+FEDERAL STATE AUTONOMOUS EDUCATIONAL INSTITUTION
OF HIGHER EDUCATION
ITMO UNIVERSITY

Report
on the practical task No. 7
“Algorithms on graphs. Tools for network analysis”

Performed by
Chernobrovkin Timofey (412642)
Academic group J4133c
Accepted by
Dr Petr Chunaev

St. Petersburg
2023

Goal

The use of the network analysis software Gephi.

Problem

1. Download and install Gephi from https://gephi.org/.

2. Choose a network dataset from https://snap.stanford.edu/data/ with number of nodes at most 10,000. You are free to choose the network nature and type (un/weighted, un/directed).

3. Change the format of the dataset for that accepted by Gephi (.csv, .xls, .edges, etc.), if necessary.

4. Upload and process the dataset in Gephi. Check if the parameters of import and data are correct.

5. Obtain a graph layout of at least two different types.

6. Calculate available network measures in Statistics provided by Gephi.

7. Analyze the results for the network chosen

Theory

Let us introduce the characteristics of the degrees of vertices of an unweighted graph:

- d(v), the degree of v, - the number of incoming and outgoing edges of vertex v;
- $d_{in}(v)$, half-degree of inlet v, - number of incoming edges of vertex v;
- $d_{out}(v)$, the semidegree of outgoing v, - the number of outgoing edges of vertex v;
- $\bar{d} = \frac{1}{|V|}\sum_{v \in V} d(v)$ is the average degree of the vertices.

For a weighted graph, similar quantities can be introduced, with the difference that they take into account the sum of weights on (incoming/outgoing) edges rather than the number of (incoming/outgoing) edges.

Now let us introduce graph characterizations in the sense of distances:

- dist(v, u) - distance (length of the shortest path) between v and u (G is connected);
- eccentricity $\in(v) = \max_{u \in V}$ dist(v, u) - the largest distance between v and other vertices;
- radius $r = \min_{v \in V} \in(v)$ - the smallest eccentricity across all vertices;
- diameter $D = \max_{v \in V} \in(v)$ - the largest eccentricity over all vertices, i.e., the largest distance between a pair of vertices;
- average path length $l = \frac{1}{|V| \cdot (|V|-1)} \sum_{v \neq u} dist(v, u)$

Further, recall that the density ρ of a graph is the quotient of |E| and the number of possible edges with the same |V|, i.e., the number of edges in a complete graph with |V|. of possible edges with the same |V|, i.e., the number of edges in a complete graph with |V| vertices:

$$\rho = \frac{2|E|}{|V|(|V|-1)}.$$

When ρ ≈ 0, the graph is usually called sparse. Note that by convention the density of any complete graph is equal to one.

The above and many other quantities allow us to characterize a graph from different points of view. This is extremely useful when analyzing graphs modeling complex networks of the real world.

Gephi also offers the following values for calculation:

- HITS calculates two independent values for each node. The first value (called Authority) corresponds to the importance of the node itself. The second value (called Hub) corresponds to the importance of the edges of the node.
- PageRank ranks nodes (as pages) according to how often a "user" clicking on links (edges) will reach a given node.
- The connected components defines the number of connected components of the graph.
- Modularity - Community Identification Algorithm.
- The clustering coefficient, along with the average shortest path, can serve to test the Small Wond hypothesis. It represents how densely nodes are clustered in their immediate neighborhood. The average corresponds to the clustering degree of the entire
- Eigenvector Centrality - Calculates the weight of a node in the network based on the links of this node.

Materials and methods

In this task, all calculations were performed on the student's personal laptop. The work was performed in the network analysis software Gephi.

Results

For this assignment, a graph representing a network of people trading using Bitcoin on the Bitcoin Alpha platform was taken. Since Bitcoin users are anonymous, there is a need to keep track of the reputation of users to prevent transactions with scammers and risky users. Bitcoin Alpha participants rate other participants on a scale from -10 (complete distrust) to +10 (complete trust)

in increments of 1. This is the first explicit weighted signed signed directed network available for research.

The available graph was represented in two different types (Figure 1, 2).

