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

Pakulev Aleksandr

J4133c

Accepted by

Dr Petr Chunaev

St. Petersburg

2021

Goal

The use of the network analysis software Gephi

Problems and methods

1. Download and install Gephi from https://gephi.org/.
2. Choose a network dataset from hittps://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 two different types.
6. Calculate available network measures in Statistics provided by Gephi.
7. Analyze the results for the network chosen.

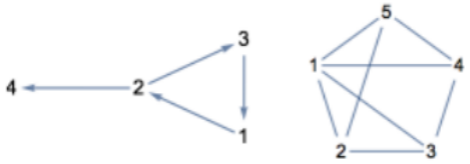
While performing the work, screenshot the main steps you are doing and insert in the report.

Brief theoretical part

Basic measures:

$|V|$, the number of vertices

$|E|$, the number of edges



Degree measures:

$d(v)$, **degree of** v, i.e. the number of edges for vertex v

$d_{\text{in}}(v)$, **in-degree of** v, i.e. the number of in-edges for vertex v

$d_{\text{out}}(v)$, **out-degree of** v, i.e. the number of out-edges for vertex v

$\overline{d} = \frac{1}{|V|} \sum_{v \in V} d(v)$, **average degree** over all vertices

Given a connected G, $\text{dist}(v, u)$ is the distance (shortest path length) between v and u

The **eccentricity** $\epsilon(v)$ of v is the greatest distance between v and any other vertex: $\epsilon(v) = \max_{u \in V} \text{dist}(v, u)$ ("how far a node is from the node most distant from it").

The **radius** r is the minimum eccentricity of any vertex: $r = \min_{v \in V} \epsilon(v) = \min_{v \in V} \max_{u \in V} \text{dist}(v, u)$.

The **diameter** D is the maximum eccentricity of any vertex, i.e. the greatest distance between any pair of vertices: $D = \max_{v \in V} \epsilon(v)$.

The **average path length** $\ell = \frac{1}{|V| \cdot (|V|-1)} \sum_{v \neq u} \text{dist}(v, u)$ ("the efficiency of information or mass transport on a network").

The **density** ρ of an undirected G is the ratio of $|E|$ and the number of possible edges, i.e. the number of edges in the complete graph with the same $|V|$:

$$\rho = \frac{2|E|}{|V|(|V|-1)} \qquad \text{(if } \rho \approx 0 \Rightarrow \text{graph is sparse)}$$

Modularity Q measures the strength of division of a graph into clusters (subgraphs, modules). Graphs with high $Q > 0$ have dense connections between the vertices within clusters but sparse between those in different clusters.

Q compares the number of edges within clusters in G with **the expected number of edges in a random graph** regardless of clusters.

Result:

For this task I select dataset bellow:

Bitcoin OTC trust weighted signed network

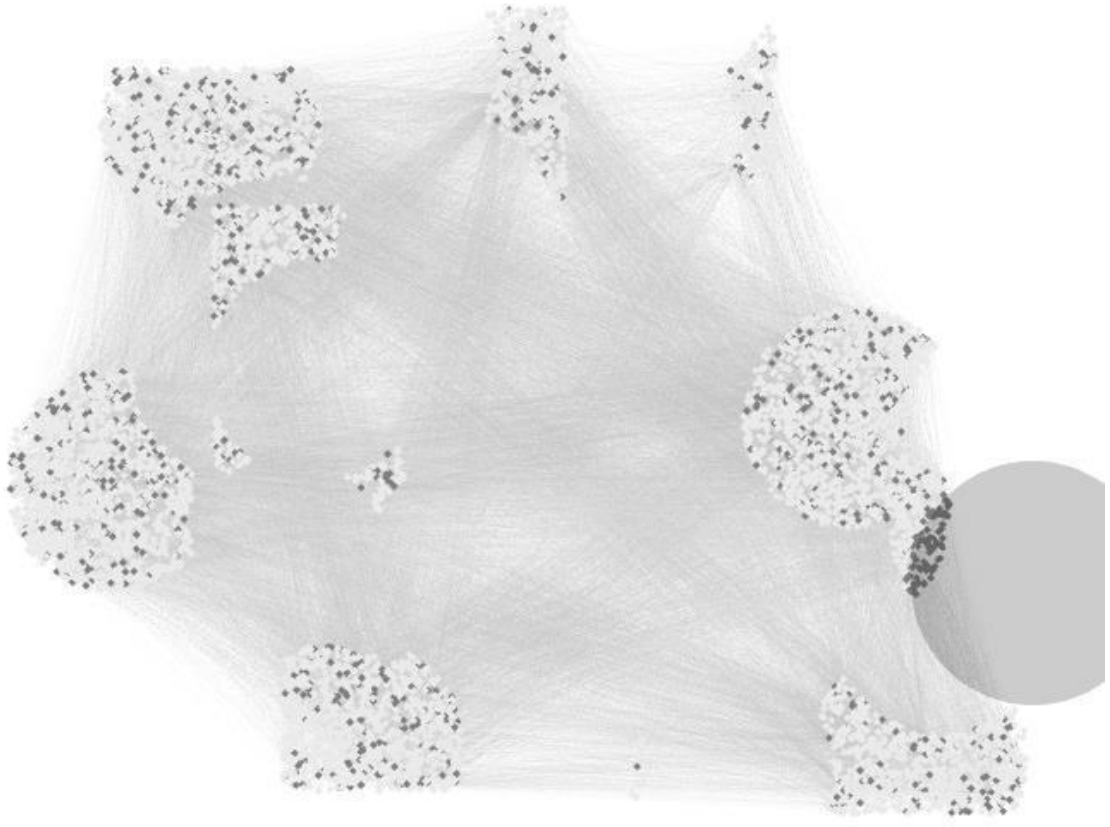
Dataset information

This is who-trusts-whom network of people who trade using Bitcoin on a platform called Bitcoin OTC. Since Bitcoin users are anonymous, there is a need to maintain a record of users' reputation to prevent transactions with fraudulent and risky users. Members of Bitcoin OTC rate other members in a scale of -10 (total distrust) to +10 (total trust) in steps of 1. This is the first explicit weighted signed directed network available for research.

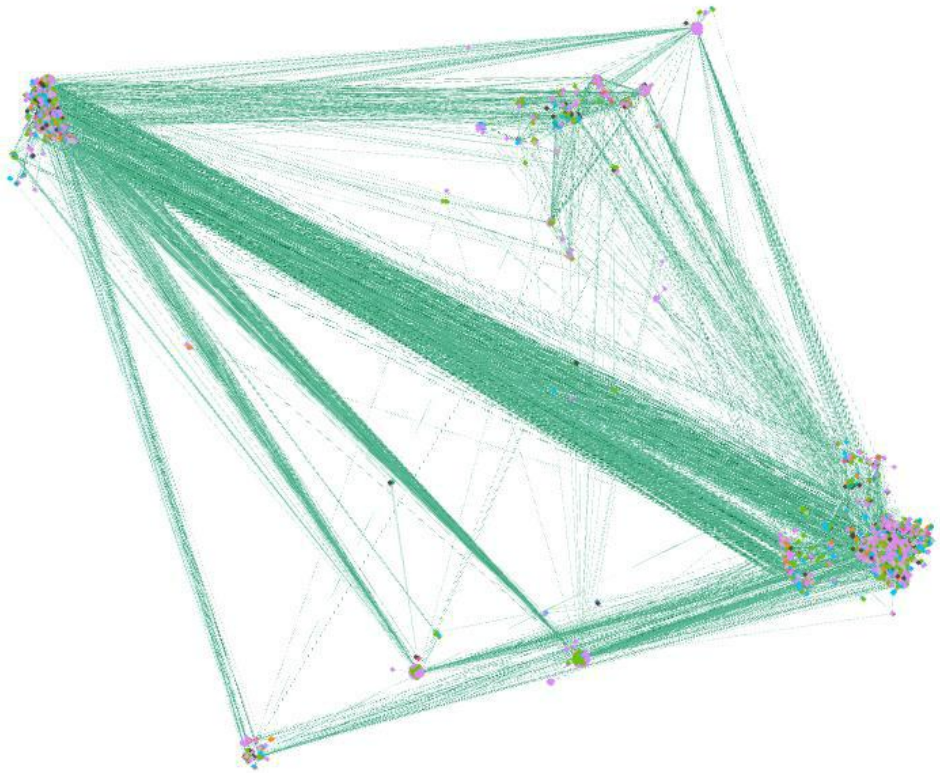
Dataset statistics	
Nodes	5,881
Edges	35,592
Range of edge weight	-10 to +10
Percentage of positive edges	89%

