2018 Fall CSE 472 Social Media Mining Project I

Summary:

1. To create the social network of authors from Twitter, the Twitter data is scraped and saved. These authors are those who have tweeted about ‘Artificial Intelligence’ in Twitter for the period July 26, 2018 to August 26, 2018. Each node of the network represents a user and each edge is a follower-followee relation. The scraping is done using the following packages in Python:
    a. Pandas
    b. Selenium
    c. Beautiful Soup
2. After fetching the data, authorship graph is created using Gephi and NetworkX by specifying the input in edgelist format. **The Visualization of The Network** is obtained as a graph where the connections and clusters of the users are shown in the following sections.
3. The following **Network Measures** are calculated for the network and are then calculated and reported in the following sections:
    a. Degree distribution
    b. Clustering Coefficient
    c. PageRank
    d. Diameter
    e. Closeness Centrality
    f. Betweeness Centrality
    g. Eigenvector Centrality
    h. HITS
    i. Graph density
    j. Modularity

Visualization of the Network:

A twitter network of users talking about Artificial Intelligence

No of nodes: 107 **No of edges**: 210 **Graph Type**: Undirected

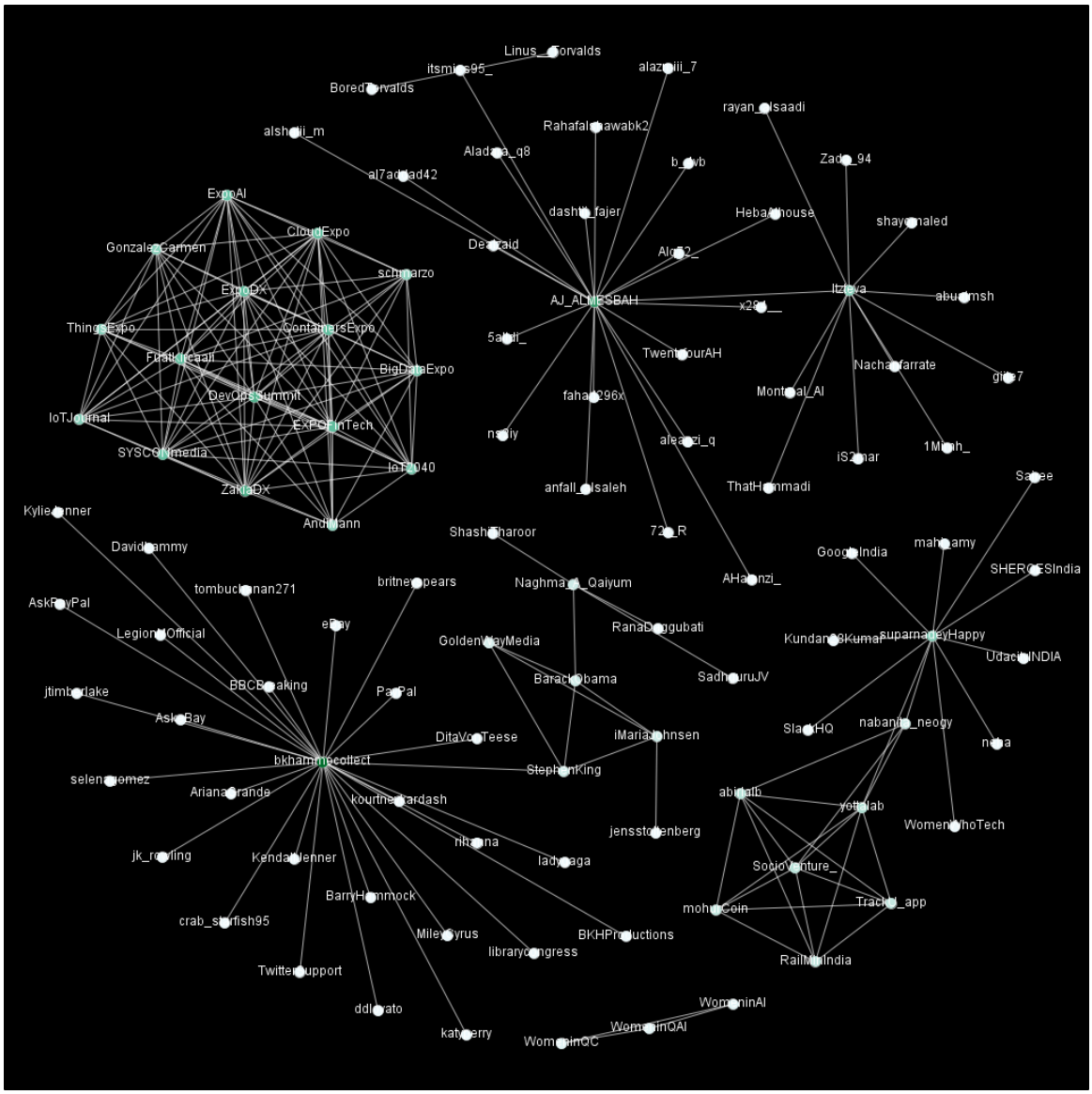

Figure 1. *Visualization of Twitter Network*

Network Measures:

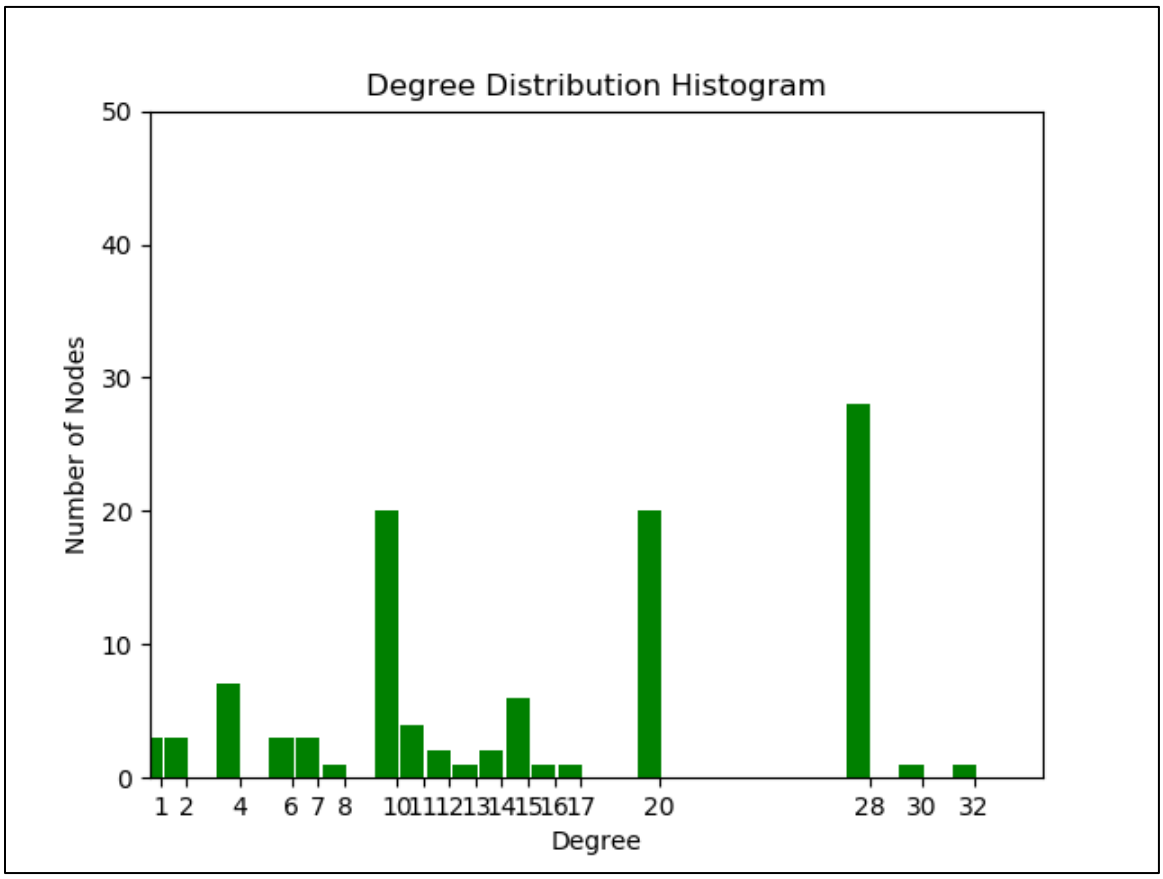

***Figure 2**. Degree distribution of the Twitter network (Histogram)*

Clustering Coefficient Metric Report

Parameters:

Network Interpretation: undirected

Results:

Average Clustering Coefficient: 0.739
Total triangles: 420
The Average Clustering Coefficient is the mean value of individual coefficients.

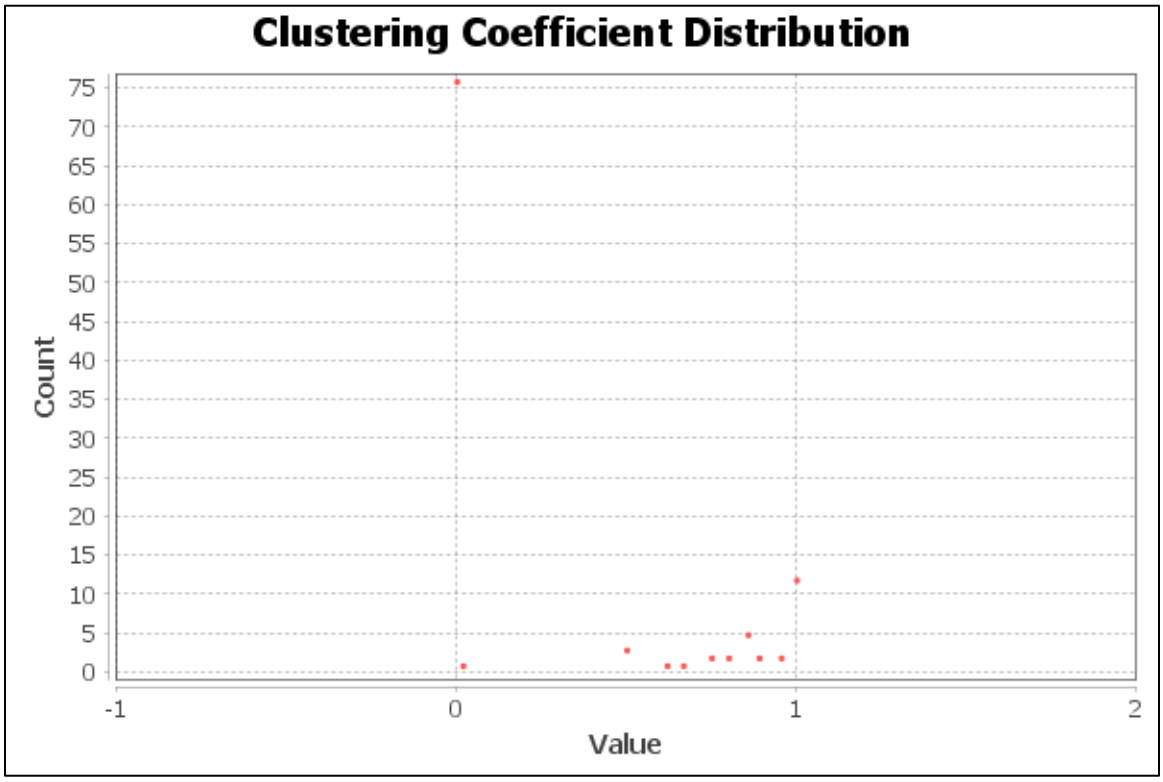

***Figure 3**. Clustering Coefficient of nodes*

Eigenvector Centrality Report

Parameters:

Network Interpretation: undirected
Number of iterations: 100
Sum change: 0.007473845256609284

Results:

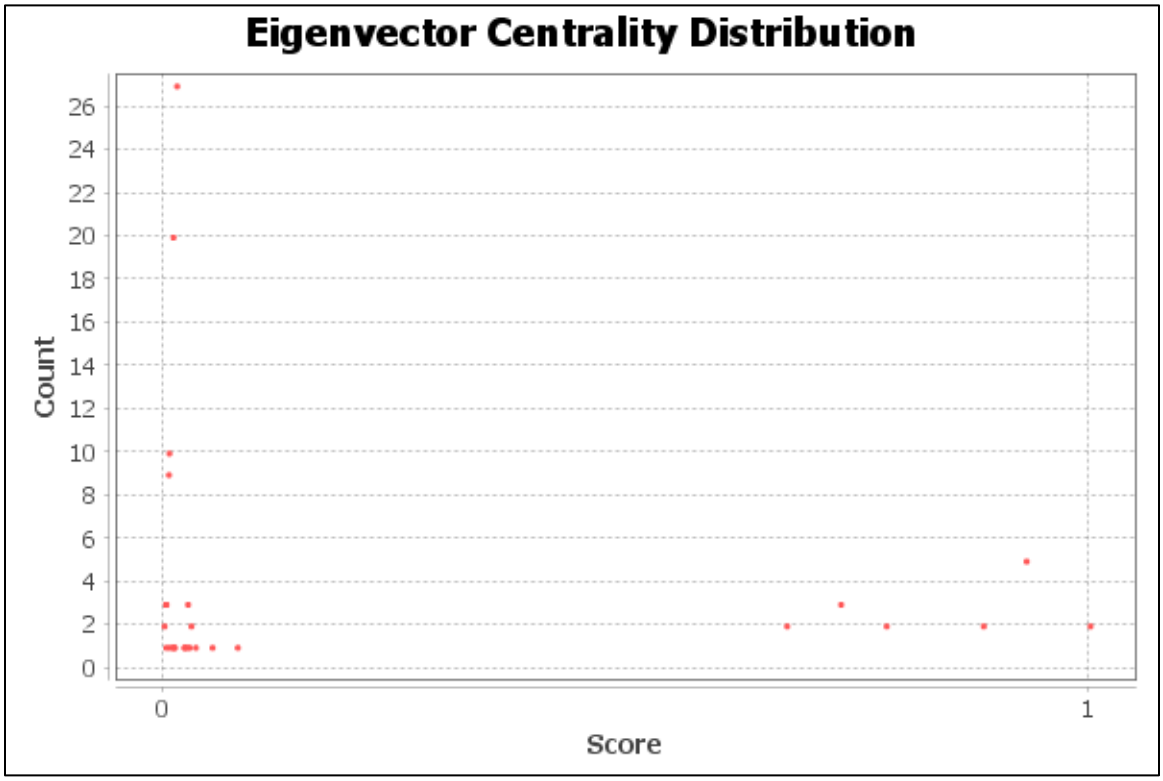

***Figure 4**. Eigenvector Centrality of the nodes*

Graph Distance Report

Parameters:

Network Interpretation: undirected

Results:

Diameter: 5
Radius: 1
Average Path length: 2.3466947960618847

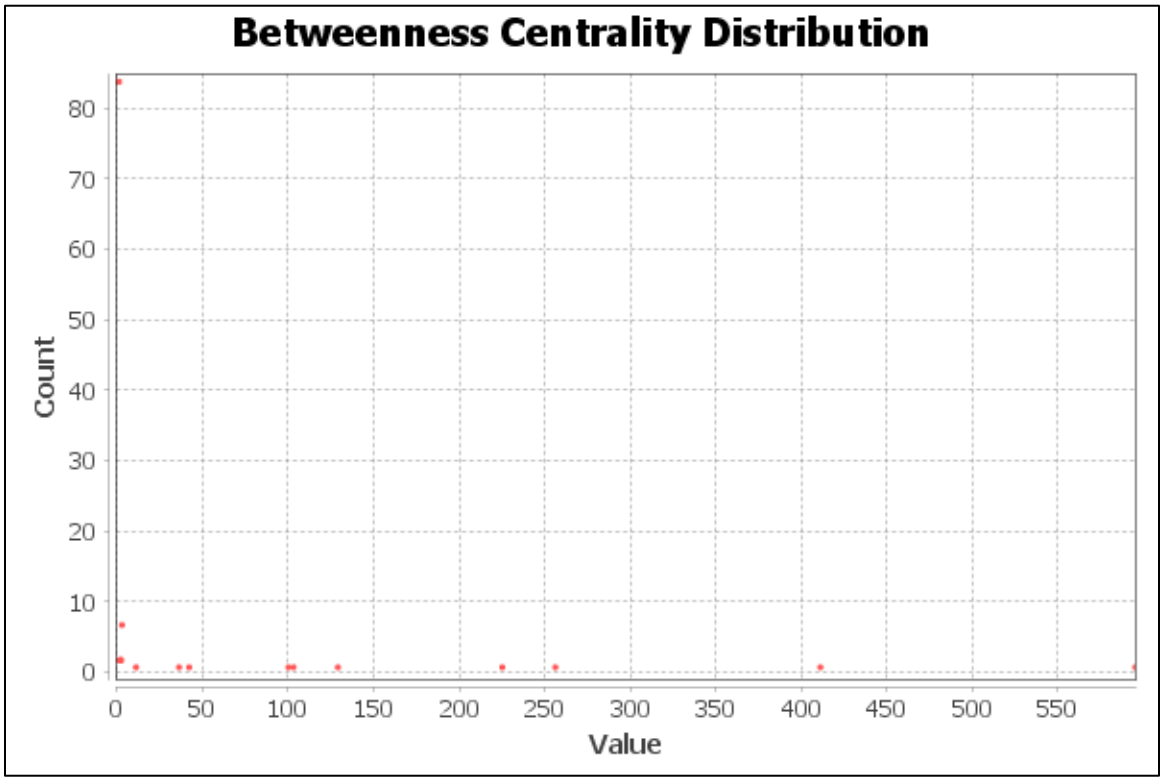

Figure 5. Betweenness Centrality of nodes

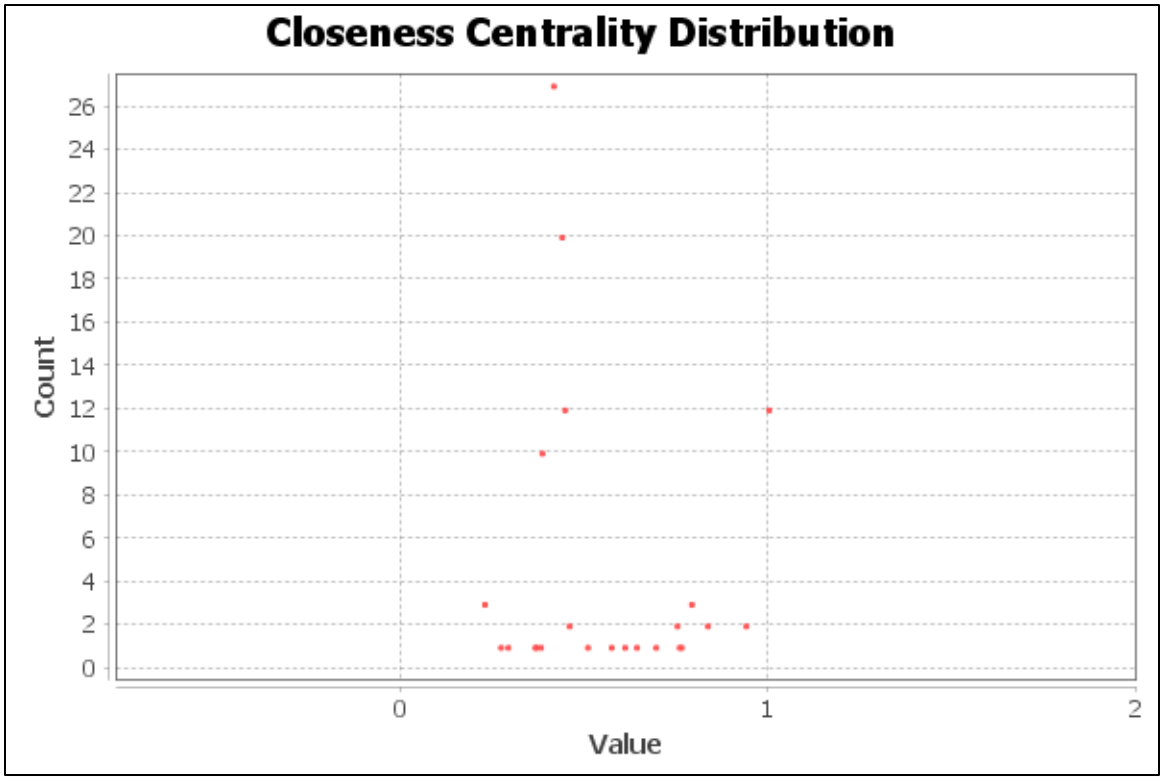

Figure 6. *Closeness Centrality of the nodes*

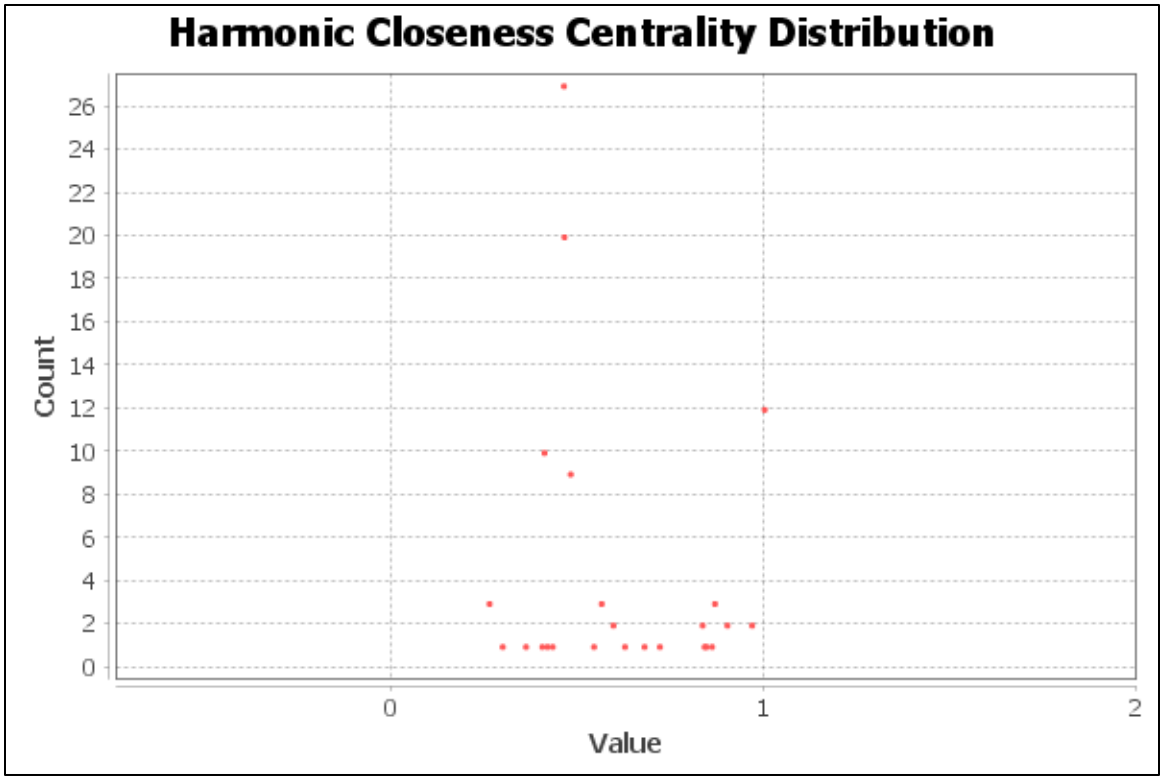

