# CS529 - Assignment 3

**Kutay Taşcı**

**22101359**

## Overview

In this assignment we are going to analyze karate club dataset, as given by Wayne Zachary in 1977. The dataset consists of 34 nodes, nodes in this dataset represents the members of a karate club, and links represent the interactions between these members. We are going to perform analysis on this dataset. For analysis we are going to use network and node level metrics to understand the given network.

This assignment consists of three experiments. In first experiment, we analyzed network and node level metrics. We will use these metrics to quantify the features of the nodes. In second experiment, we visualized the network to have a visual understanding of given network. Lastly, we are going to analyze our quantitative and visual results of our experiments.

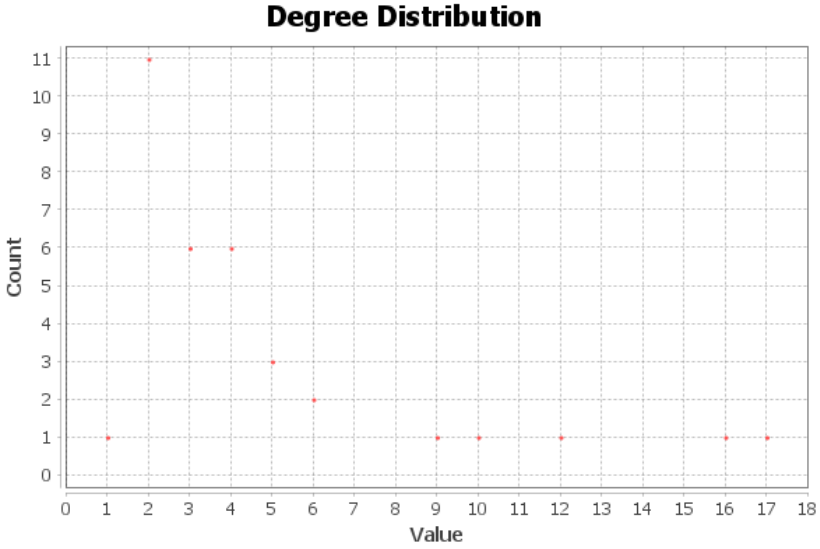
## Exercise 1: Compute Network and Node Level Metrics

In this experiment we are going to extract following metrics from the dataset:

- Degree Distribution
- Network Density
- Degree Centrality
- Betweenness Centrality
- Closeness Centrality
- Page Rank Centrality

You can find all the metrics extracted as a table in the file **Node_Features_KutayTasci_A3.csv** in the submission.

**A-** You can find the **degree distribution** in the plot bellow; average degree of the network is 4.58.

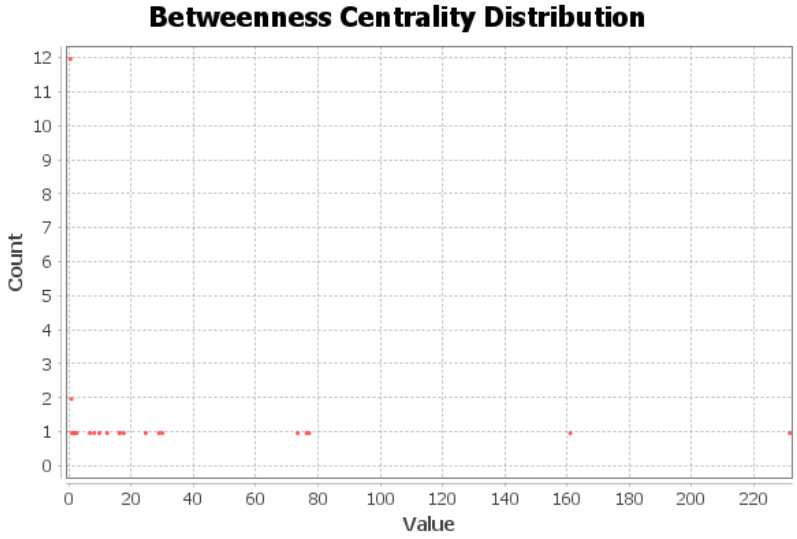


**B-** **Network density** of the karate club network is **0.139**

**C-** **Degree centrality** metric is the number of edges in a node. Here are the top 10 nodes with highest degree centrality.