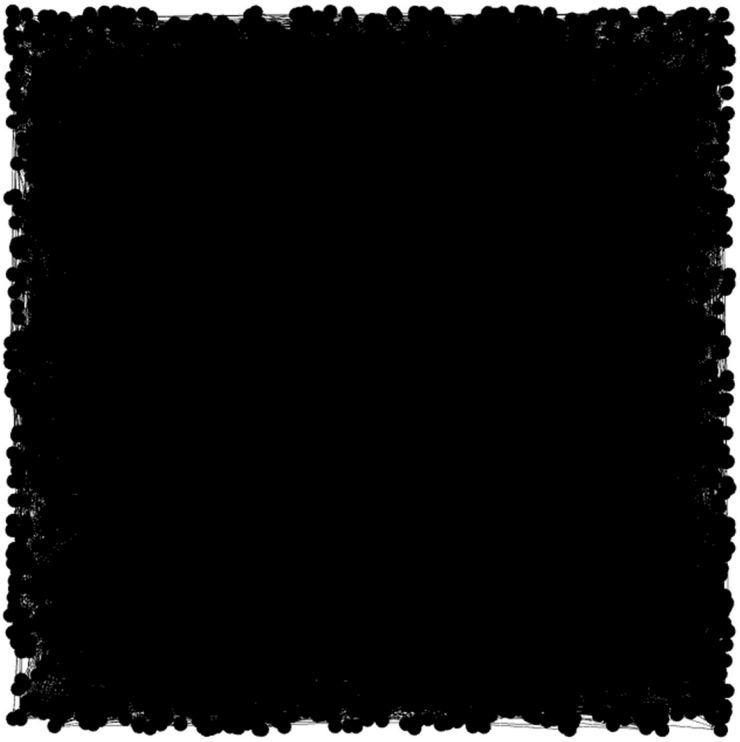

Figure 1 - First graph representation

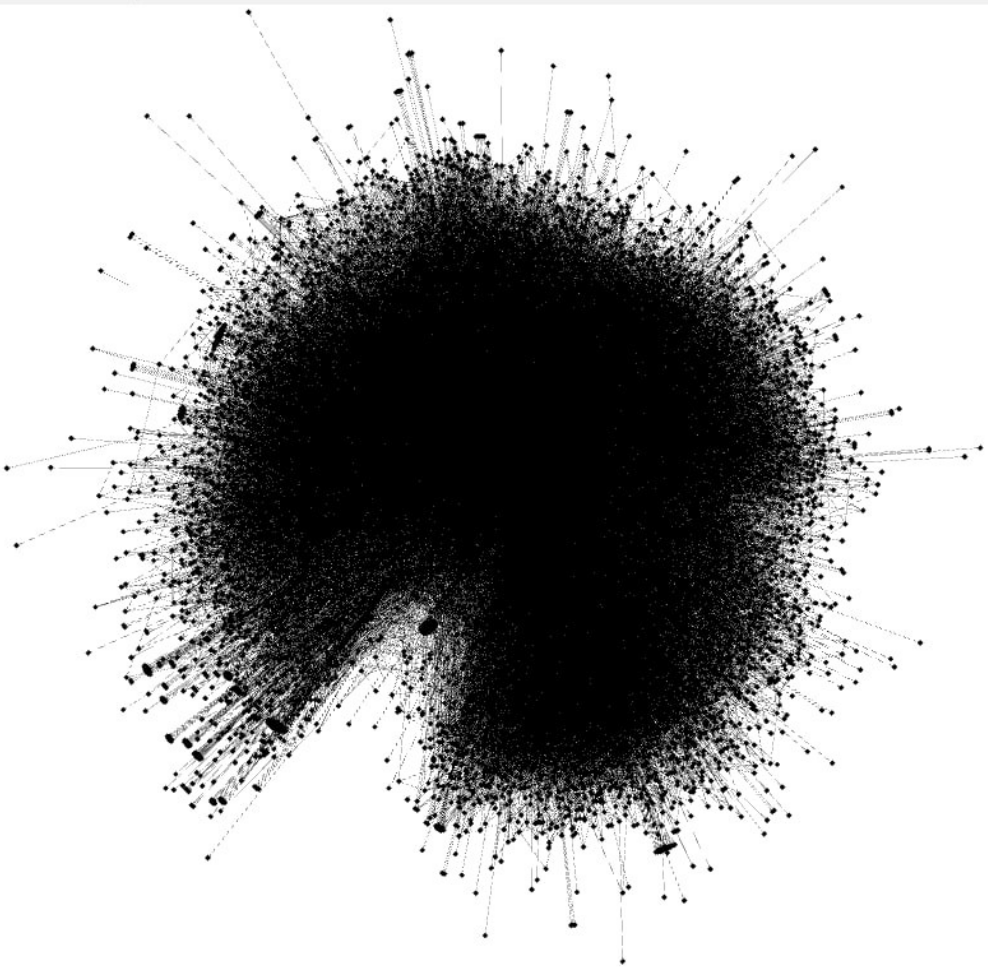

Figure 2 – Processed version of the graph

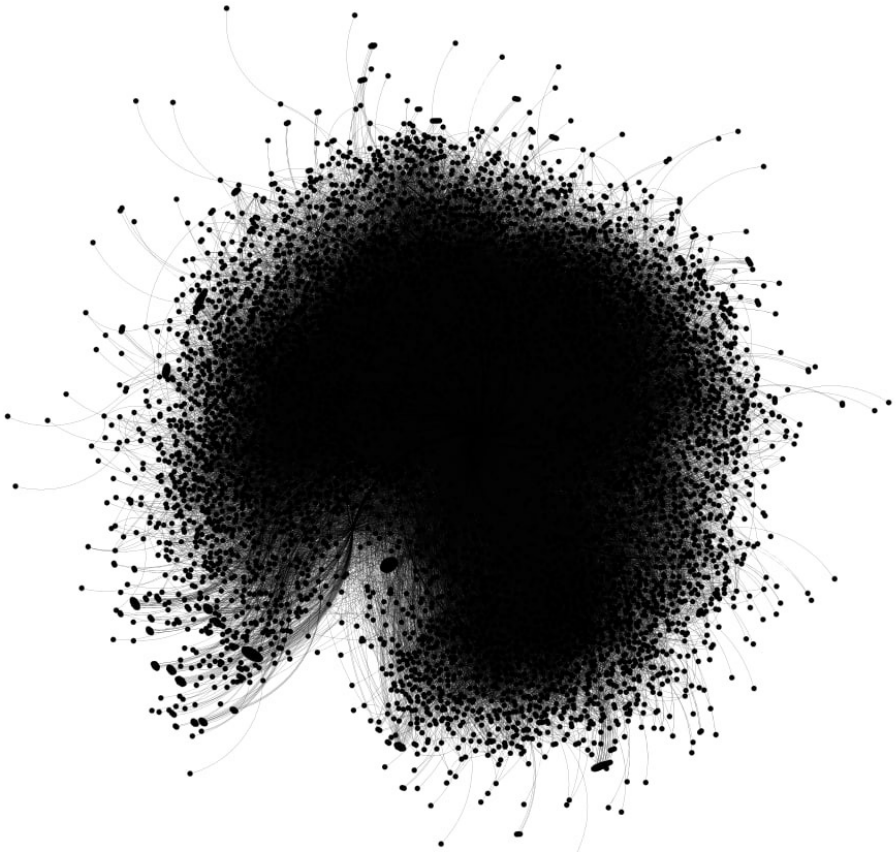

Figure 2 - Second graph representation

Next, we calculated the available network metrics in the statistics provided by Gephi.

The average value in the graph is 8.988 (Figure 3). We also plotted the graph degree distribution, in-degree distribution and in-degree distribution shown in Figures 4-6 respectively.