Figure 7. *Harmonic closeness Centrality of the nodes*

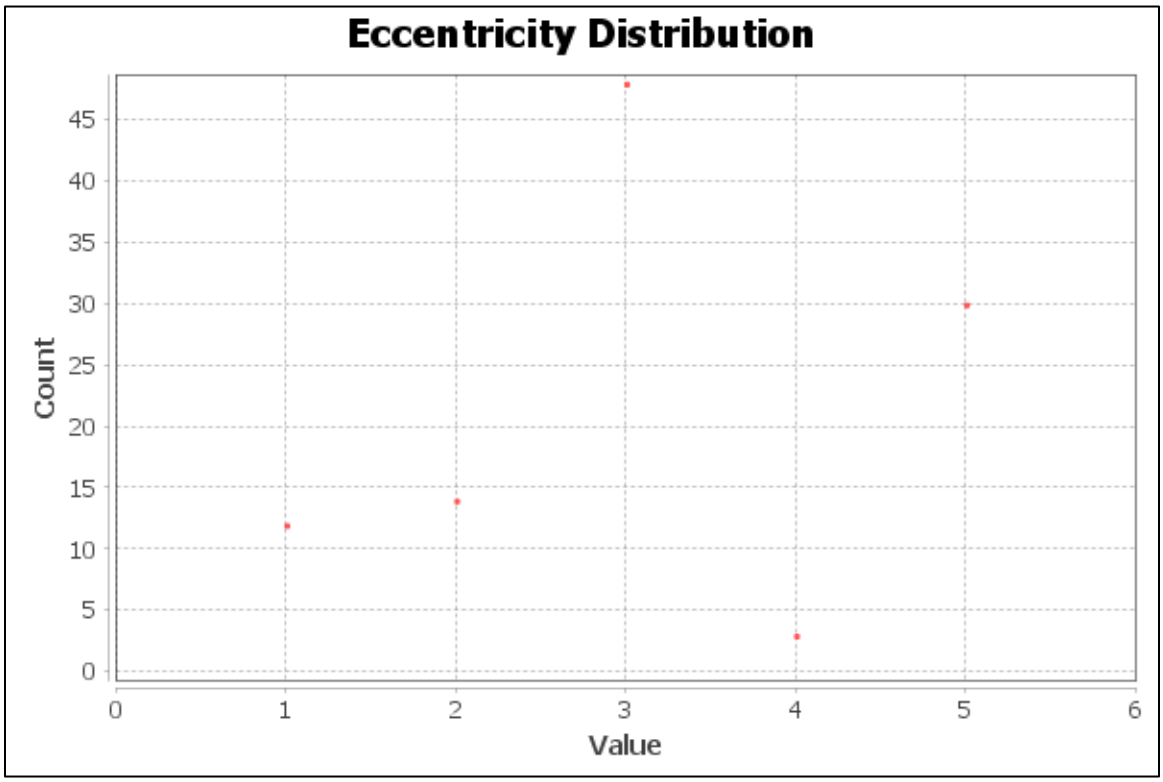

Figure 8. *Eccentricity of the nodes*

Degree Report

Results:

Average Degree: 3.925

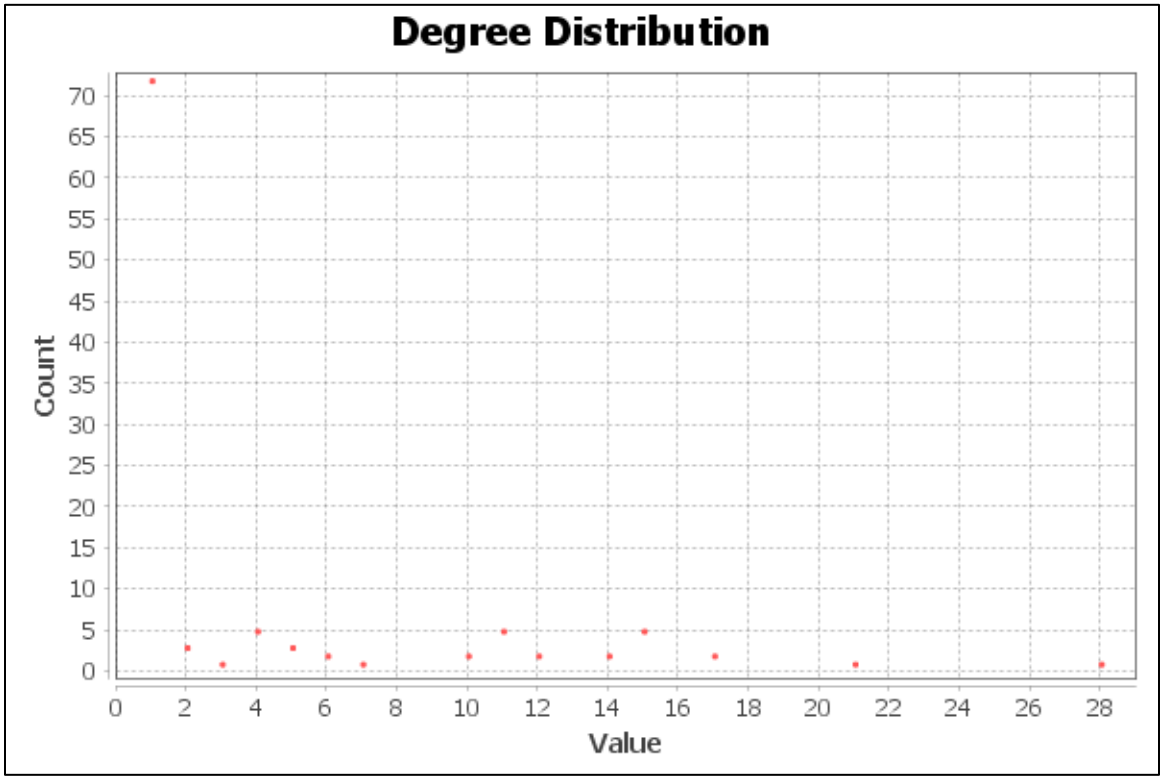

Figure 9. *Degree Distribution of the nodes*

Weighted Degree Report

Results:

Average Weighted Degree: 3.925

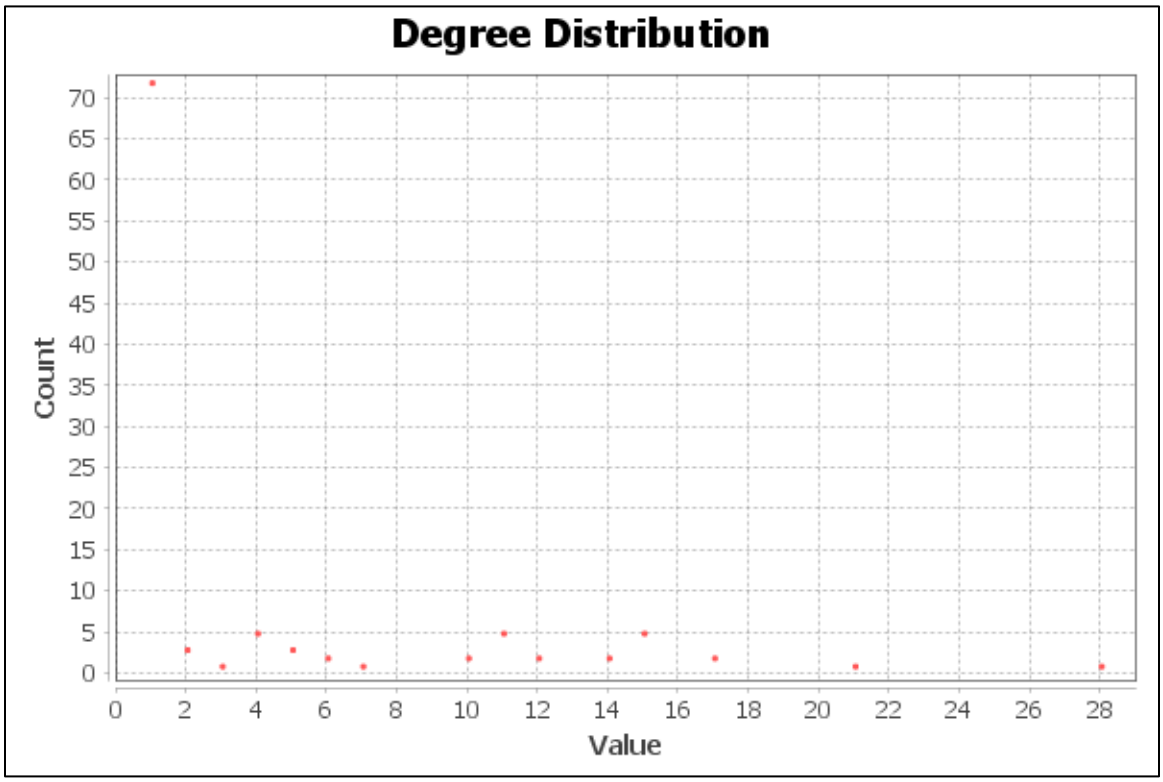

***Figure 10**. Weighted Degree Distribution of the nodes*

Graph Density Report

Parameters:

Network Interpretation: undirected

Results:

Density: 0.037

HITS Metric Report

Parameters:

E = 1.0E-4

Results:

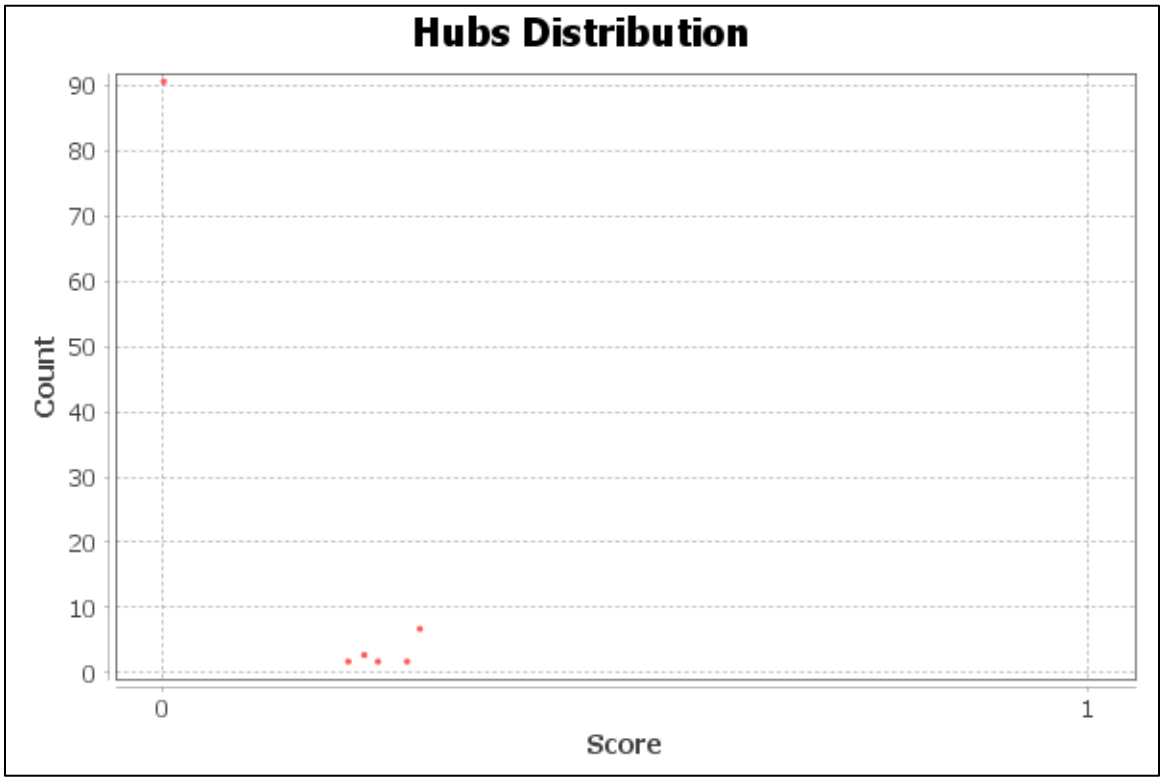

Figure 11. *HITS of the nodes (Hub)*

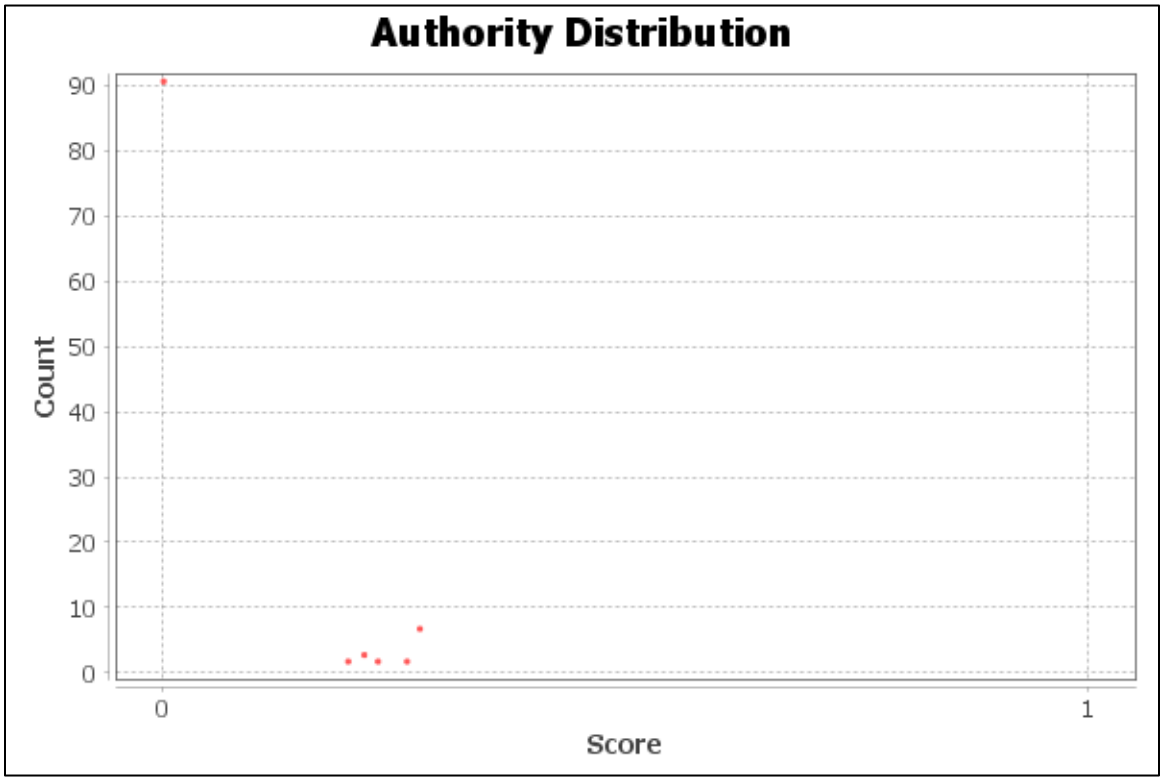

Figure 12. *HITS of the nodes (Authority)*

Modularity Report

Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

Modularity: 0.683
Modularity with resolution: 0.683
Number of Communities: 8

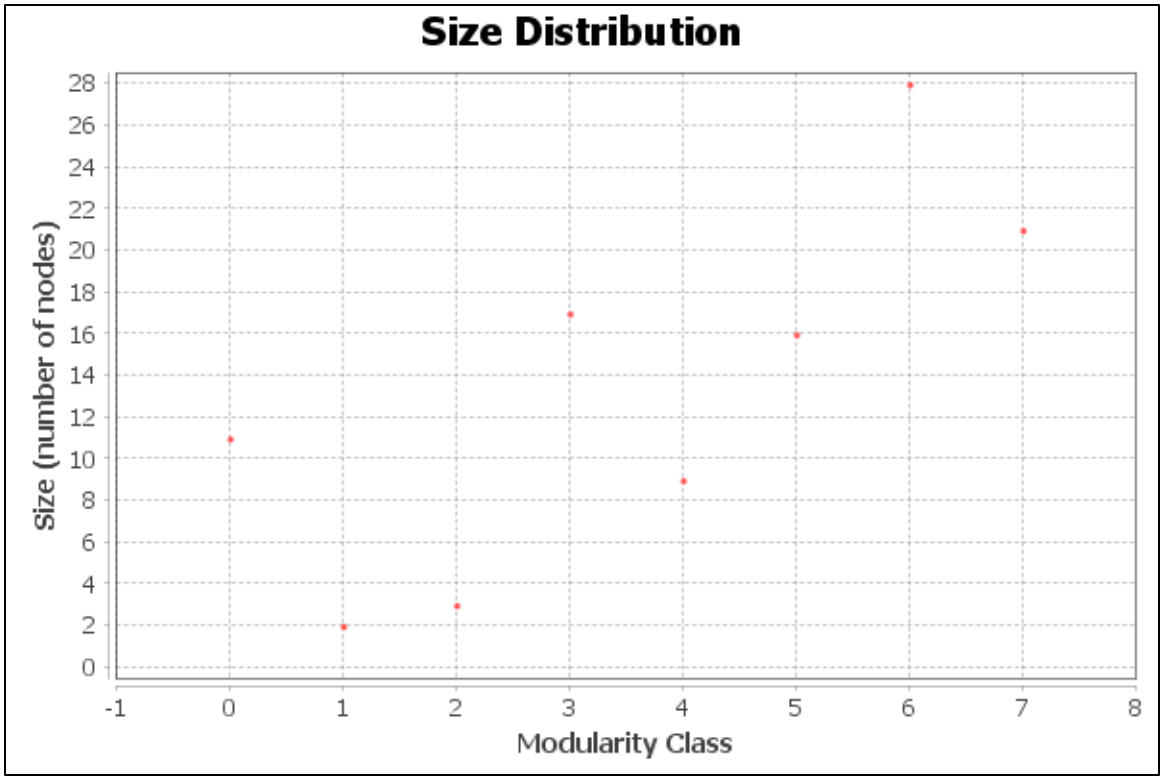

Figure 13. *Modularity of the nodes*

PageRank Report

Parameters:

Epsilon = 0.001
Probability = 0.85

Results:

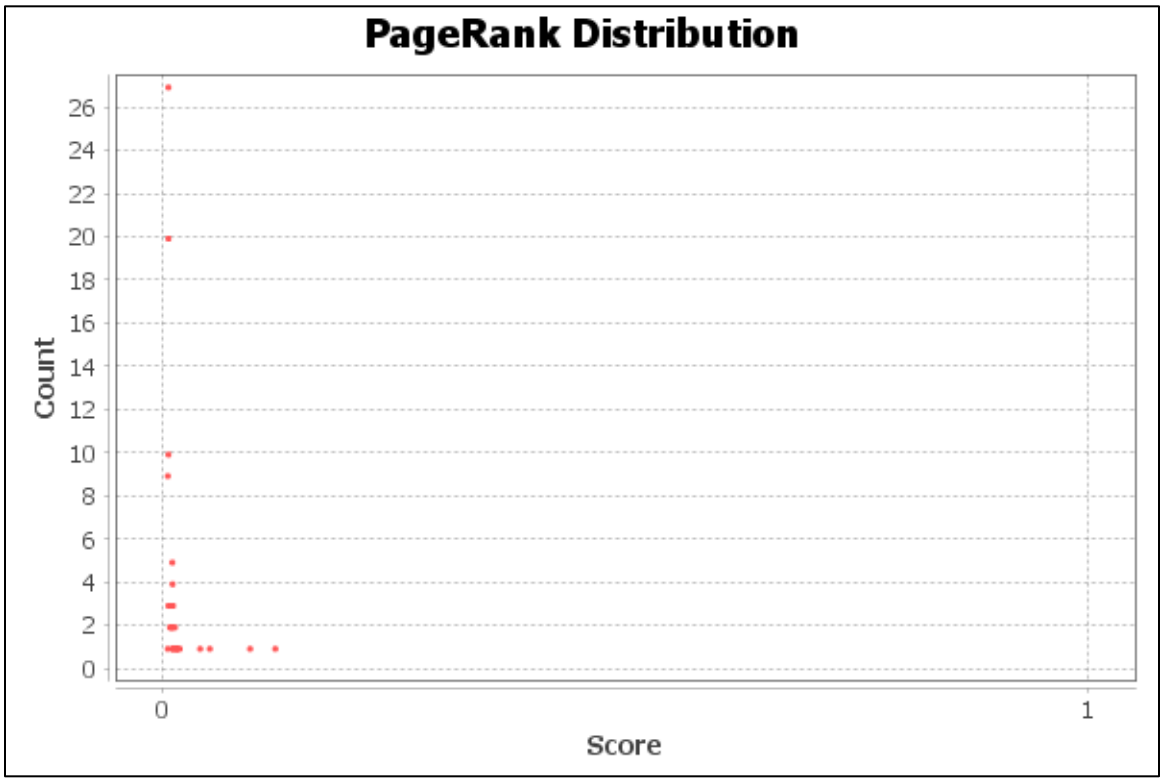

Figure 14. *PageRank of the nodes*

Connected Components Report

Parameters:

Network Interpretation: undirected

Results:

Number of Weakly Connected Components: 6

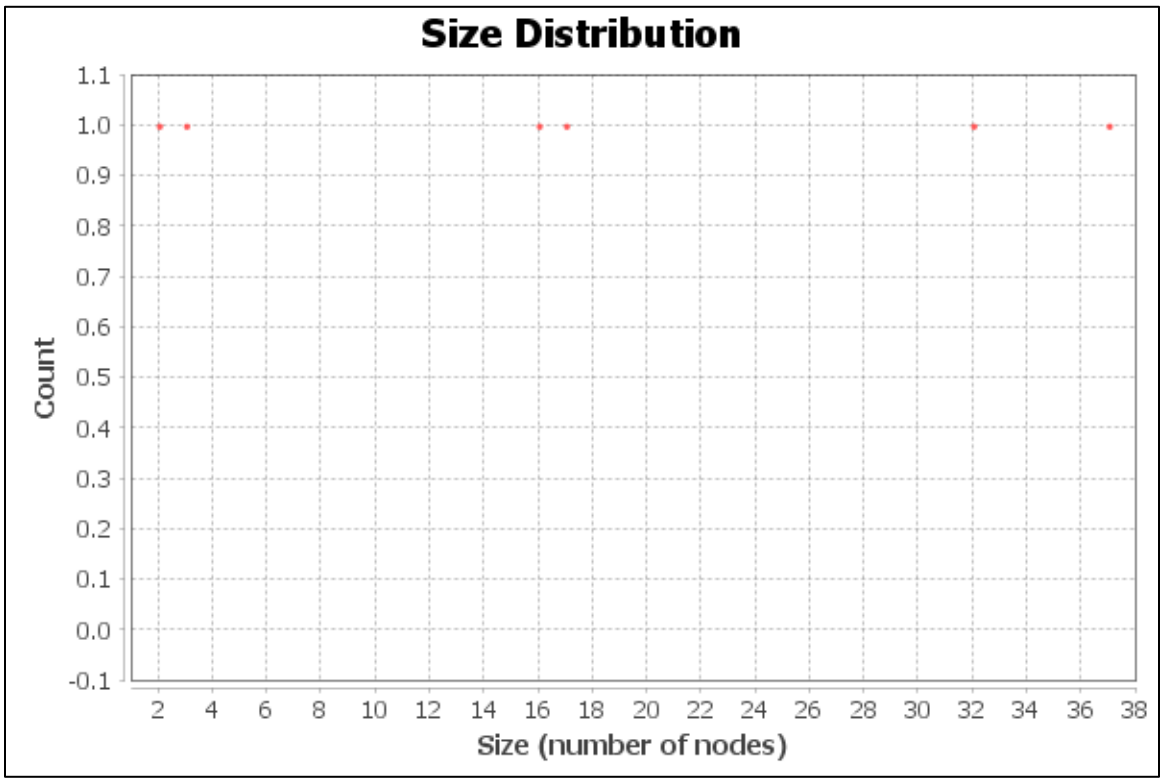

Figure 15. *Connected Components of the network*