# Complex Network - Assignment 1

NGUYEN T. Hoang - SID: 15M54097

Fall 2015, W833 Mon. Period 5-6

Due date: 2015/11/30

## Problem

1. Visualize the network of Zachary's karate club. (GML file is available at http://www-personal.umich.edu/ mejn/netdata/).
2. Select two central vertices. Why do you think they are central?
3. Show the diameter, density, average path length, and clustering coefficient of the (undirected) network.
4. Draw a degree distribution (a histogram of the degrees of vertices) of the network.
5. Select two vertices whose PageRank values are the highest.
6. Divide the network into small groups and answer its modularity.

# Answer

**1.** Zachary's karate club undirected network is visualized using Gephi [1]. The nodes in the graph are labeled by its id in the GML file, and they are also colored by the node degree (deeper blue indicates higher node degree).

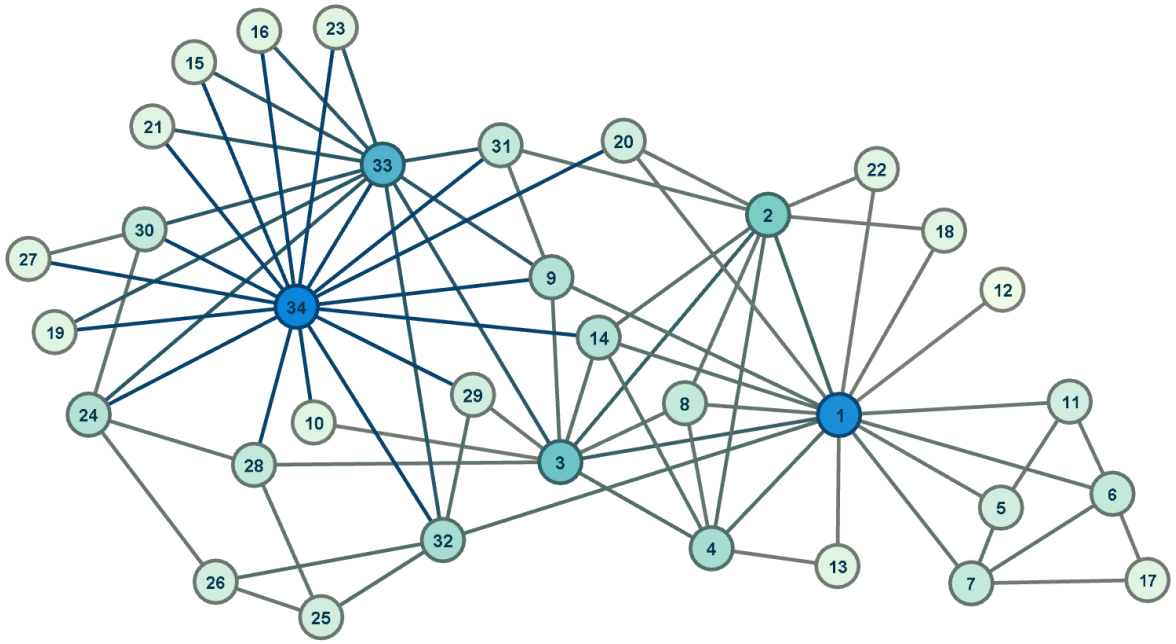


Figure 1: *Visualization of Zachary's karate club network. Layout: Yifan Hu. Colored by edge degree.*

**2.** Figure 1 shows that there is two nodes with highest degree: **node 1** and **node 34**. I think that in term of *intuitive centrality*, node 1 and node 34 are the top-2 most central nodes. In the network, node 1 and node 34 have the highest number of connection to other nodes. By observation, we can also see that any other nodes in the network must have a direct connection to node 1 or node 34. Therefore, the two most central nodes in the network are node 1 and node 34.

Listing 1: Create the graph in SNAP.PY and print top-k highest degree centrality node.

```
# Extracted from UnweightedUndirectedGraph class - File: get_data.py
...
self._graph = snap.LoadEdgeList(snap.PUNGraph, edge_list_file, 0, 1, '␣')
...
# Get the degree list
DegreeCentr = {}
for NI in self._graph.Nodes():
  deg = snap.GetDegreeCentr(self._graph, NI.GetId())
  DegreeCentr[NI.GetId()] = deg
sorted_ranking = sorted(ranking.items(), key=operator.itemgetter(1), reverse=True)
return sorted_ranking[0:1]
```

**3.** The metrics of the given network is summarized as below: [2]

- Diameter of the network: 5
- Density: $\approx 0.14$
- Average path length: $\approx 2.41$
- Clustering coefficient: $\approx 0.57$

Listing 2: Graph metrics properties in SNAP.PY

```
...
# Create a container for the graph
karate_graph = UnweightedUndirectedGraph(edge_list, 'karate')
...
# Python CLI
>>> karate_graph.approx_diameter(20) # 20 is number of sample
5
>>> graph.density
0.13903743315508021
>>> graph.avg_path_length
2.408199643493761
>>> graph.get_quick_clust_coeff()
0.5706384782076823
```

**4.** The graph degree distribution visualization is obtained by creating frequency value *csv* file and plot with Google Spreadsheet.