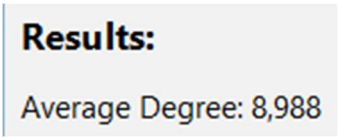

Figure 3 - Average degree in the column

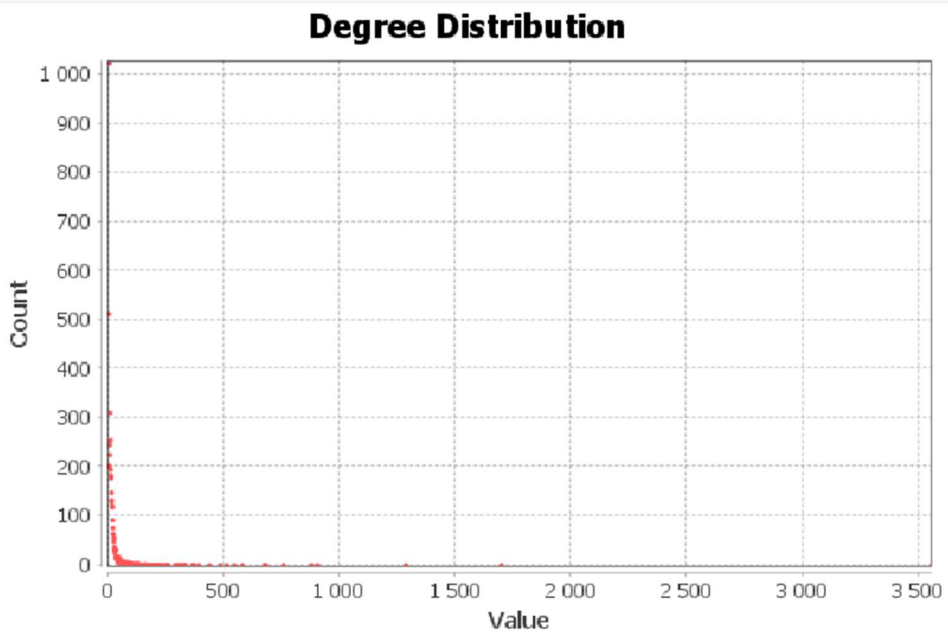

Figure 4 – Degree distribution

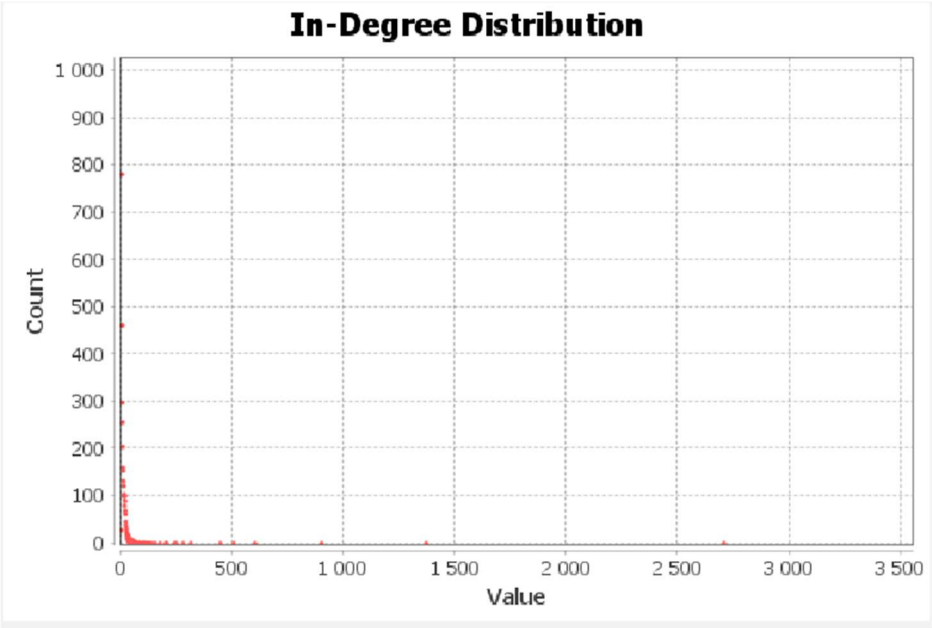

Figure 5 – In-degree distribution

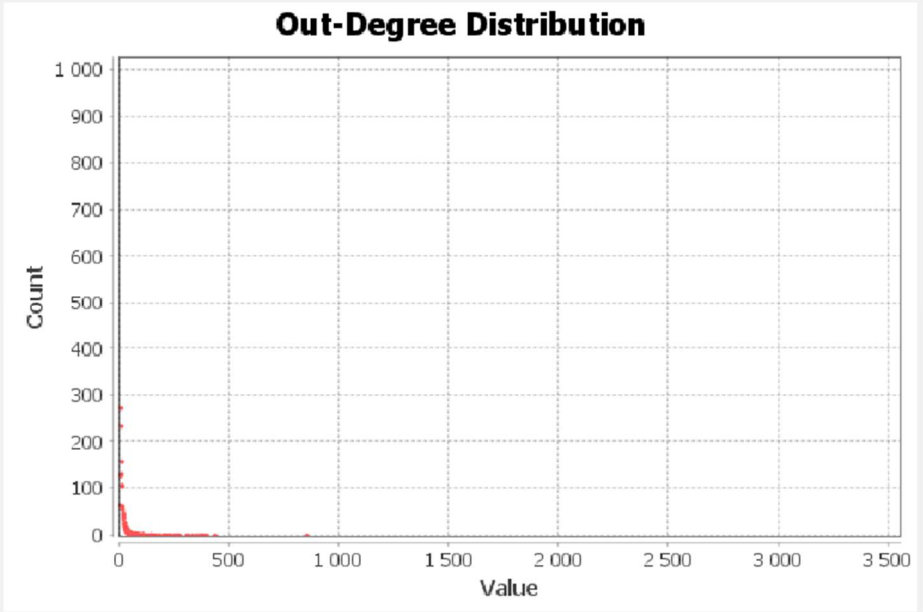

Figure 6 – Degree distribution

The average weighted degree of the graph was found to be 13.338 (Figure 7).

Results:

Average Weighted Degree: 13,338

Рисунок 7 – Average weighted degree

The diameter, radius and average path length were found. The results are presented in Figure 8.

Parameters:

Network Interpretation: directed

Results:

Diameter: 8
Radius: 0
Average Path length: 3.033007608079684

Figure 8 - Diameter, radius and average path length

The plots of betweenness centrality distribution, closeness centrality distribution, harmonic closeness centrality distribution and eccentricity distribution were also plotted. The graphs are presented in Figures 9-12, respectively.

