Web Science cs532: Assignment #5

Due on Thursday, March 3, 2016

Dr.Michael.L.Nelson 4:20pm

Zetan Li

Contents

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-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`

<div align="center">SOLUTION</div>

First, we have to get the raw data of karate club network from the official data site of igraph called Nexus.`http://nexus.igraph.org/api/dataset_info?id=1&format=html`
Then plot the original network out.

Then in the original data file (karate.graphml), we find there is a data field called "Faction". This indicates the original subgroup architecture. So we set the node to different colour according to the faction number they have. We will use this to compare with the prediction made by algorithm.

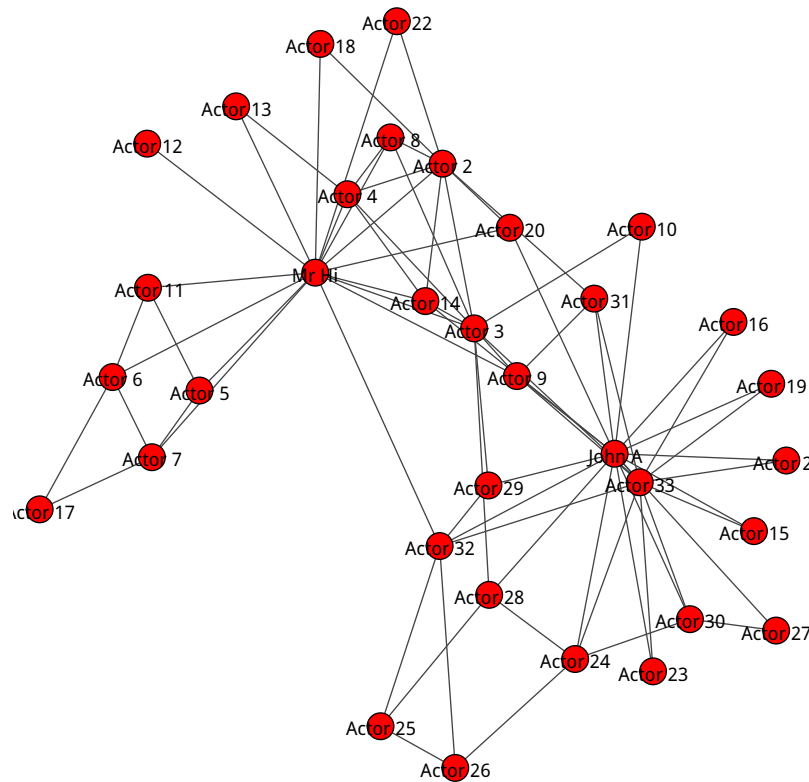Figure 1: original karate club network

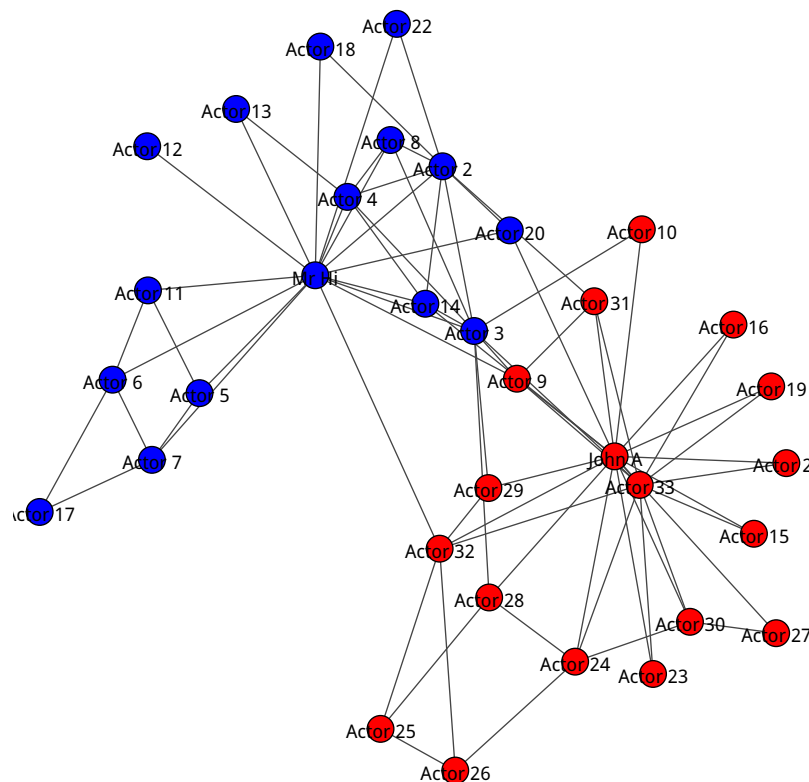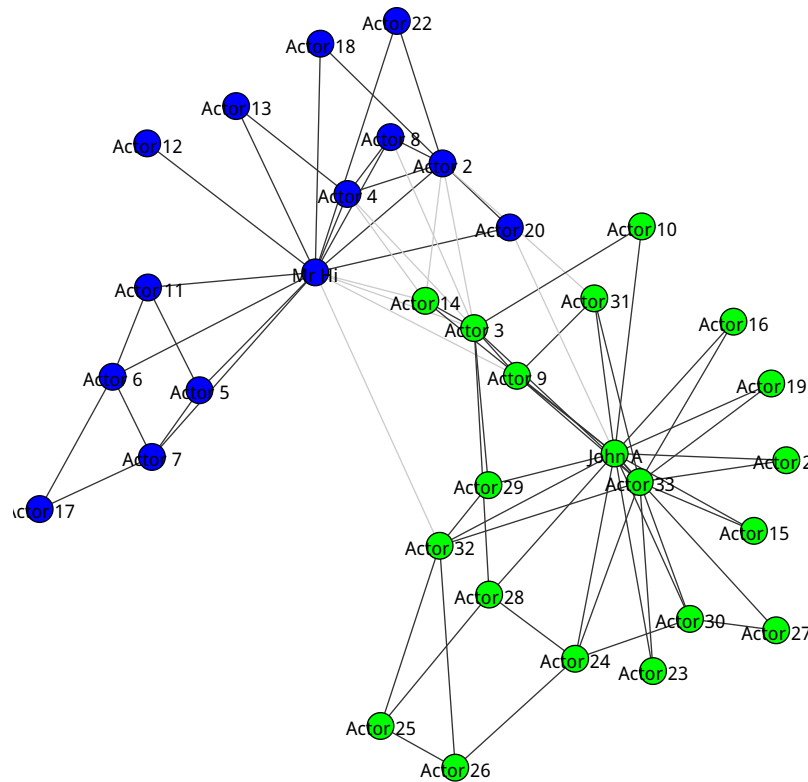Figure 2: original karate club network subgroups

Figure 3: Predicted karate club network

Now comes the betweenness algorithm, the igraph library has this feature in the graph module. Above is the graph output split by the betweenness algorithm.

Listing 1: python code to plot original graph and predicted graph

```python
from igraph import *
g=Graph.Read_GraphML("karate.GraphML")
layout=g.layout('fr')
plot(g,'p1_org.pdf',layout=layout,vertex_label=g.vs['name'])
color_table={1:"blue",2:'red'}
g.vs['color']=[color_table[group] for group in g.vs['Faction']]
plot(g,'p1_org_split.pdf',layout=layout,vertex_label=g.vs['name'])

cut_graph=g.community_edge_betweenness(clusters=2,directed=False,weights=g.es['
    weight'])
cutted_g=cut_graph.as_clustering()

plot(cutted_g,"p1_cut.pdf",layout=layout,vertex_label=g.vs['name'])
```

From the predicted graph, we can see that there are only two node in wrong group: Actor14 and Actor3. So the total accuracy is 94.11%

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

The betweennes algorithm in igraph have an option called 'clusters', which enable us to specify the certain number of subgroups we want in splitting.

