INTRO. TO WEB SCIENCE: CS 532: A5

Due on Thursday, March 16, 2017

Dr. Nelson

Udochukwu Nweke

Contents

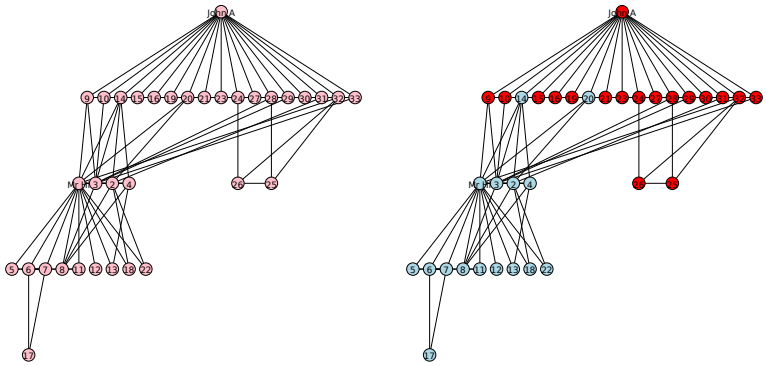
Problem 1

Listing 1: Girvan and Newman Algorithm Implementation Code

```
import igraph


def setEdgeProps(graph, color):

    for edges in graph.es:
        edges['edge_color'] = color
        graph.es["edge_width"] = 1

    return graph

def GirvinNewmanAlg(graph, maxClusterCount):

    vertex_colors = {0: 'black', 1: 'light blue', 2: 'red',
    3: 'green', 4: 'pink', 5: 'brown'}
    clusters = graph.clusters('weak')
    clusterCount = len(clusters)
    iterationCounter = 1

    while graph.ecount() > 0 and clusterCount < maxClusterCount:

        clusters = graph.clusters('weak')
        clusterCount = len(clusters)

        for cluster in clusters:
            for vertex_index in cluster:
                graph.vs[vertex_index]['vertex_color'] = 0

        style = {}
        style['vertex_color'] = [vertex_colors[node['vertex_color']]
        for node in graph.vs]


        edgeBetweenness = graph.edge_betweenness()


        indexOfEdgeWithMaximumBetweenness = max(xrange(len(edgeBetweenness)),
       key = edgeBetweenness.__getitem__)


        graph.es[indexOfEdgeWithMaximumBetweenness]['edge_color'] = 'gold'
        graph.es[indexOfEdgeWithMaximumBetweenness]['edge_width'] = 5
        drawGraph(graph, vertex_colors, str(iterationCounter)+ '-' +
       str(clusterCount) + '.pdf' )


        graph.delete_edges(indexOfEdgeWithMaximumBetweenness)
        iterationCounter = iterationCounter + 1

def drawGraph(graph, vertex_colors, outfilename):
```

Problem 1 continued on next page...