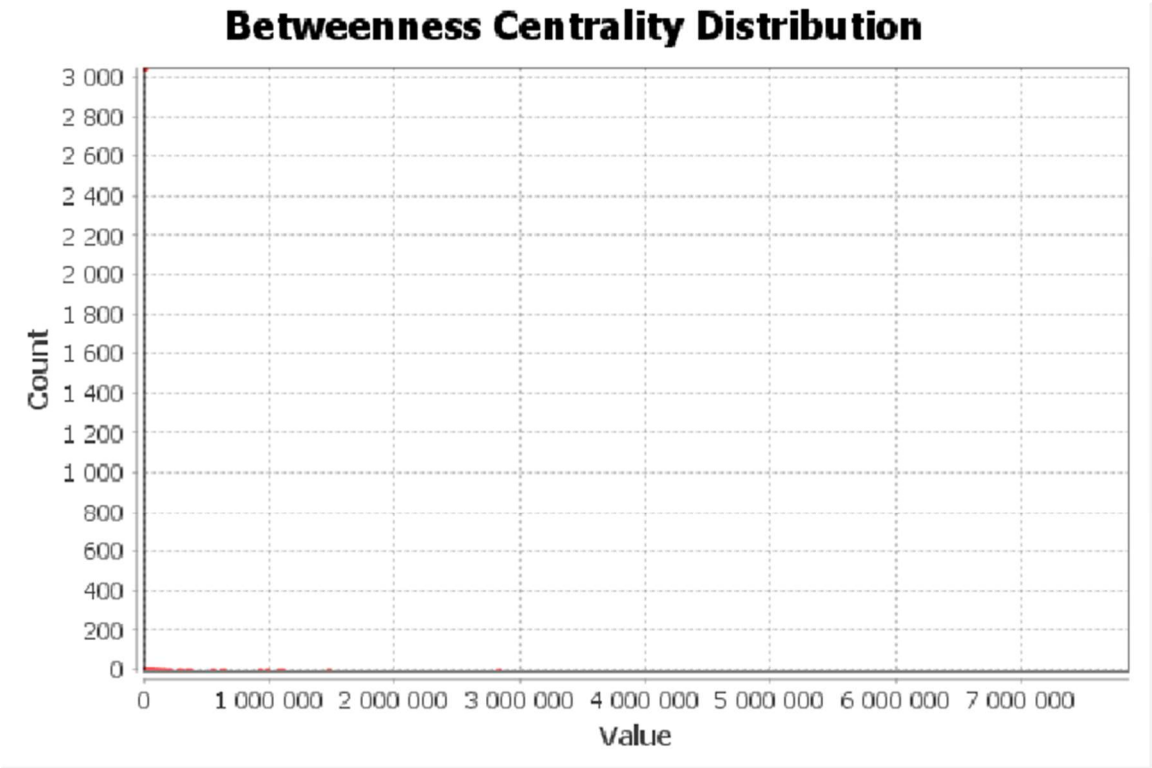

Figure 9 – Betweenness centrality distribution

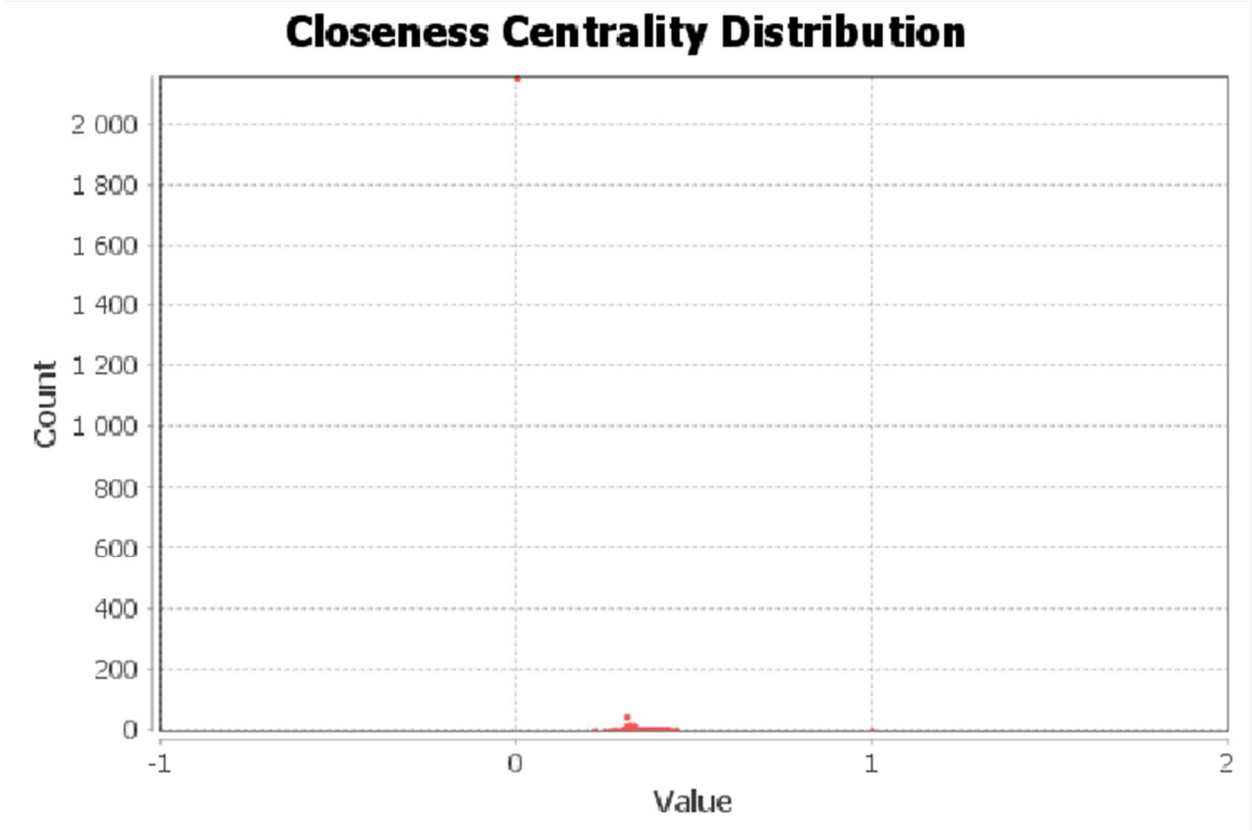

Figure 10 – Closeness centrality distribution

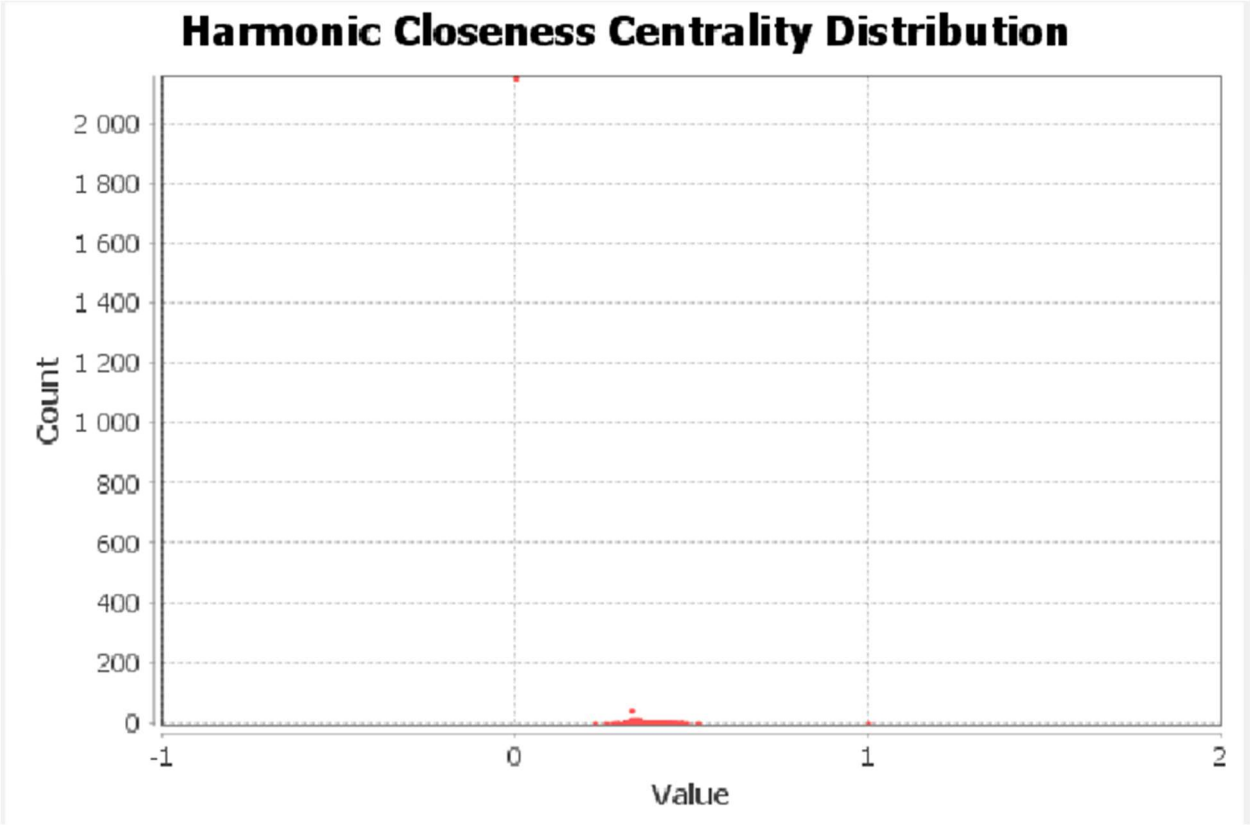

Figure 11 – Harmonic closeness centrality distribution

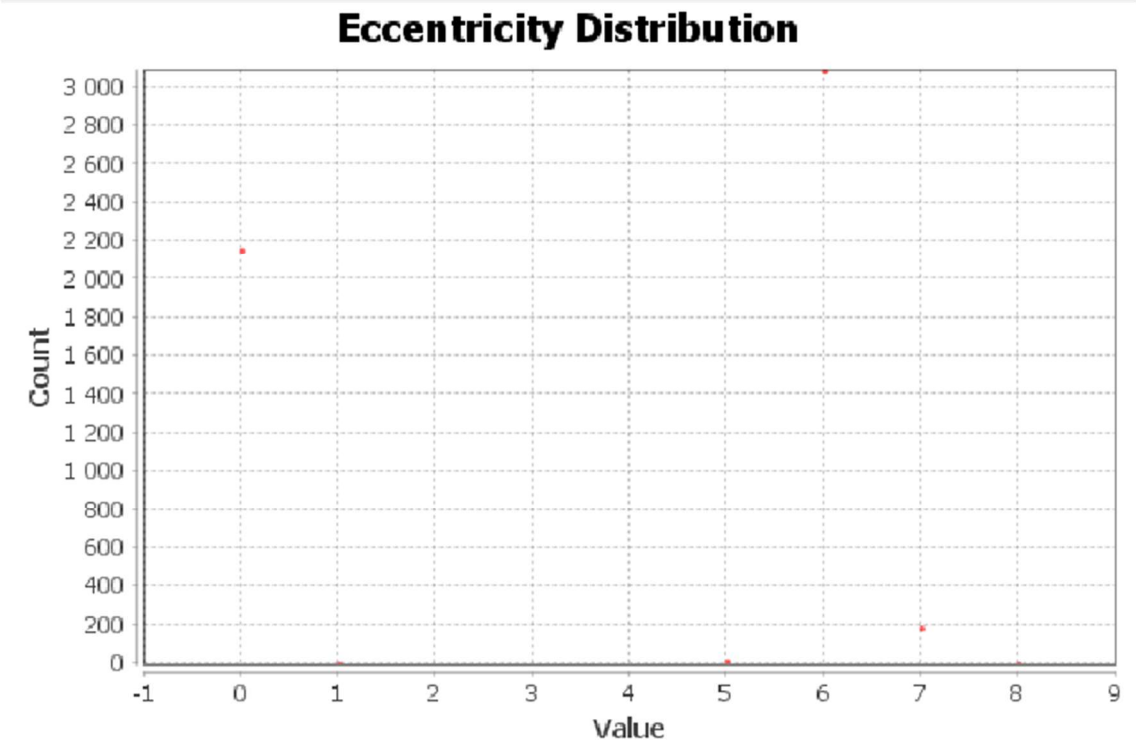

Figure 12 – Eccentricity distribution

The graph density was calculated and found to be equal to 0.002 (Figure 13).