Similar network from another Bitcoin platform, Bitcoin Alpha, is available here.

And how it is present in Gephi:

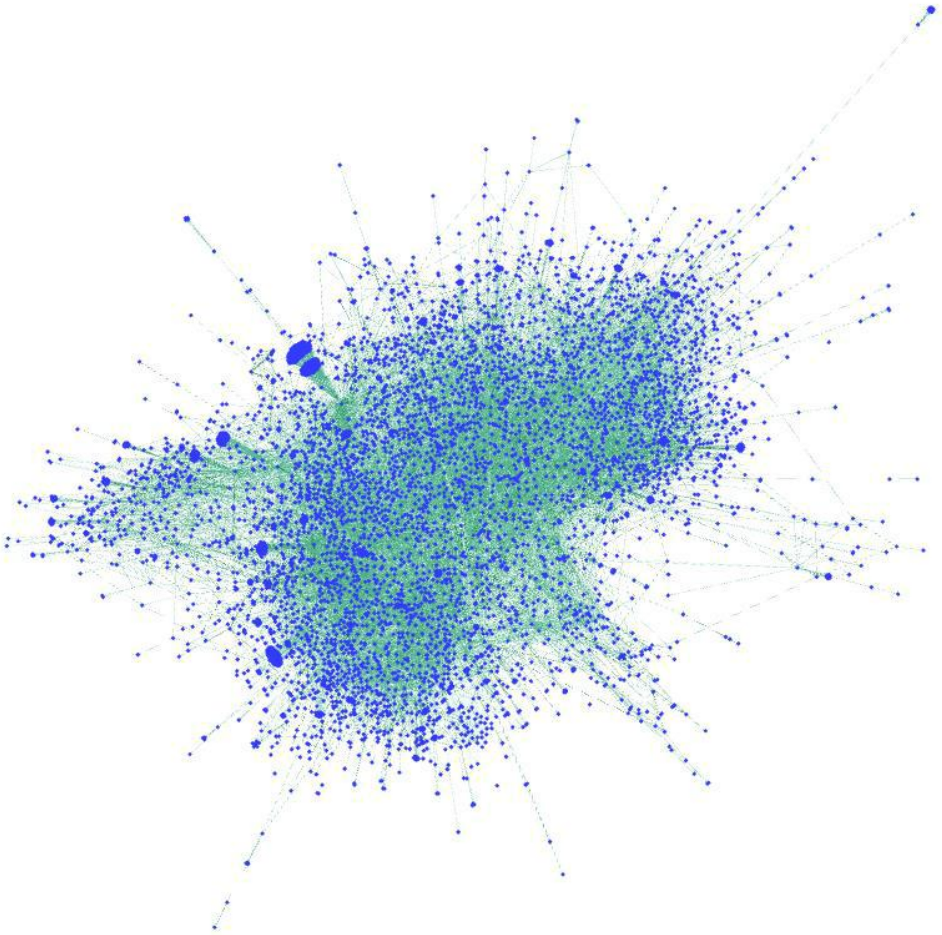


Then I launch three different layout algorithms



Openord

Force atlas



Force atlas 2

After that I calculate some statistics:

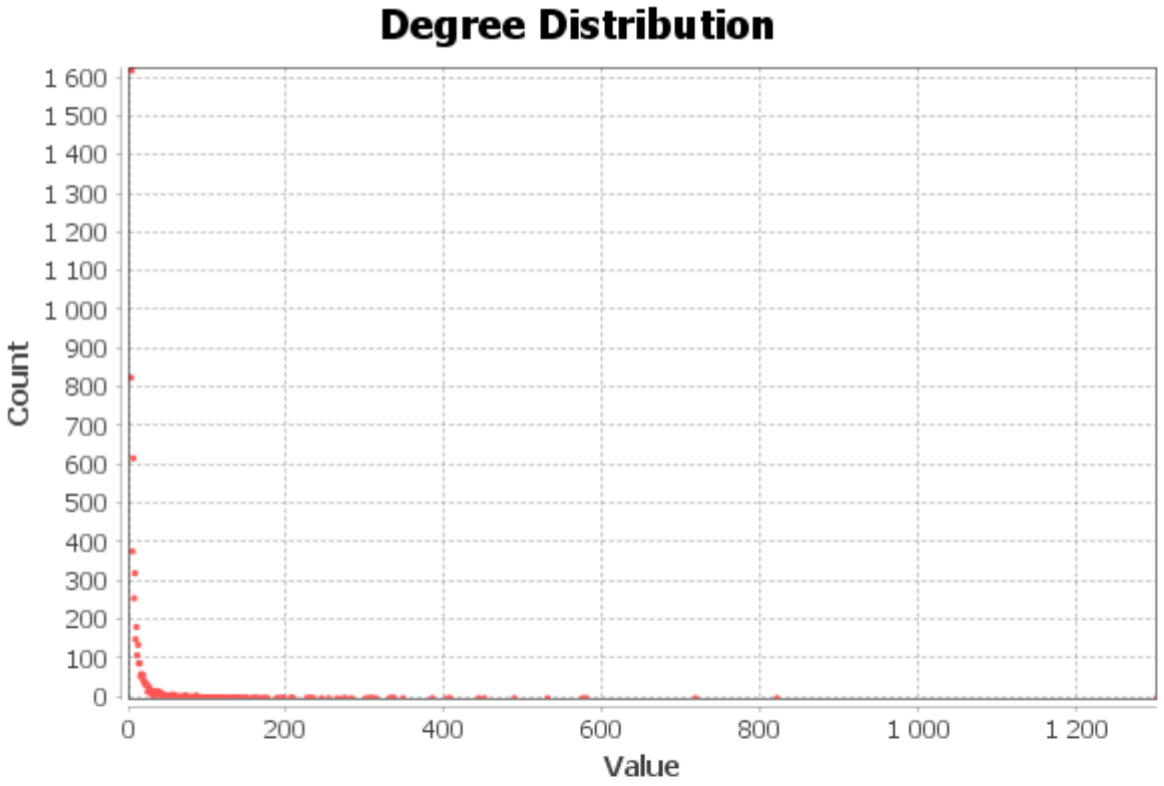
Results:

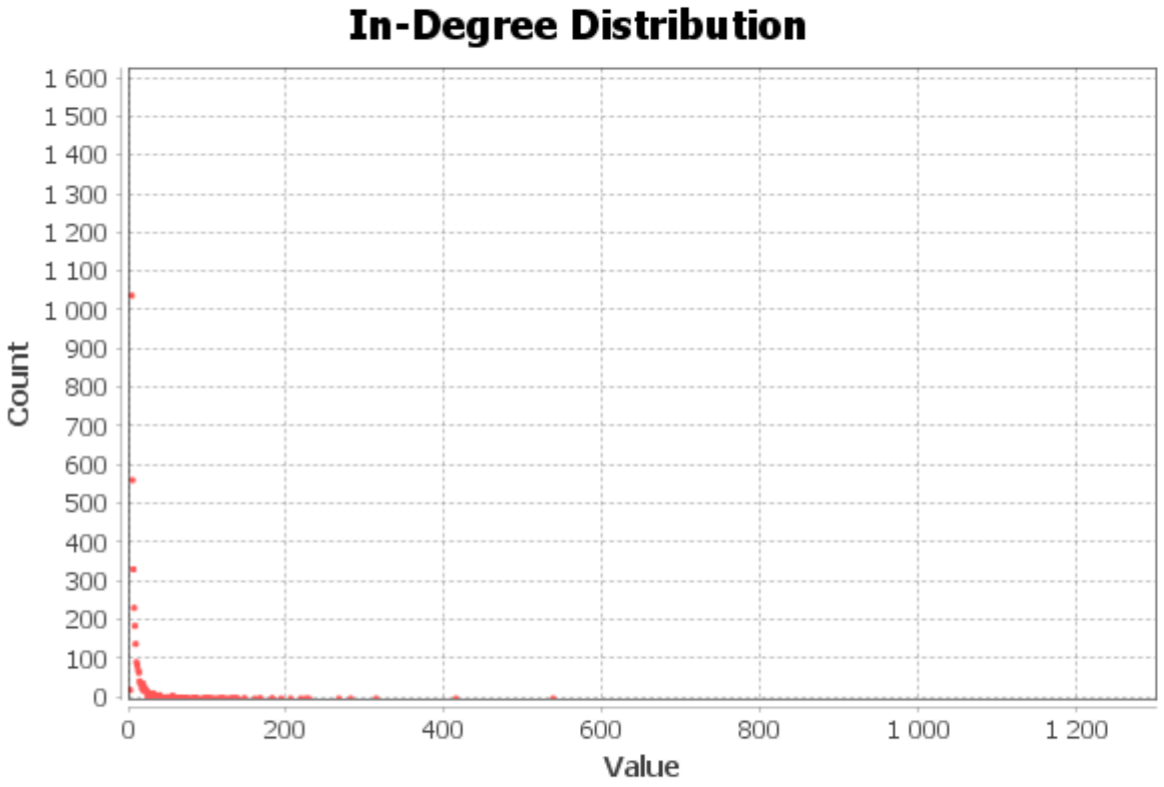
Diameter: 11
Radius: 0
Average Path length: 3.7189130700273005

