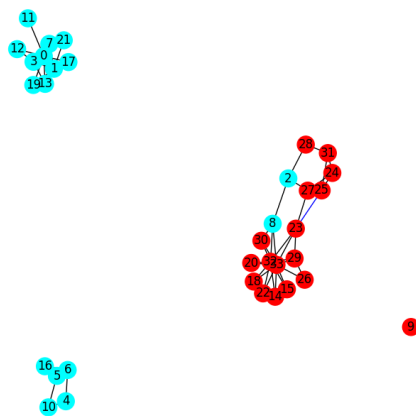
Solution:

This is determined by running the Girvan-Newman algorithm until the community breaks into 3, 4, and 5 respective communities. The same process will be repeated from splitting the same communities, which will provide an accurate representation of the club after the splits.
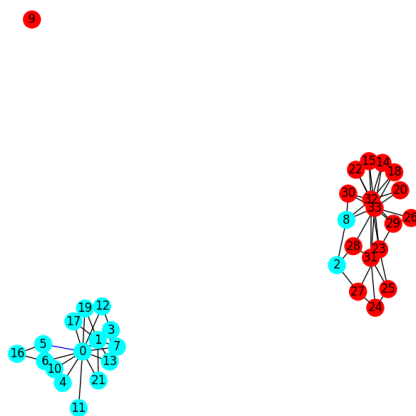
(a) 3 Groups
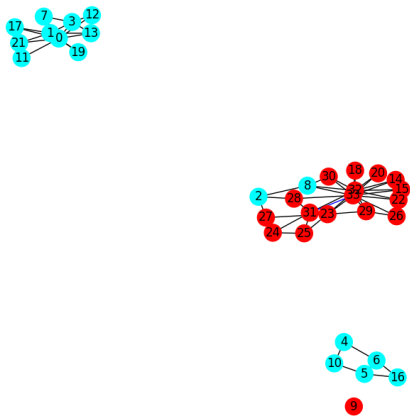
(b) 4 Groups

(c) 5 Groups

Figure 0.4: Graphs for groups 3, 4, and 5