Figure 1: Girvan and Newman Karate club split graph

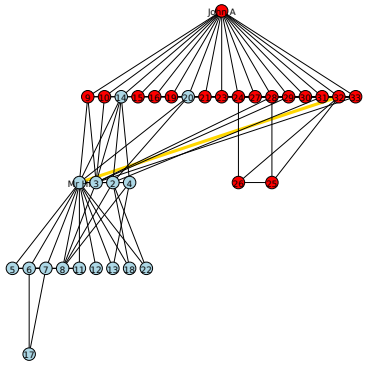


(a) Karate club graph before split (b) Karate club graph after split

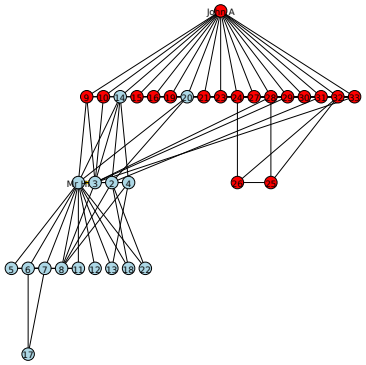
```
    style = {}

    layout = graph.layout('rt')
    style['layout'] = layout
    style['margin'] = 25

    style["vertex_color"] = [vertex_colors[node['Faction']] for node in graph.vs]
    style['edge_color'] = [edgeColor for edgeColor in graph.es['edge_color']]
    style['edge_width'] = [edgeWidth for edgeWidth in graph.es['edge_width']]
    style['vertex_label'] = graph.vs['name']

    igraph.plot(graph,'./graph_plots/' + outfilename, **style)

karateClub = igraph.Graph.Read_GraphML('karate.graphml')
karateClub = setEdgeProps(karateClub, 'black')

drawGraph(karateClub, {1.0: 'pink', 2.0: 'pink'}, 'karateClub.beforesplit.pdf')
drawGraph(karateClub, {1.0: 'light blue', 2.0: 'red'},
 'karateClub.postsplit.pdf')
GirvinNewmanAlg(karateClub, 5)
```

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

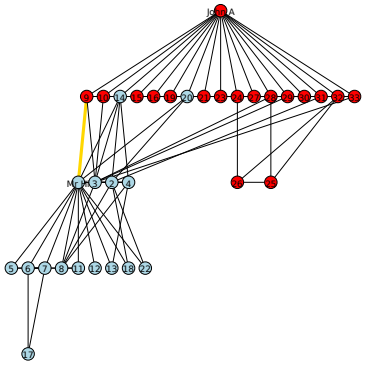
Figure 2: Iterative application of Girvin and Newman Algorithm