Parameters:

Network Interpretation: directed

Results:

Density: 0,002

Figure 13 - Graph density

The parameters by HITS were calculated (Figure 14-16).

HITS Metric Report

Parameters:

E = 1.0E-4

Figure 14 – Epsilon parametr for HITS metric

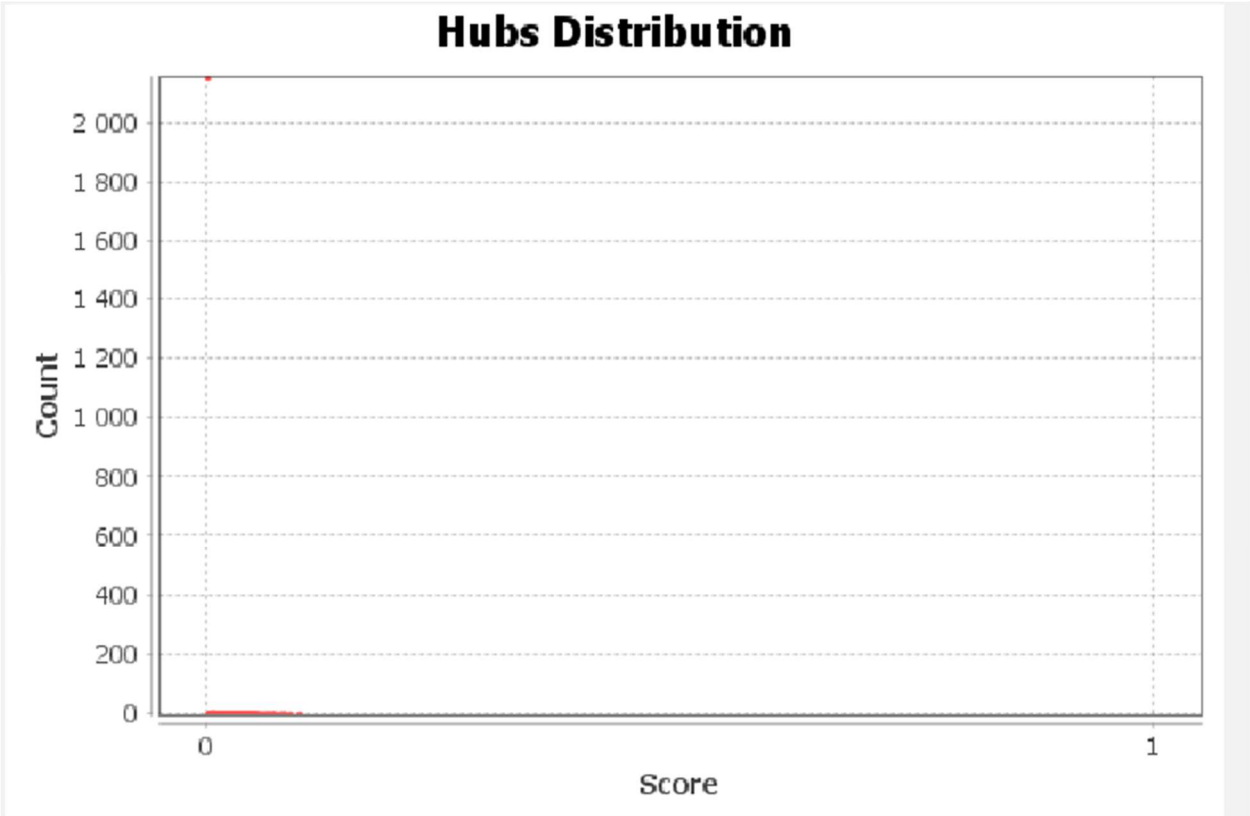

Figure 15 – Hubs distribution

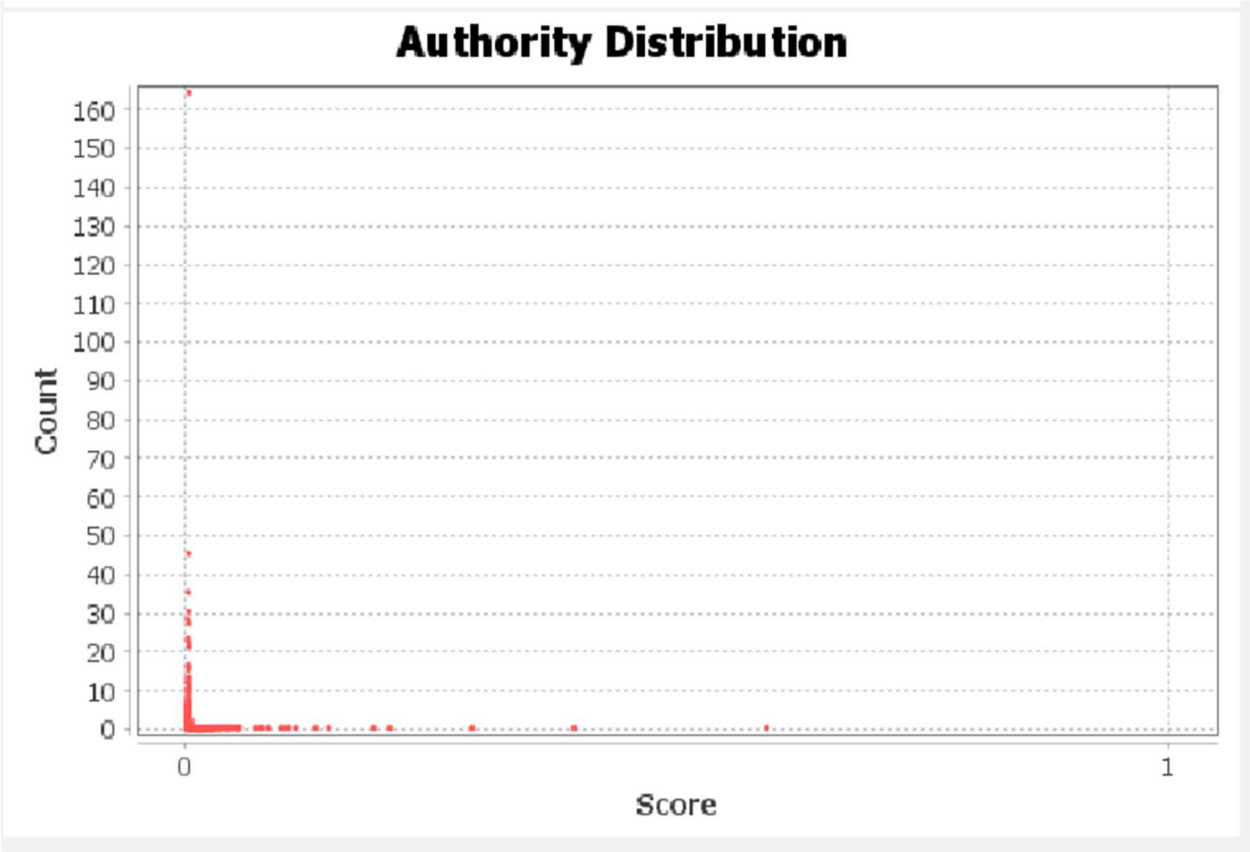

Figure 16 – Authority distribution

The PageRank distribution shown in Figure 18 with the parameters shown in Figure 17 was calculated.