Listing 2: python code to plot predicted graph in 3 4 5 subgroups

```python
from igraph import *
g=Graph.Read_GraphML("karate.GraphML")
layout=g.layout('fr')

for c in range(3,6):
    cut_graph=g.community_edge_betweenness(clusters=c,directed=False,weights=g.es
        ['weight'])
    cutted_g=cut_graph.as_clustering()

    plot(cutted_g,"p2_cluster"+str(c)+".pdf",layout=layout,vertex_label=g.vs['
        name'])
```

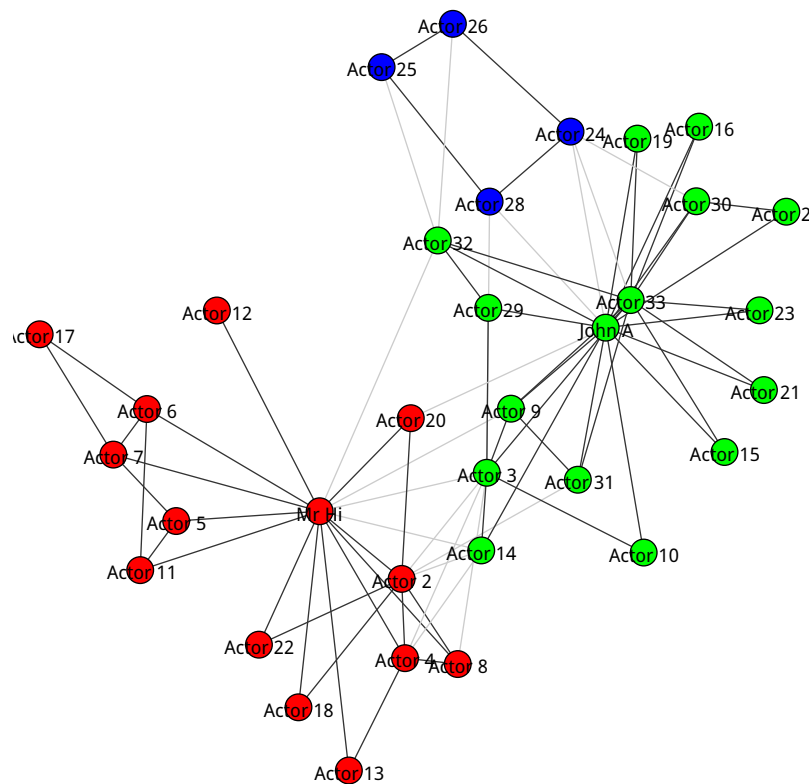Figure 4: Predicted karate club network in 3 subgroups
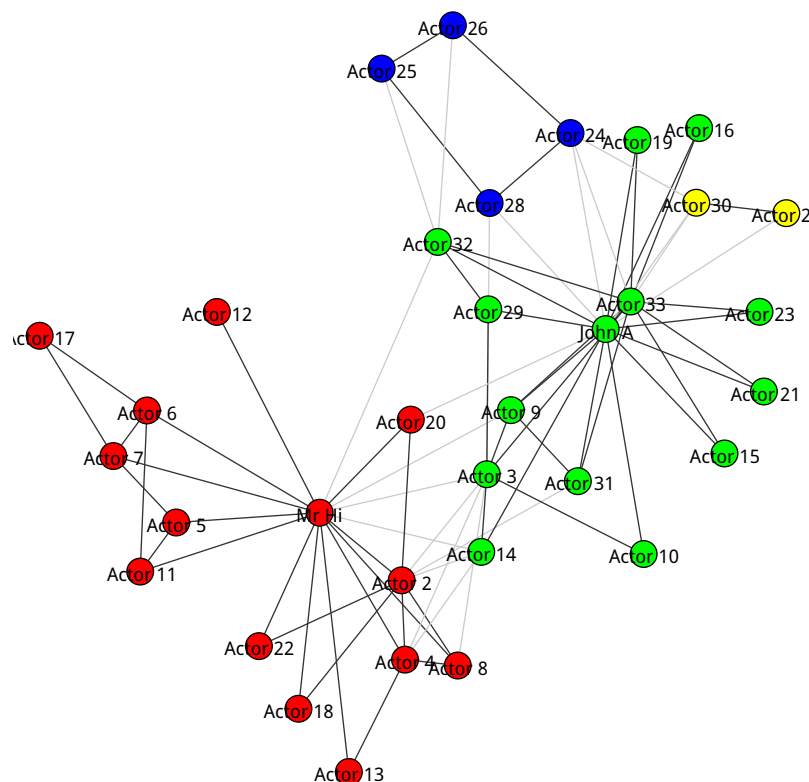
Figure 5: Predicted karate club network in 4 subgroups

Figure 6: Predicted karate club network in 5 subgroups