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
Dmitry Pogrebnoy
J4132c
Accepted by
Dr Petr Chunaev

St. Petersburg
2021

Goal

The use of the network analysis software Gephi.

Formulation of the 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.

Brief theoretical part

The **degree** of of a vertex of a graph is the number of edges that are incident to the vertex. The **in-degree** of a vertex is the number of in-edges for vertex v, the **out-degree** of vertex v, is the number of out-edges for vertex v.

The **eccentricity** $\epsilon(v)$ of v is the greatest distance between v and any other vertex:

$\epsilon(v) = max_{u \in V} dist(v,u)$ ("how far a node is from the node most distant from it").

The **radius** r is the minimum eccentricity of any vertex:

$r = min_{v \in V} \epsilon(v) = min_{v \in V} max_{u \in V} dist(v,u)$.

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** $l = \frac{1}{|V| \cdot (|V|-1)} \sum_{v \neq u} dist(v,u)$ ("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|.

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.

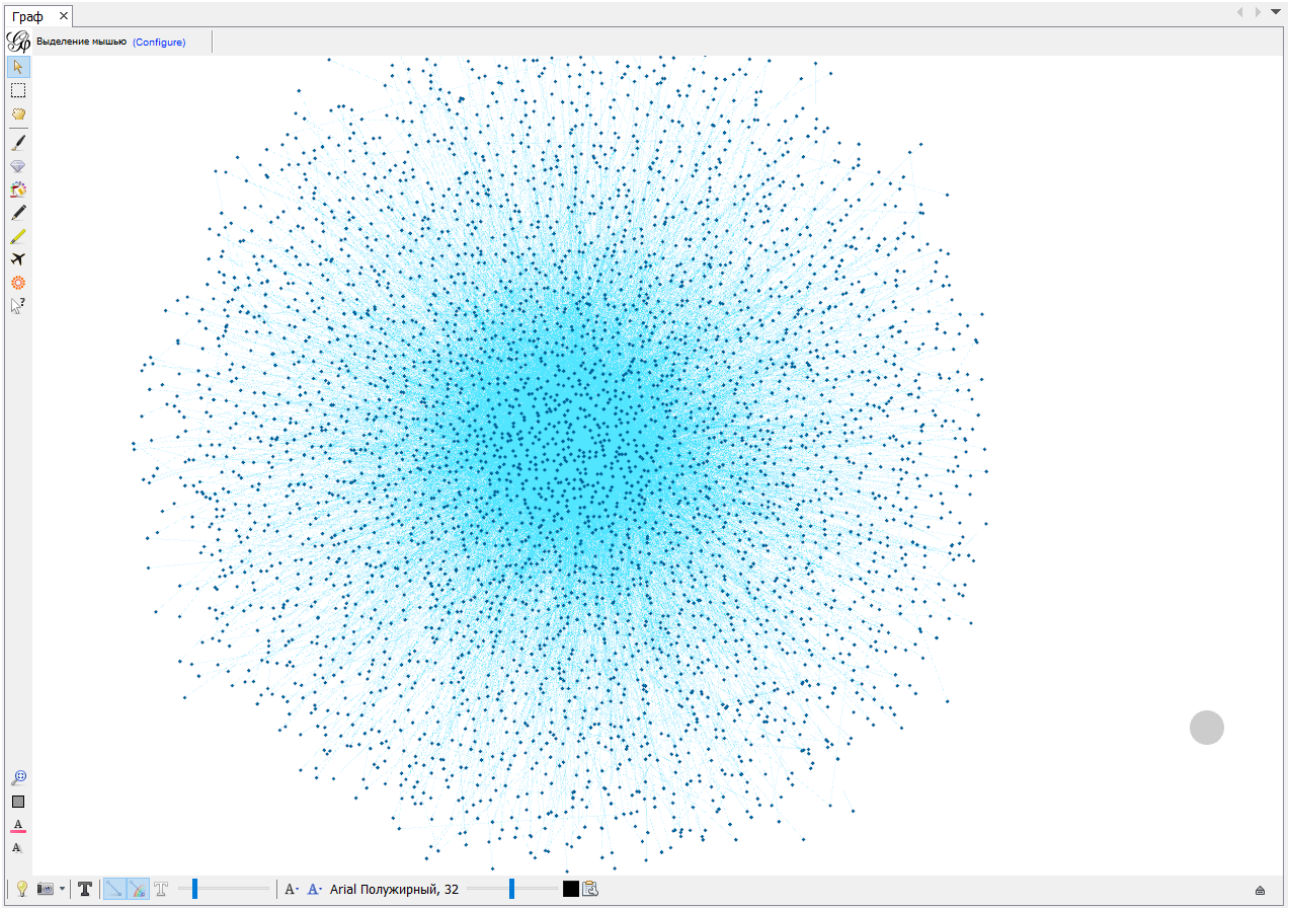
Results

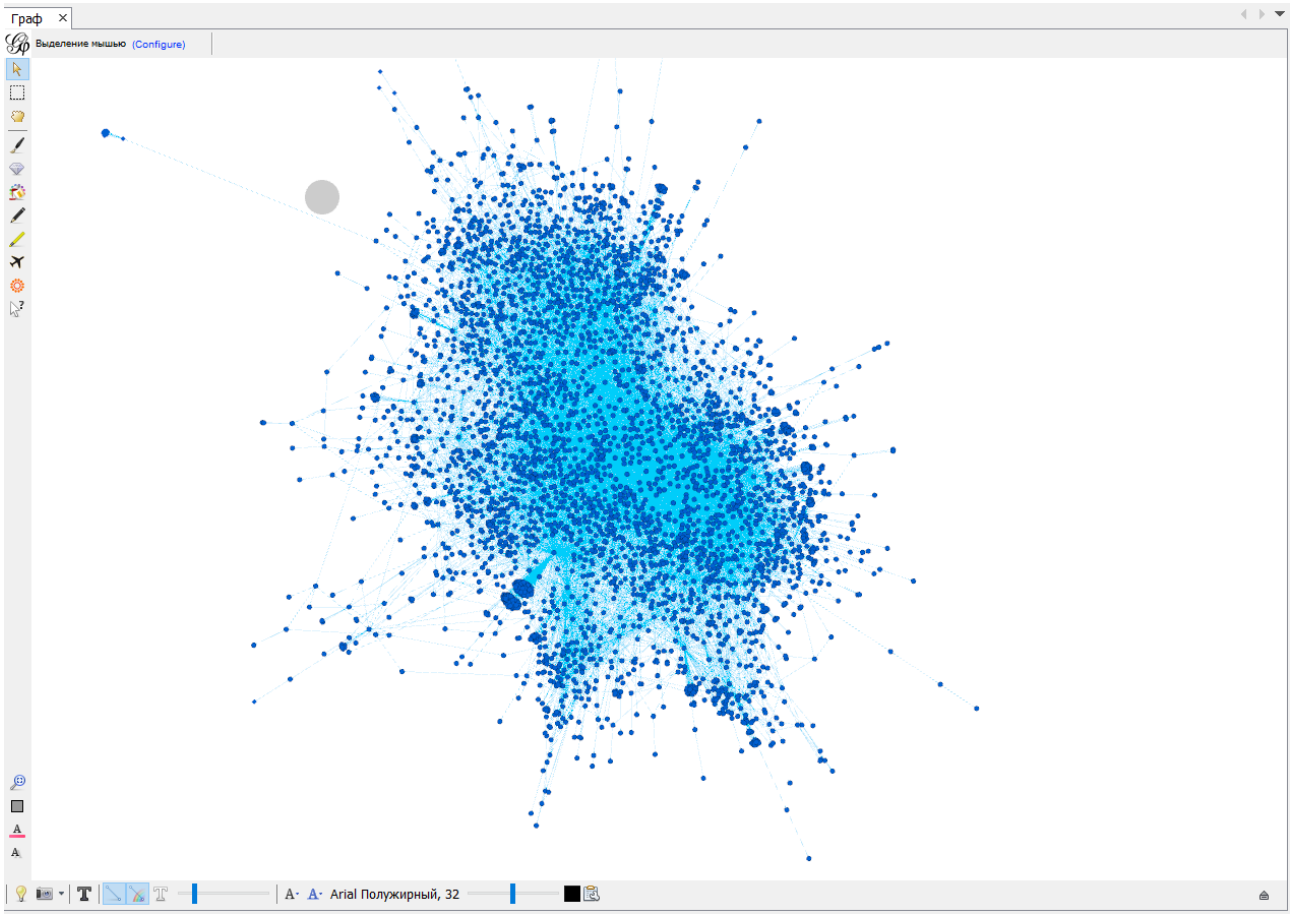
The program *gephi 0.9.2* was used to analyze the graph *Bitcoin Alpha trust weighted signet network* (https://snap.stanford.edu/data/soc-sign-bitcoin-alpha.html). This is

who-trusts-whom network of people who trade using Bitcoin on a platform called Bitcoin Alpha. 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 Alpha rate other members in a scale of -10 (total distrust) to +10 (total trust) in steps of 1.