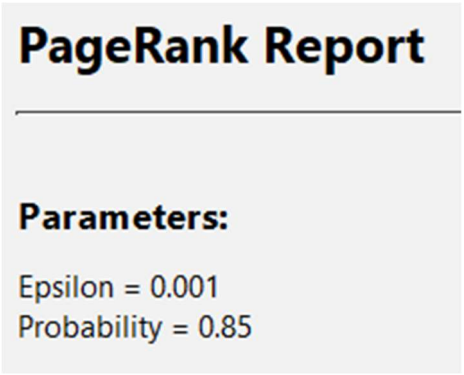

Figure 17 - PageRank Parameters

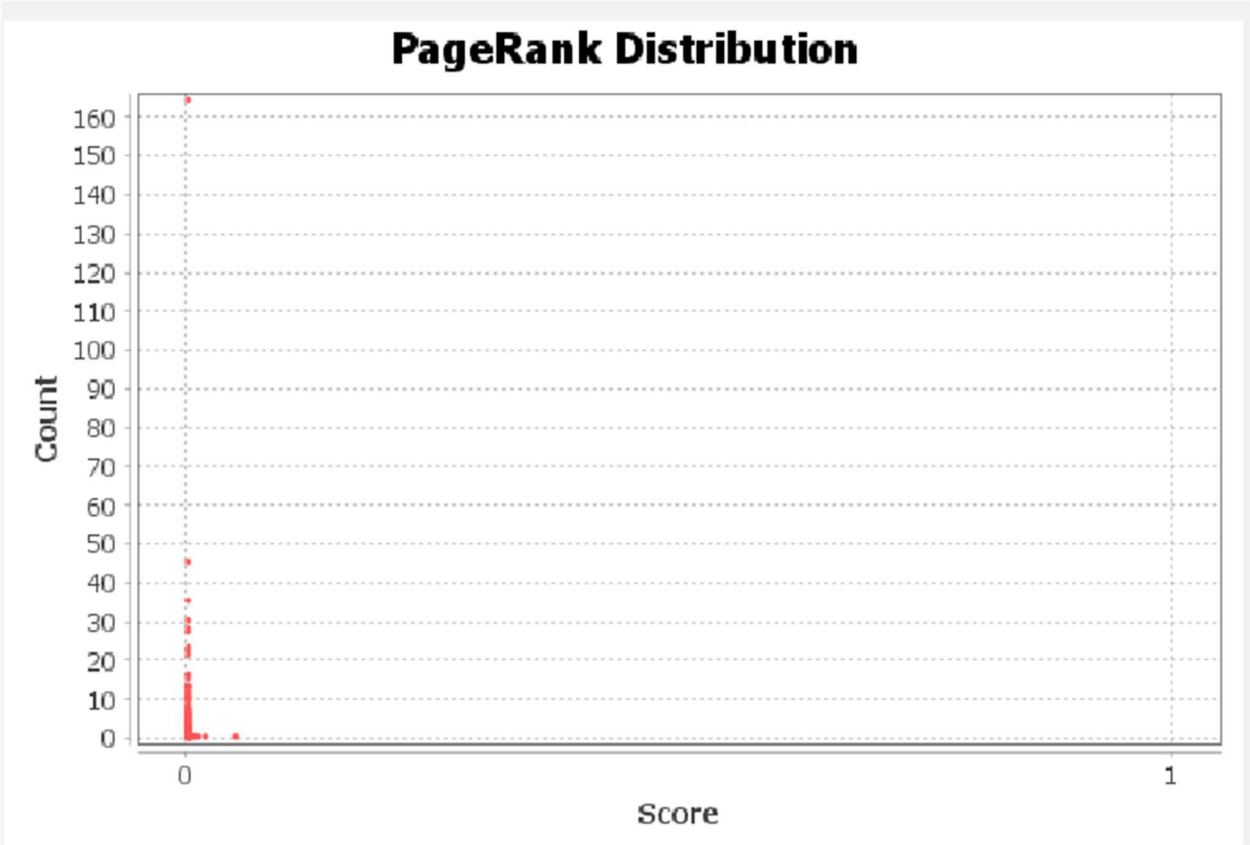

Figure 18 – PageRank Distribution

A connected components report was made to find the number of weakly connected components and the number of strongly connected components (figure 19).

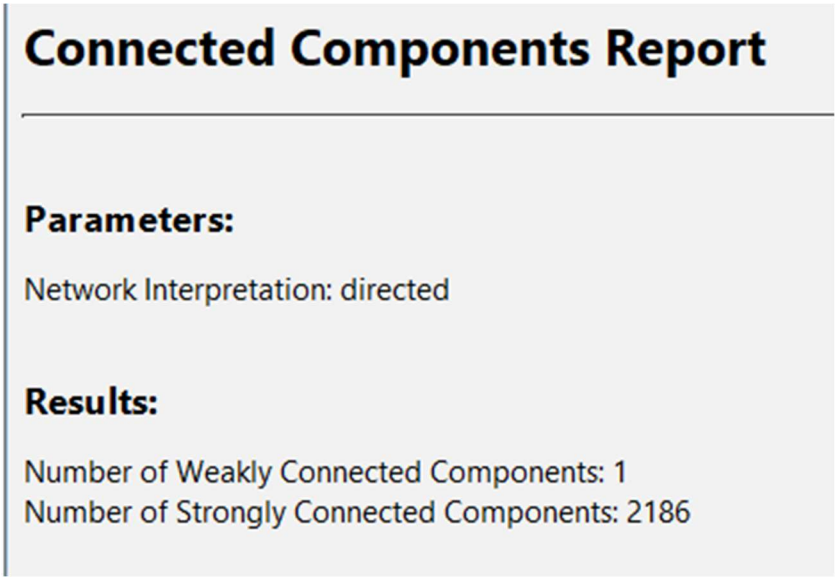

Figure 19 – connected components report

A modularity report was generated to calculate modularity, modularity with resolution, number of communities and size distribution. The results are presented in Figures 20 and 21 respectively.