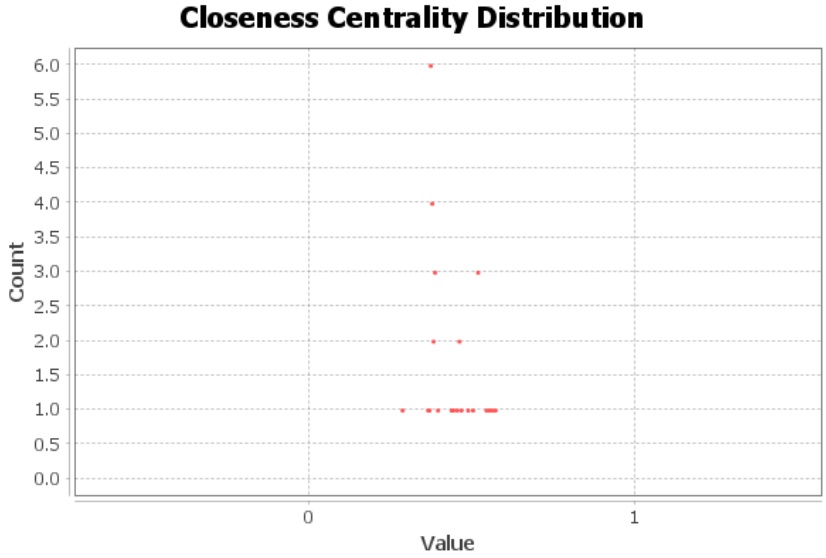
| Node_Id | Degree Centrality |
|---|---|
| 34 | 17 |
| 1 | 16 |
| 33 | 12 |
| 3 | 10 |
| 2 | 9 |
| 4 | 6 |
| 32 | 6 |
| 9 | 5 |
| 14 | 5 |
| 24 | 5 |

**D-** **Betweenness Centrality** is how often a node lies along the shortest path between two other nodes. Here is the visualization on the distribution and the top 10 nodes with highest betweenness centrality.



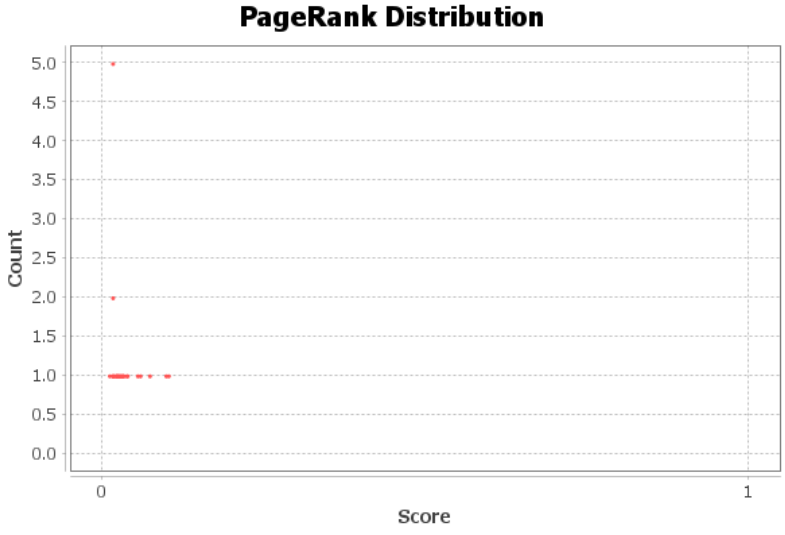

| Node_Id | Betweenness Centrality |
|---|---|
| 1 | 231.071429 |
| 34 | 160.551587 |
| 33 | 76.690476 |
| 3 | 75.850794 |
| 32 | 73.009524 |
| 9 | 29.529365 |
| 2 | 28.478571 |
| 14 | 24.215873 |
| 20 | 17.146825 |
| 6 | 15.833333 |

**E-** **Closeness centrality** measures the shortest distance between a given node to all other nodes in the network. Higher values indicate that given node is closer to every other node in the network. Here is the visualization on the distribution and the top 10 nodes with highest closeness centrality.




| Node_Id | Closeness Centrality |
|---|---|
| **1** | **0.568966** |
| **3** | **0.559322** |
| **34** | **0.55** |
| **32** | **0.540984** |
| **33** | **0.515625** |
| **9** | **0.515625** |
| **14** | **0.515625** |
| **20** | **0.5** |
| **2** | **0.485294** |
| **4** | **0.464789** |