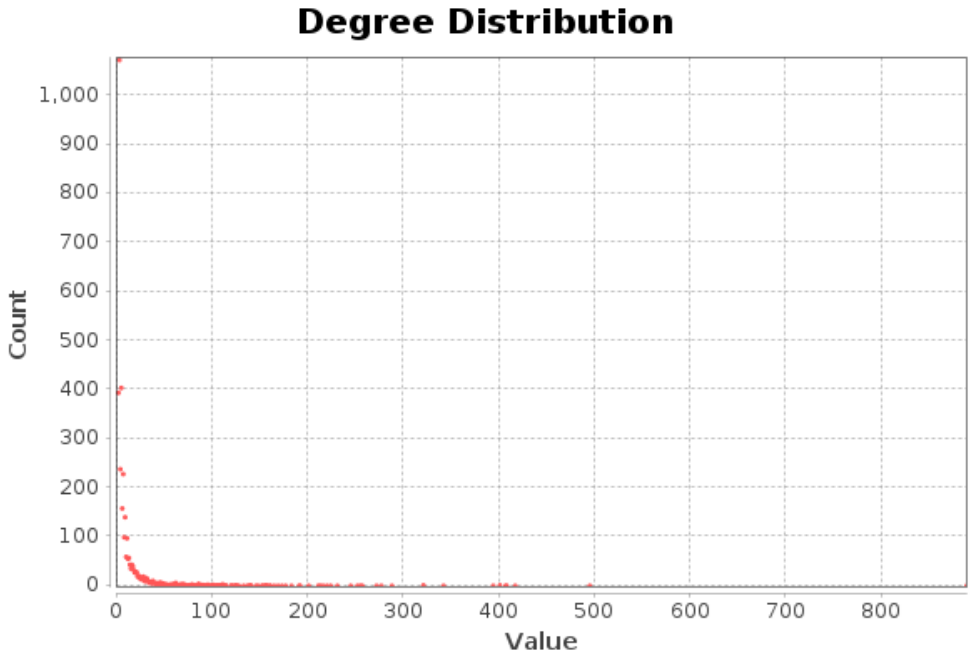
This oriented weighted graph contains **3783 nodes** and **24186 edges**.

Two layouts were obtained for the graph – “Fruchterman Reingold” and “Force Atlas 2”. They were the most informative and allowed to clearly visualize the nodes which have the highest degrees.