Modularity Report

Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

Modularity: 0,262
Modularity with resolution: 0,262
Number of Communities: 6

Figure 20 – Modularity report

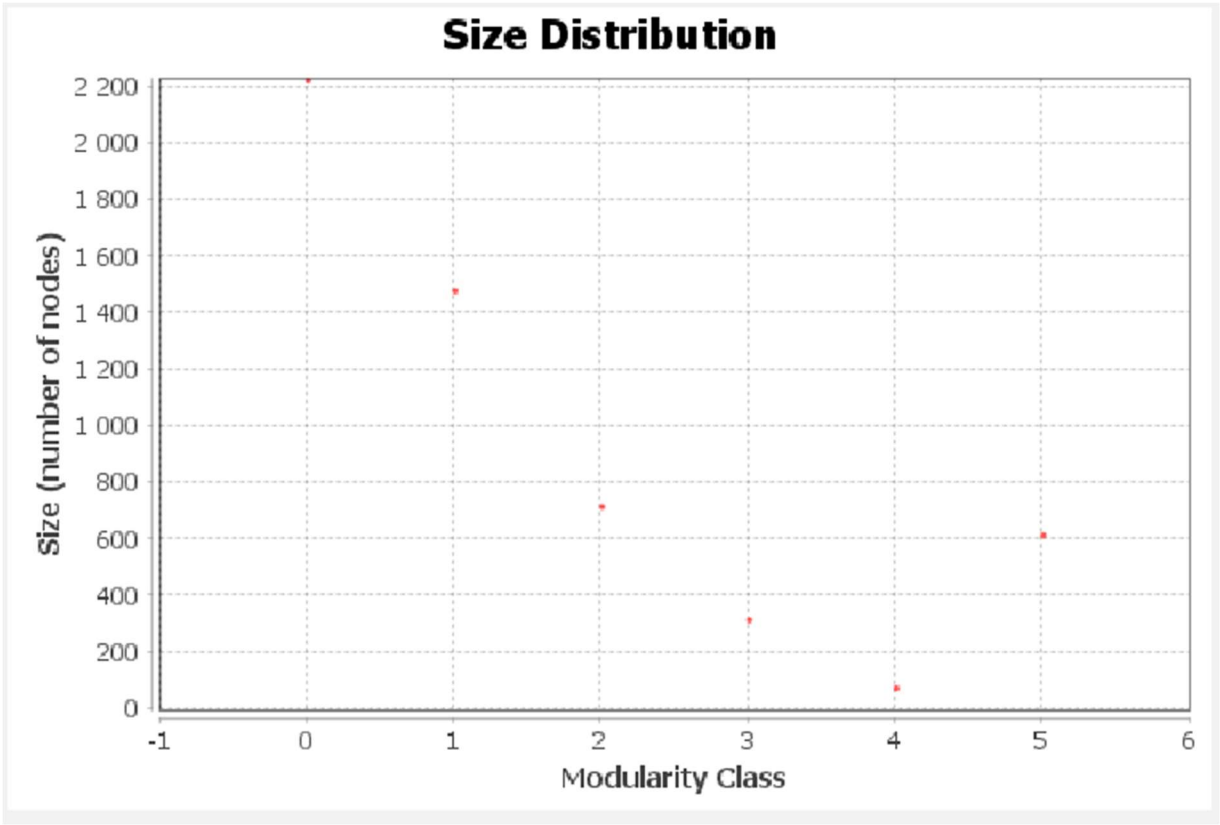

Figure 21 – Size distribution

A statistical inference report was generated to find the description length, number of communities, and size distribution (Figures 21, 22).

Figure 22 – Statistical inference report

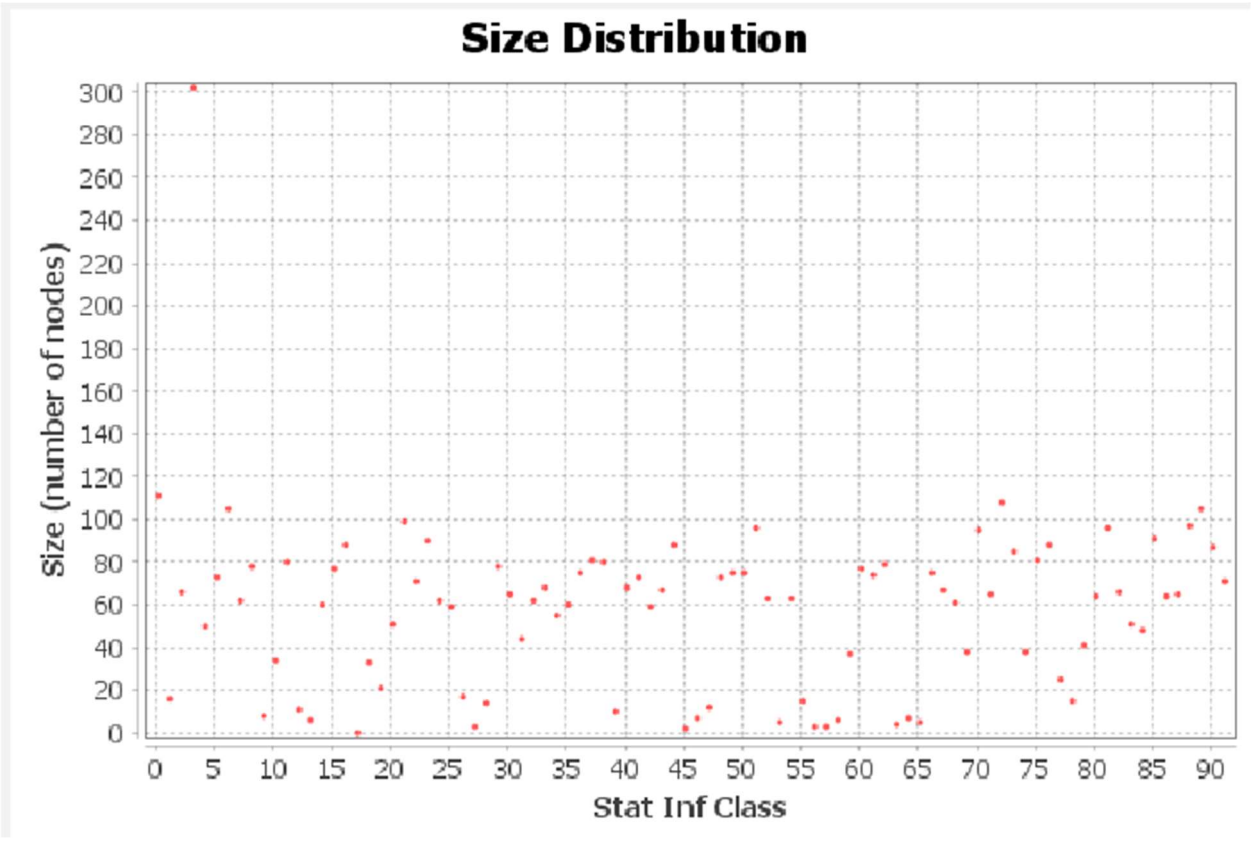

Figure 23 – Size distribution

A clustering coefficient metric report was made to find the average clustering coefficient (figure 24).

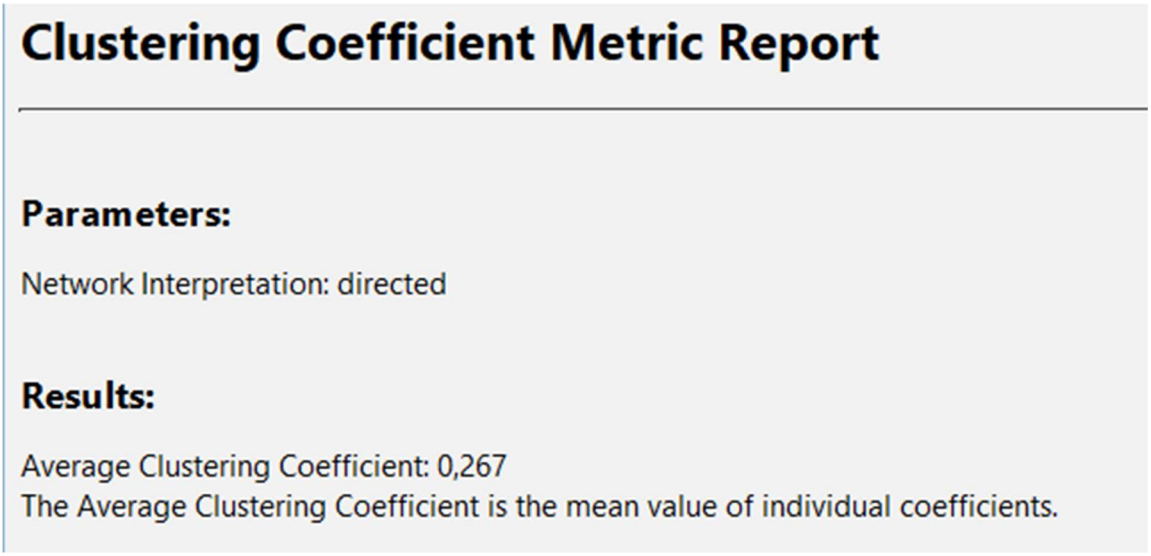

Figure 24 – Clustering coefficient metric report

An eigenvector centrality report and eigenvector centrality distribution was made (figure 25, 26).