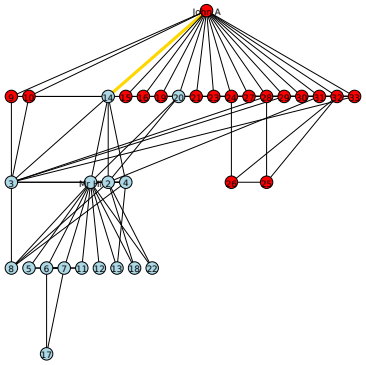
(a) G-N-Iteration1-subgraph1



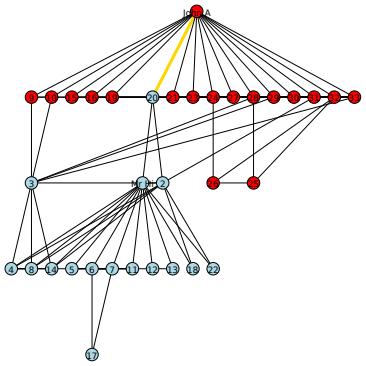
(b) G-N-Iteration2-subgraph1



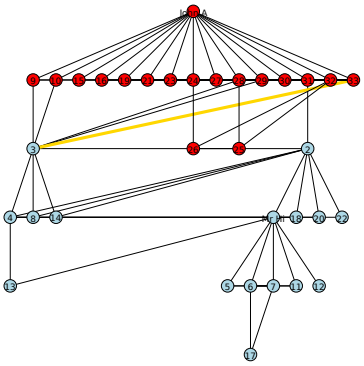
(c) Girvan-Newman Iteration3-subgraph1



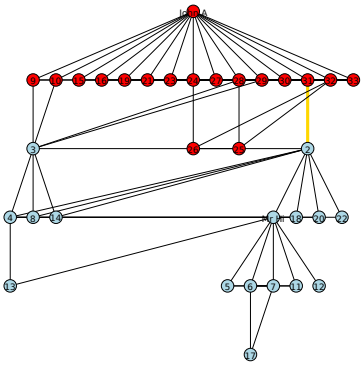
(d) G-N-Iteration4-subgraph1



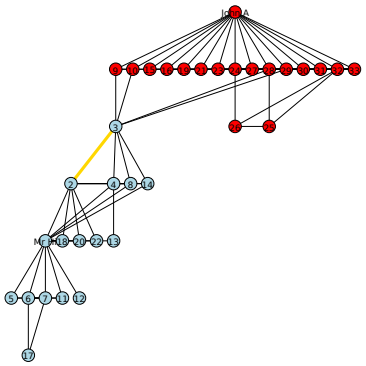
(e) G-N-Iteration5-subgraph1



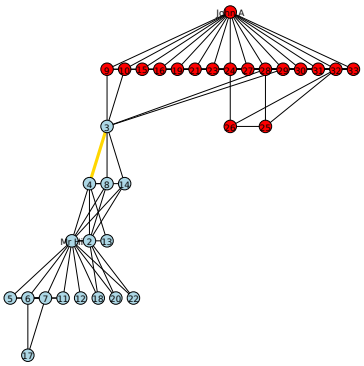
(f) G-N- Iteration6-subgraph1



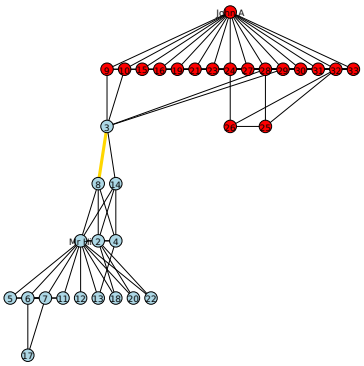
(g) G-N-Iteration7-subgraph1



(h) G-N-Iteration8-subgraph1

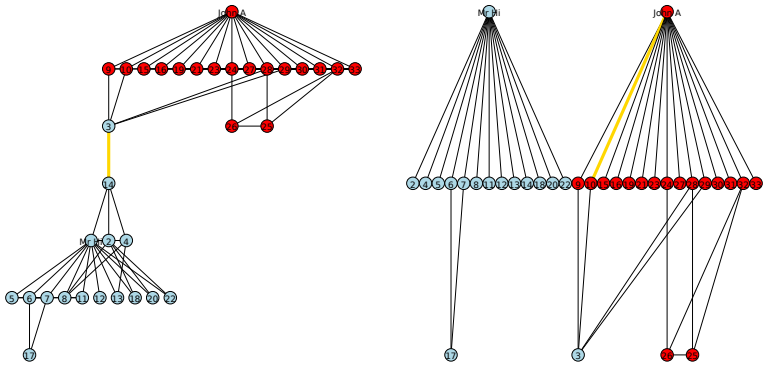


(i) G-N-Iteration9-subgraph1



(j) G-N-Iteration10-subgraph1

Figure 3: Iterative application of Girvin and Newman Algorithm Graph Split into Two Sub graphs



(a) G-N-Iteration11-subgraph1 (b) G-N-Iteration12-subgraph2

* 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-dete
9478989#9478989

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

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

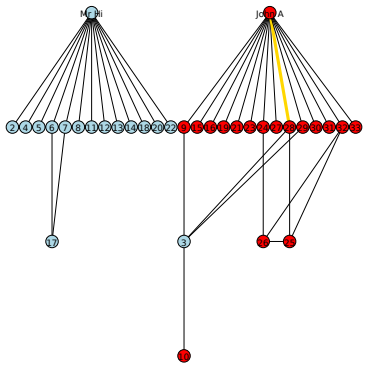
Solution 1:

Karate club contains 34 club members with Mr. Hi as the club instructor and John A. as the club president. After the disagreement between Mr. Hi and John A., the club split into 2 different groups, leaving Mr. Hi as the head of one group and John A. as the head of the second group.

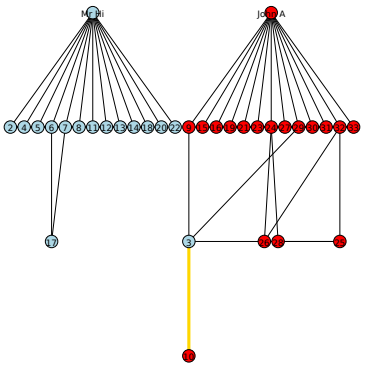
In order to prove if split could be predicted based on the weighted graph of social interactions, let us consider centrality of the graph. According to Girvan and Newman, The edge betweeness centrality is the shortest path that goes through a graph or network. The centrality gives the measure of important nodes in a graph. If the Zarachy Karate club graph shows two or more important central nodes, this could indicate the possibility of the split.