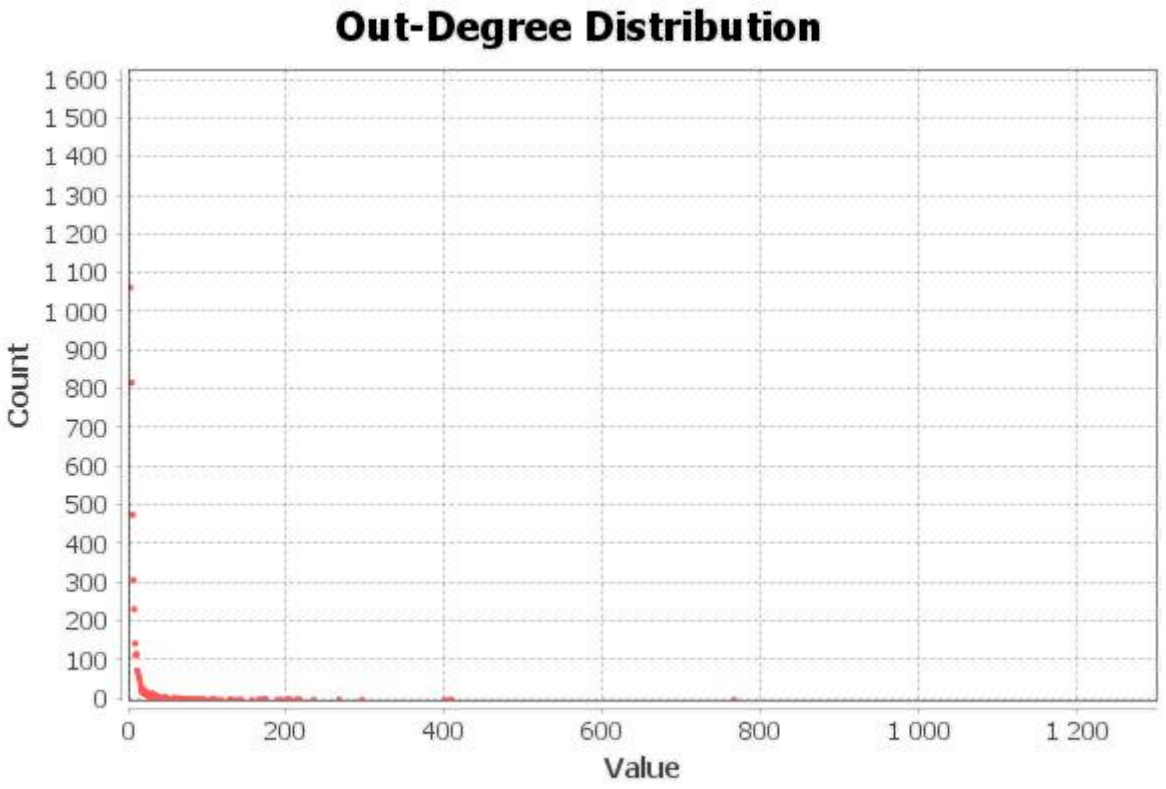
Density: 0,001

Density about zero means, that graph is sparce.