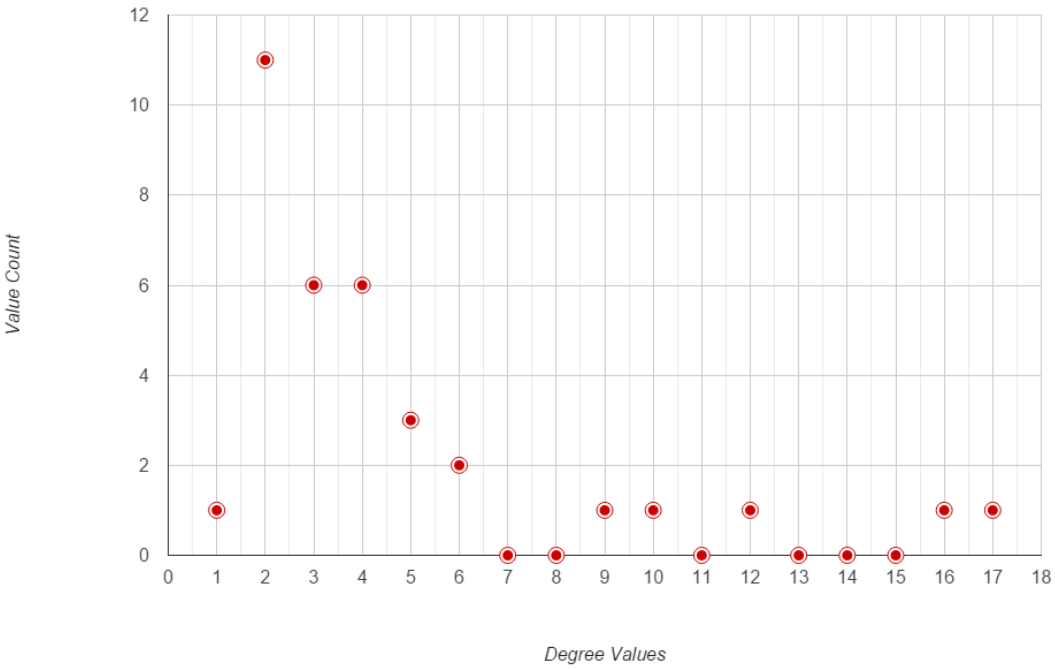


Figure 2: *Degree distribution in the Zachary's Karate network.*

The highest degree value in the network is 17 (1 node) and the lowest value is 1 (1 node). The most frequent degree value in the network is 2 with 11 nodes. For a social network in a sport club setting, I believe this distribution is natural.

**5.** Two vertices with the highest PageRank values are: [2]

- Node 34 - PageRank value ≈ 0.100
- Node 1 - PageRank value ≈ 0.097

Listing 3: Graph PageRank calculation

```
# Python CLI
>>> pr_dict_sorted = karate_graph.rank_pagerank()
>>> pr_dict_sorted[0] # 0 is the highest score PageRank
(34, 0.10091876156358327)
>>> pr_dict_sorted[1] # 1 is the second highest score PageRank
(1, 0.09700636750128044)
```

**6.** I ran the Girvan-Newman community detection algorithm with resolution 1.0 and 2.0 to obtain 2 different graph partitionings as in figure 3 and figure 4.

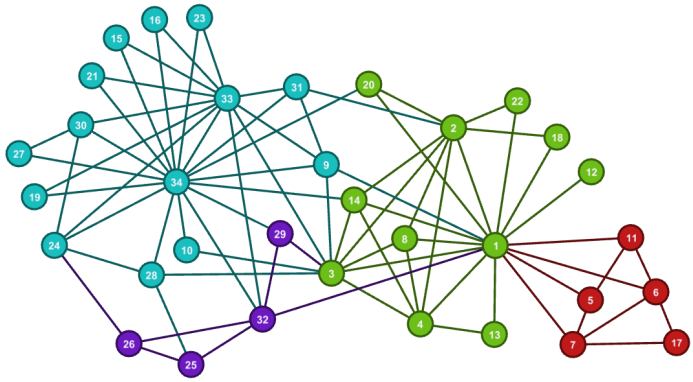

Figure 3: *Girvan-Newman community output. Resolution = 1.0.* ***Modularity = 0.415***

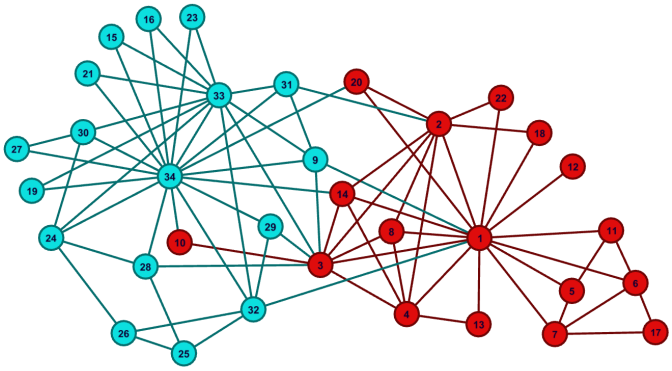

Figure 4: *Girvan-Newman community output. Resolution = 2.0.* ***Modularity = 0.371***

# References


[1] Gephi.org. Gephi - the open graph viz platform. http://gephi.github.io/. Accessed: 2015-11-25.

[2] Nguyen, H. Zachary snap.py. https://github.com/gear/zachary-snap.py. Accessed: 2015-11-29.