Girvan and Newman Algorithm states as follows:

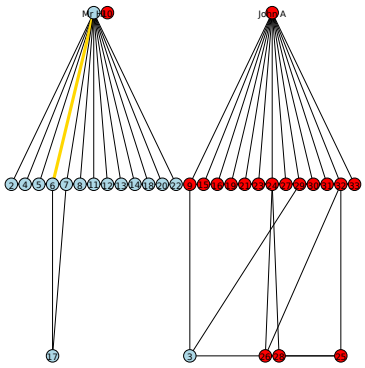
Figure 4: Iterative application of Girvin and Newman Algorithm



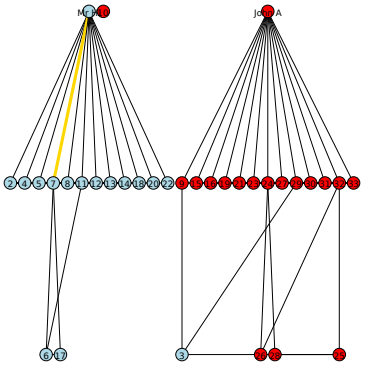
(a) G-N-Iteration13-subgraph2



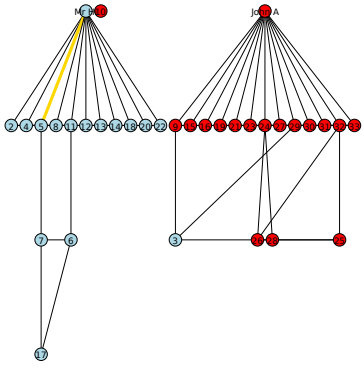
(b) G-N-Iteration14-subgraph2



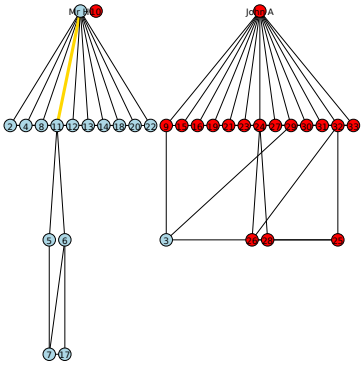
(c) G-N-Iteration15-subgraph3



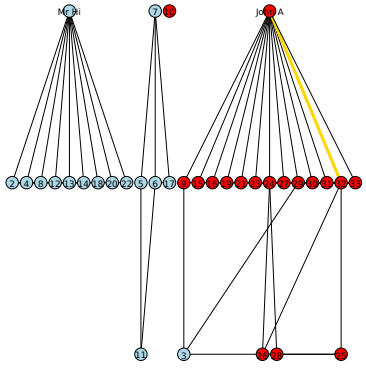
(d) G-N-Iteration16-subgraph3



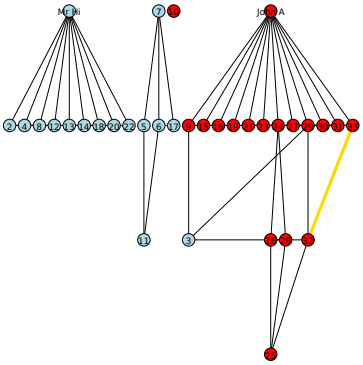
(e) G-N-Iteration17-subgraph3



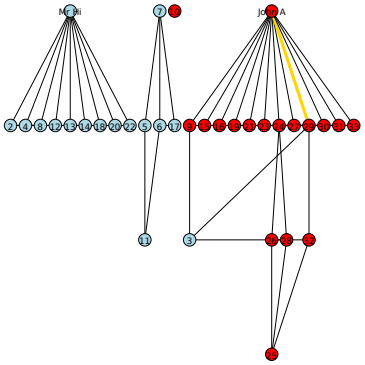
(f) G-N-Iteration18-subgraph3



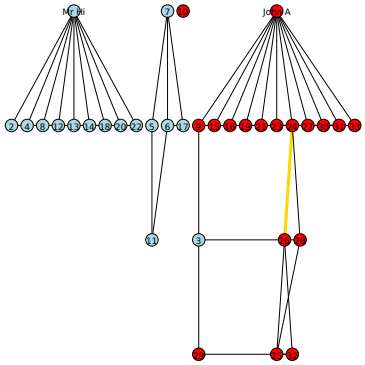
(g) G-N-Iteration19-subgraph4



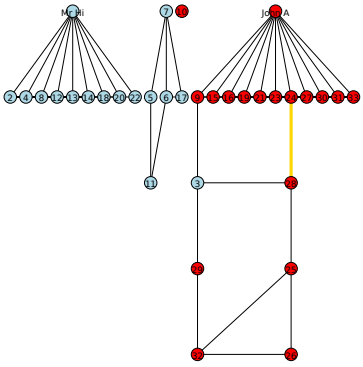
(a) G-N-Iteration20-subgraph4



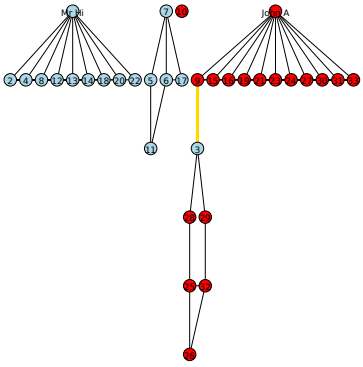
(b) G-N-Iteration21-subgraph4



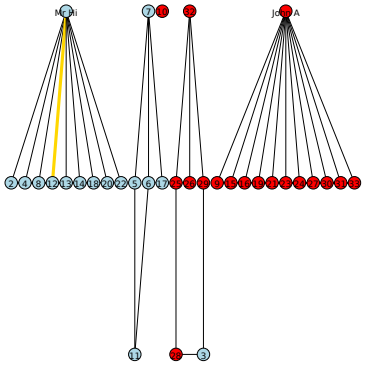
(c) G-N-Iteration22-subgraph4



(d) G-N-Iteration23-subgraph4



(e) G-N-Iteration24-subgraph4



(f) G-N-Iteration25-subgraph5

1. Calculate the betweenness for all edges in the network.
2. Remove the edge with the highest betweenness.
3. Recalculate betweennesses for all edges affected by the removal.
4. Repeat from step 2 until no edges remain.

I implemented the Girvan-Newman algorithm iteratively to see if I could possibly get the split showing different subgraphs . Specifically, the John A and Mr. Hi subgraphs. *function GirvinNewmanAlg* in listing 1 shows how I implemented Girvan-Newman algorithim by iteratively removing the edge with the highest betweeness until the graph split into 2 subgraphs. The result of the first split is seen in Figure 3(b) G-N-Iteration12-subgraph2

Comparing my result with Zachary's model: From my implementation and graph in Figure 3(b) G-N-Iteration12-subgraph2 I missed node 3 and Zachary missed node 9. This shows that Zachary Karate club split is possible with some level of accuracy.

Problem 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?
Solution 2:

Since I already predicted that Zachary karate club can be split into 2 by using Girvan and Newman's algorithm, *function GirvinNewmanAlg* in listing 1, I applied the same method to see if the graph could be

split into 3, 4, and 5 different subgraphs.
After iteratively applying Girvan and Newman Algorithm, I was able to get the different subgraphs.

The graph after 3 split is seen in Figure 4(c) G-N-Iteration15-subgraph3
The graph after 4 split is seen in Figure 4(g) G-N-Iteration19-subgraph4
The graph after 5 split is seen in Figure 4(f) G-N-Iteration25-subgraph5

References

[1] Igraph. http://lists.nongnu.org/archive/html/igraph-help/2008-11/msg00047.html. Accessed: 2017-10-03.

[2] An Information Flow Model for Conflict and Fission in Small Groups. Accessed: 2017-10-03.

[3] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[4] LongJason Lu and Minlu Zhang. Edge betweenness centrality. In *Encyclopedia of systems biology*, pages 647–648. Springer, 2013.