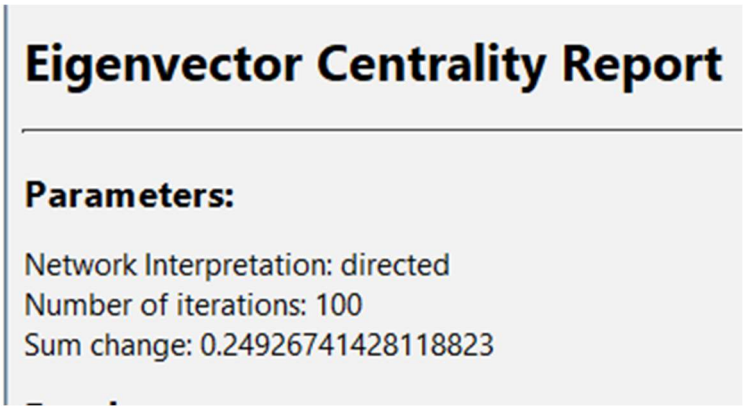

Figure 25 – Eigenvector centrality report

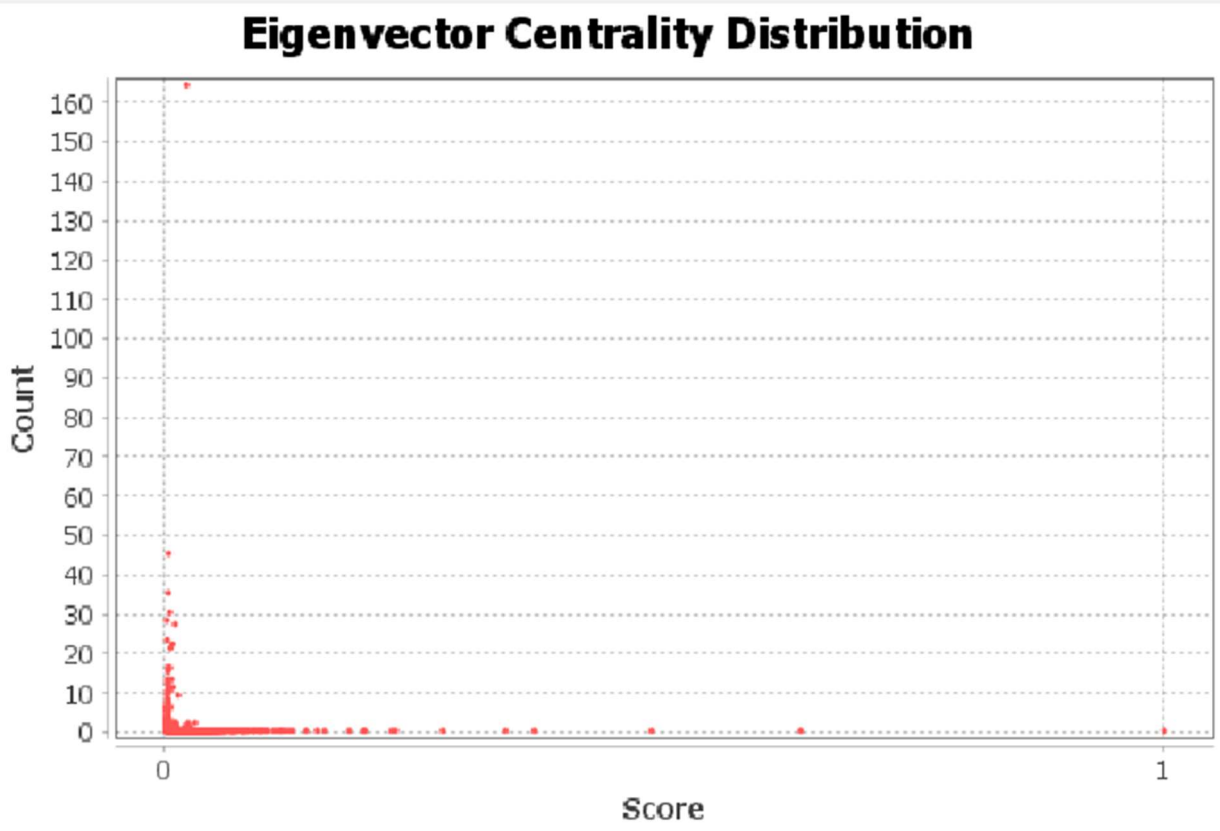

Figure 26 – Eigenvector centrality distribution

Since the array under study is not dynamic, it is not possible to analyze for dynamic arrays (Figure 27).

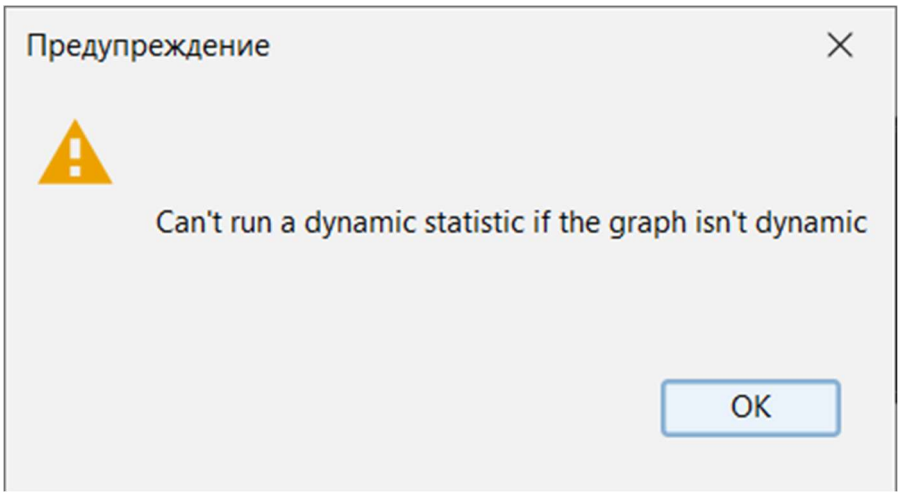

Figure 27 – Warning

Conclusion

This assignment explored the use of Gephi software to analyze networks. A graph representing a network of people trading with Bitcoin on the Bitcoin Alpha platform was analyzed. The graph was visualized in two views and its following parameters were found: average weighted

degree, diameter, radius, density, HITS calculation, PageRank calculation, number of connected components, modularity calculation, statistical inference, average clustering coefficient, eigenvector centrality, average path length, as well as the corresponding graphs.