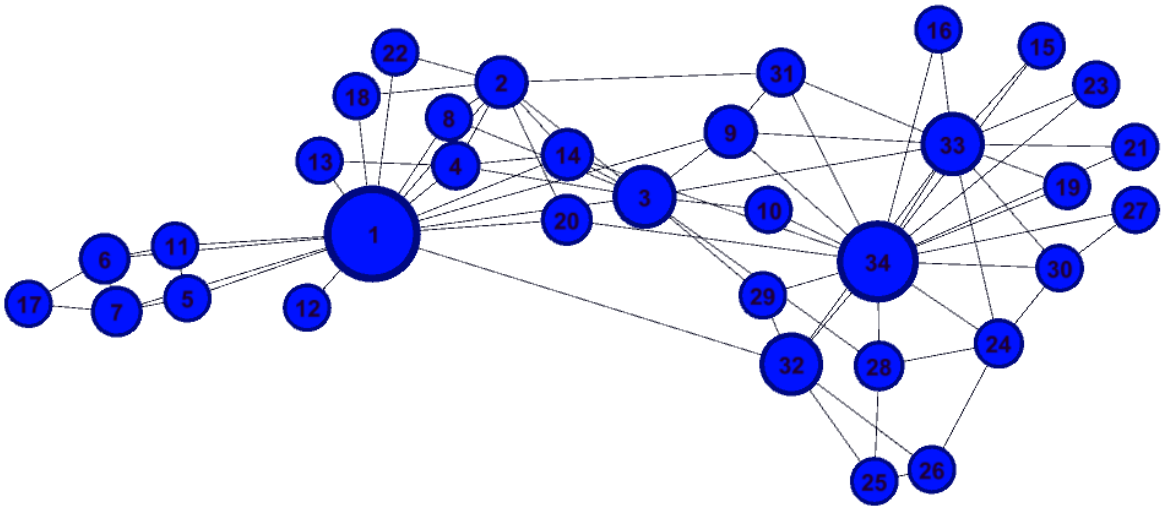
**F-** **Page rank centrality** can be defined as importance of a node based on the importance of its incoming links. This metric is derived from the PageRank algorithm proposed by Google. Higher values mean higher importance. Here is the visualization on the distribution and the top 10 nodes with highest closeness centrality.



| Node_Id | PageRank Centrality |
|---|---|
| 34 | 0.100879 |
| 1 | 0.097001 |
| 33 | 0.071663 |
| 3 | 0.057075 |
| 2 | 0.052872 |
| 32 | 0.037161 |
| 4 | 0.035862 |
| 24 | 0.031521 |
| 9 | 0.029768 |
| 14 | 0.029542 |

## Exercise – 2: Network Visualization

In this experiment, we are going to visualize the karate club network. In this visualization node sizes are adjusted based on betweenness centrality.



# Experiment – 3: Interpretation and Analysis

In this experiment we are going to interpret and analyze the results of centrality metrics. For each metric we are going to analyze top 3 nodes with highest centralities for each metric.

**Degree Centrality:**

| Node_Id | Degree Centrality |
|---|---|
| **34** | **17** |
| **1** | **16** |
| **33** | **12** |

As you can see top 3 nodes with highest degree centrality are 34, 1, and 33. These nodes have the greatest number of connections with other nodes. Since nodes correspond to people, these people are the most popular people among the karate club. They are known by most of the people.

**Betweenness centrality:**

| Node_Id | Betweenness Centrality |
|---|---|
| **1** | **231.071429** |
| **34** | **160.551587** |
| **33** | **76.690476** |

As you can see top 3 nodes with highest betweenness centrality are 1, 34 and 33. These nodes lie on the shortest paths between nodes. This can be interpreted, for communicating between any node to any other node, probability that these people lie on the best possible path is higher than other nodes.

**Closeness centrality:**

| Node_Id | Closeness Centrality |
|---|---|
| **1** | **0.568966** |
| **3** | **0.559322** |
| **34** | **0.55** |

As you can see top 3 nodes with highest betweenness centrality are 1, 3 and 34. These nodes are closer to other nodes in the network. So, we can say that they can reach all the people in the network easier than other nodes.

**Page rank centrality:**

| Node_Id | PageRank Centrality |
|---|---|
| **34** | **0.100879** |
| **1** | **0.097001** |
| **33** | **0.071663** |

As you can see top 3 nodes with highest betweenness centrality are 34, 1 and 33. According to the PageRank centrality these nodes have important incoming links. So, we can say that these nodes have important connections with other nodes.

**Overall Interpretation:**

According to the metrics we have analyzed. Most important nodes seem to be 1 and 34 since this node is always one of the top ones in each metric. Other than these 33 shows up on degree, betweenness and page rank centralities. Which means this node both many connections, also these connections to this node are important than others. For closeness centrality 3 appears on the list, it is not surprising because it is the $4^{th}$ node in every other centrality metric.

For this dataset quantitative metrics shows that 1, 3, 33 and 34 are the central nodes, that has great importance on this network. These nodes are probably the key people in the karate club. Such as teachers or popular people.