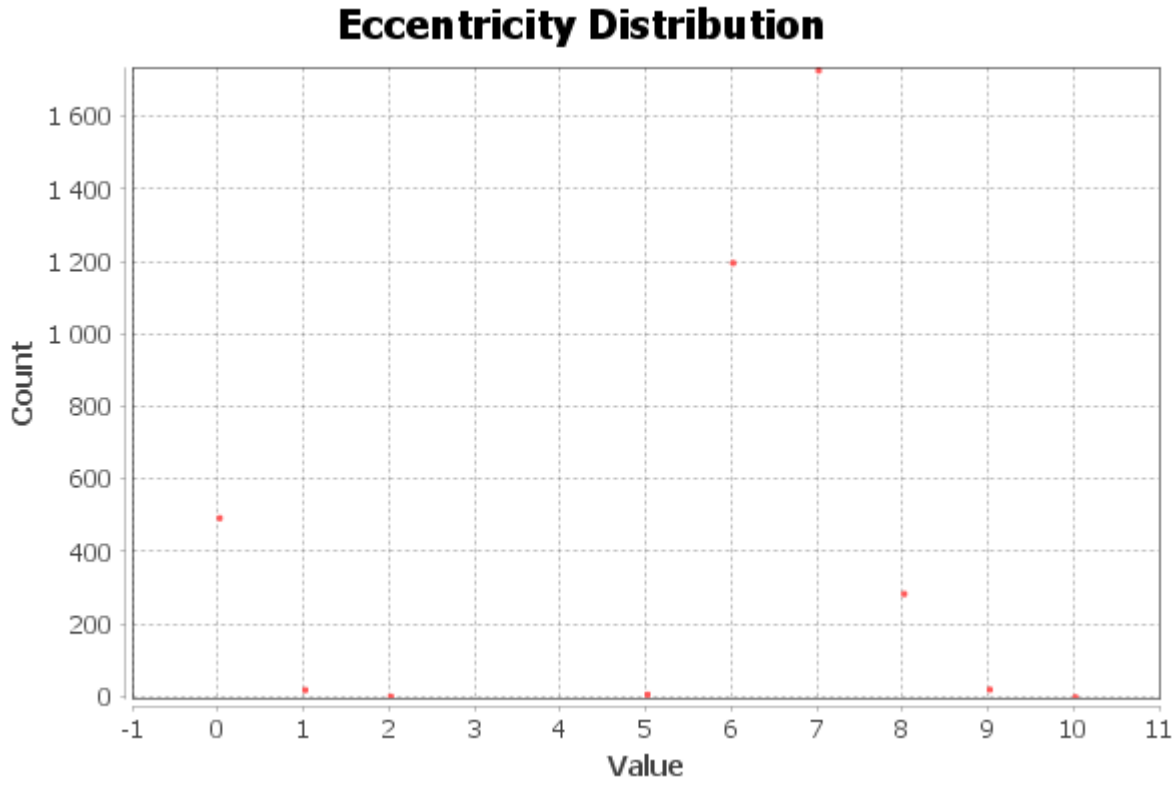




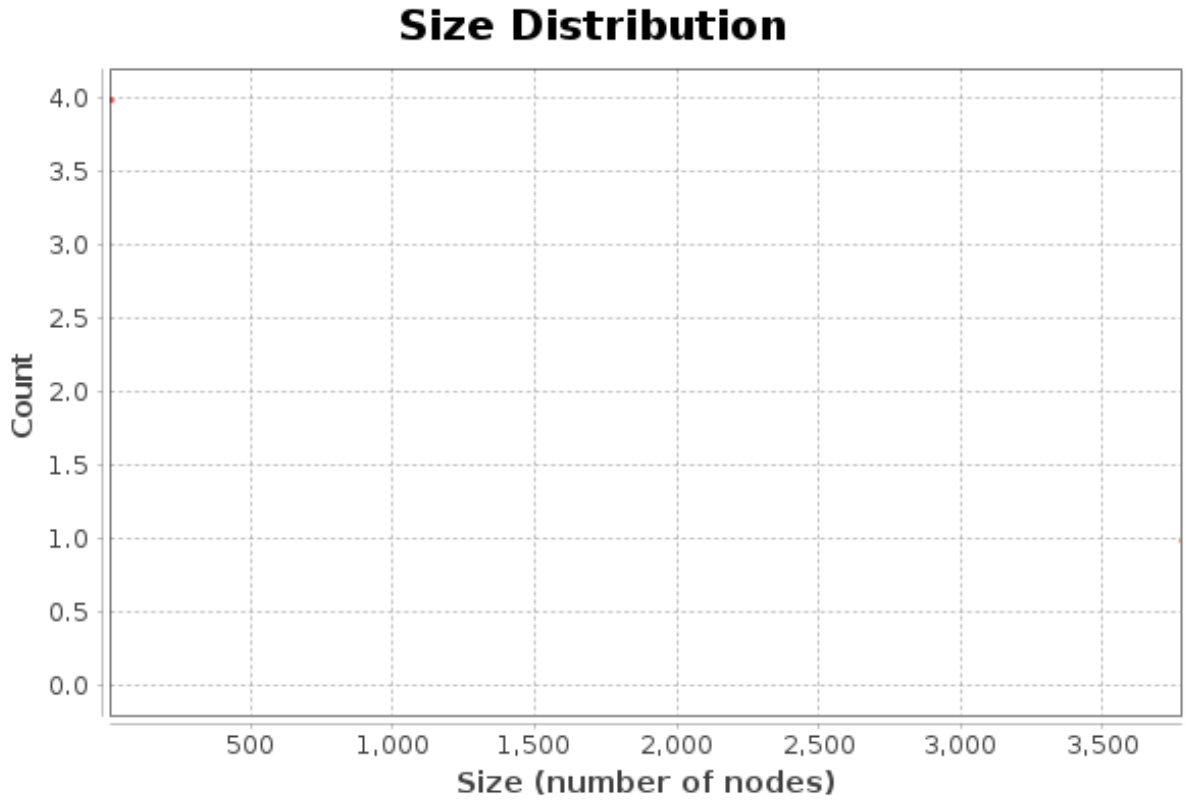
The **average degree** value was **6.393**. Some of the vertices had the degree over 100, and one had degree 1000.