Average Degree: 6,052

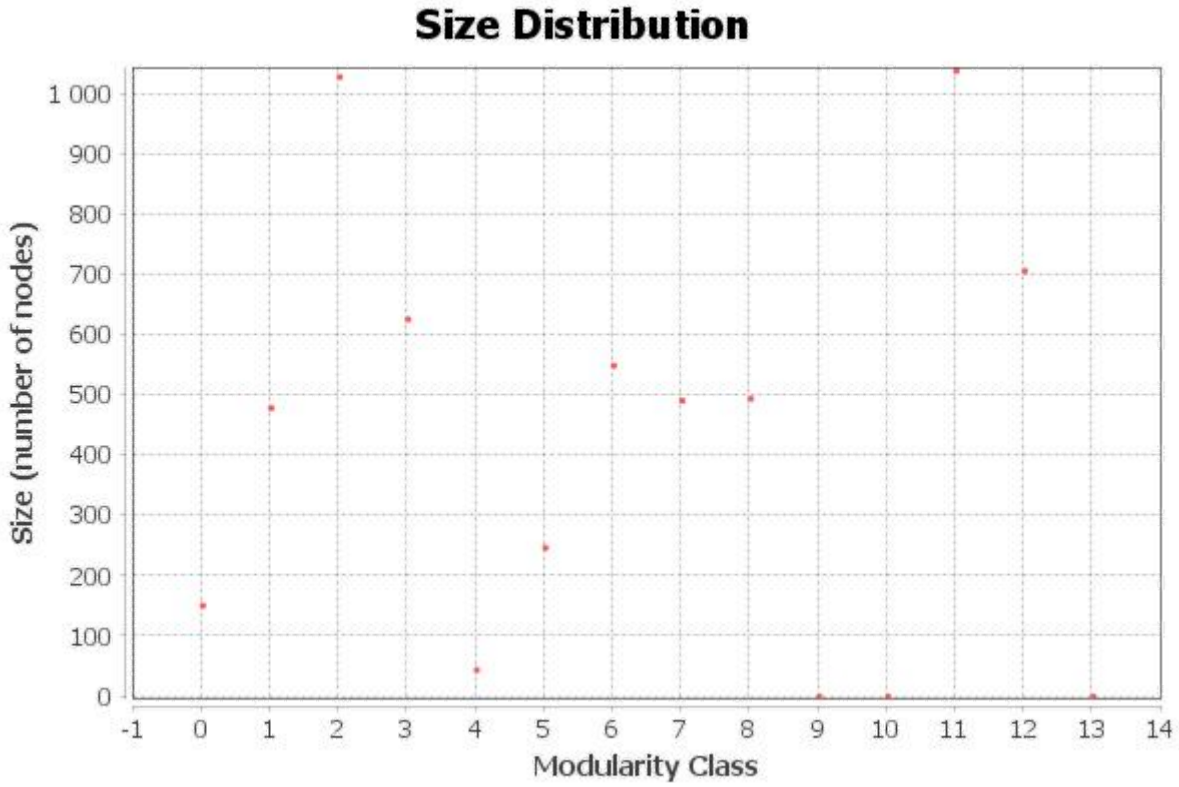






As you may notes from plots above, here not many nodes with degree more then 400. This with graph degree equals 6 meaning that graph is sparce.

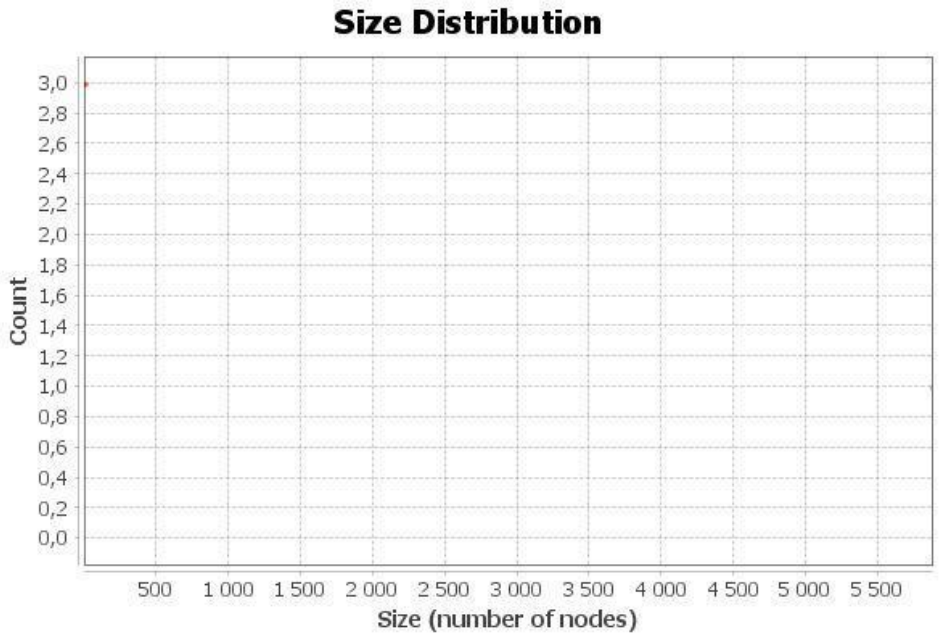
Modularity: 0,483
Modularity with resolution: 0,483
Number of Communities: 14



As you may notes on plot above there only 14 communities, and in vast majority communities contains about 500 nodes. And two of 14 communities contains more then 1000 nodes.

Connected Components:

Number of Weakly Connected Components: 4
Number of Strongly Connected Components: 1144



Conclusion:

In this task I analyses graph with very tool named Gephi, which allow visualize graphs and computate a lot of interesting statistics about passed graph.

Moreover it allow to split and merge graph by component, which may be additional analyze in future to make decision about nature feature.