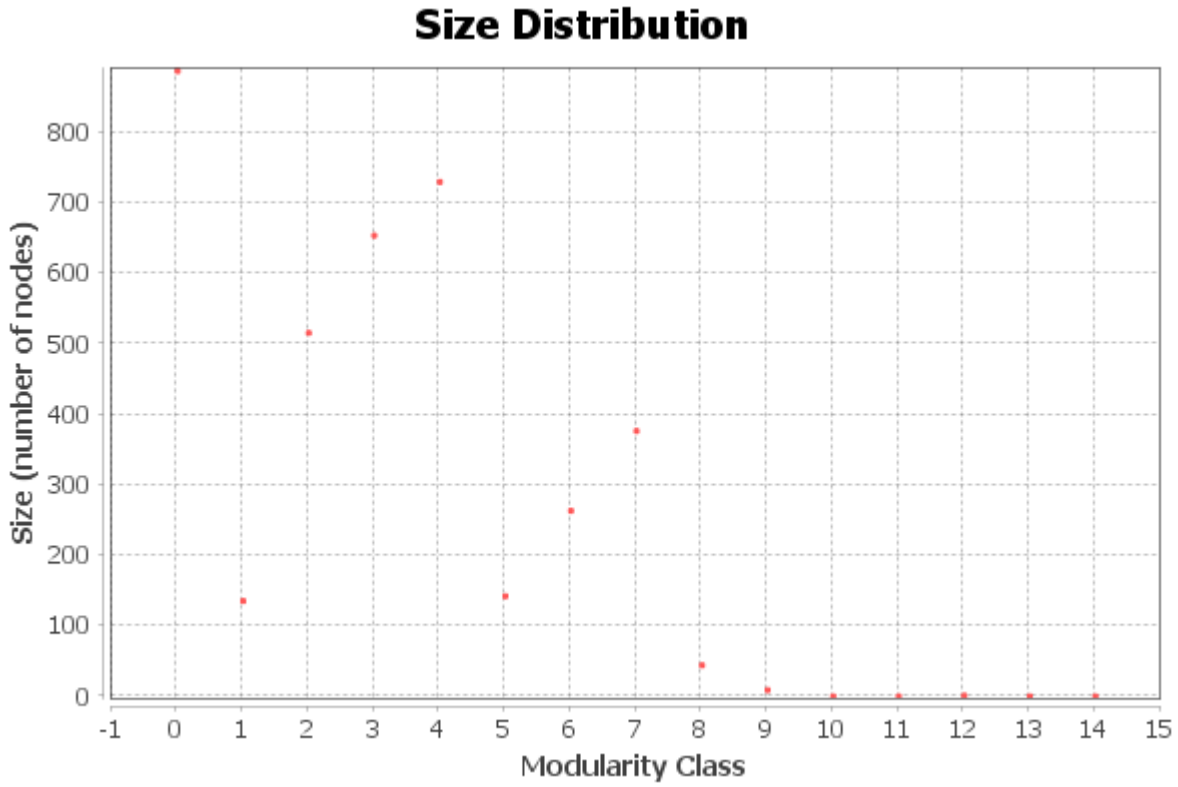
The **diameter** of the graph (longest shortest path) was **10**. The **average path length (APL)** was **3.7**. The **density** of the graph was **0.002**.



The graph had **5 connected components**, but number of **strongly connected components** was **540**.



The **modularity** of the graph was **0.452**, and the **number of communities** was **15.**



Conclusions

The graph *Bitcoin Alpha trust weighted signet network* was analyzed in program Gephi. This graph has several connected components, low density, but the value of modularity is > 0, that is, some of the components are connected more strongly than the others. The diameter of the graph was 10, which says that two completely different users were connected via at most 8 other users, but on average two users were connected with 1.7 other users (average path lenght 3.7).

Appendix

Additional screenshots can be found at https://github.com/DmitryPogrebnoy/algo_itmo/tree/main/task7.