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

Dmitriy Prusskiy

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.

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

Brief theoretical part

Graph vertex degree $d(v)$ is the number of edges that are incident to vertex *v*. **The vertex indegree** is the number of incoming edges for vertex *v*. **The vertex outdegree** is the number of outgoing edges for vertex *v*. **The average vertex degree -** $\bar{d} = \frac{1}{|V|}\sum_{v \in V} d(v)$

dist(v, u) – **distance** (shortest path length) between two vertices.

The eccentricity $\epsilon(v)$ of vertex is the greatest distance between the vertex and any other vertex: $\epsilon(v) = max_{u \in V} dist(v, u)$.

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

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: $l = \frac{1}{|V|(|V|-1)}\sum_{v \neq i} dist(v, i)$.

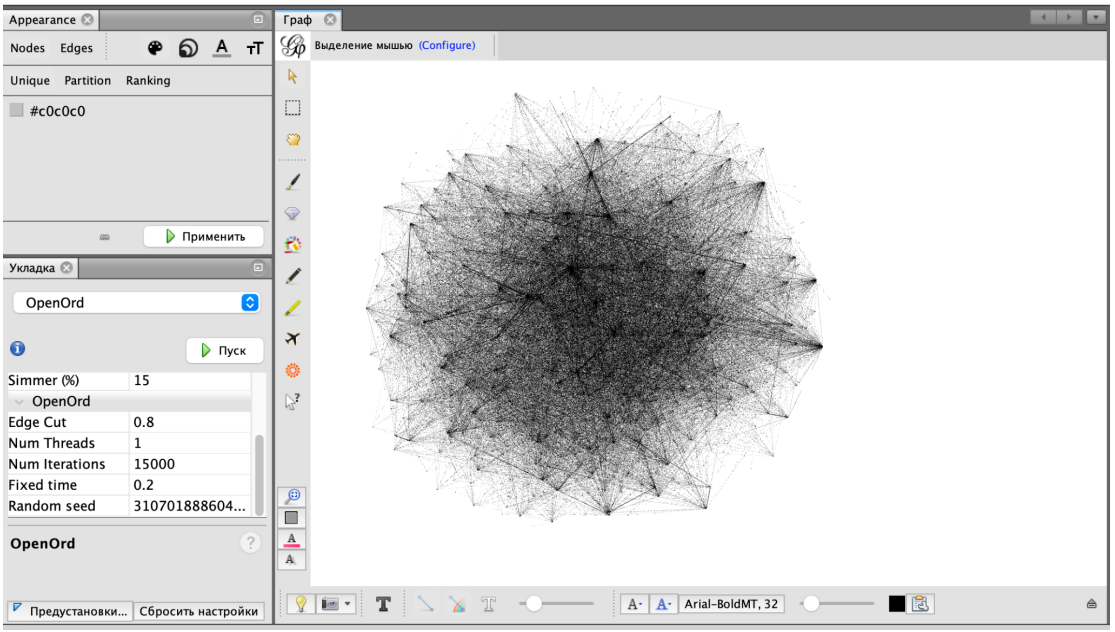
The density ρ of an undirected G is the ratio of |E| and the number of possible edges with the same |V|, i.e. the number of edges in the complete graph with |V| vertices: $\rho = \frac{2|E|}{|V|(|V|-1)}$. Sparse graph is a graph with $\rho \approx 0$.

Results

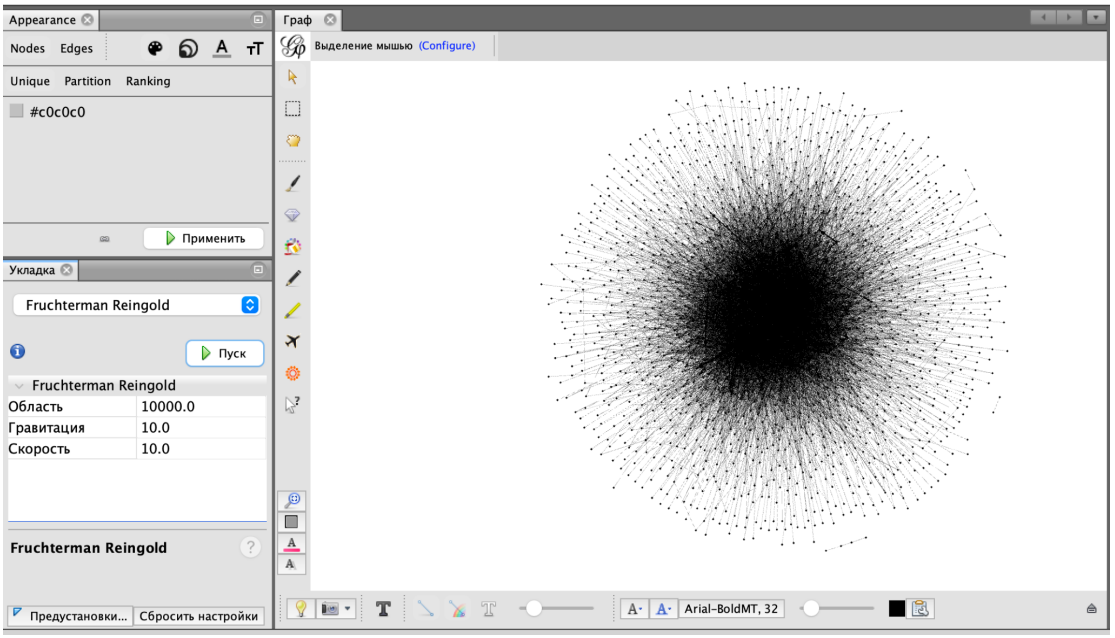
The program Gelphi 0.9.2 was used to analyze the graph “College messaging temporal network” https://snap.stanford.edu/data/CollegeMsg.html. This dataset is composed of private messages sent on an online social network at the University of California, Irvine.

This oriented unweighted graph contains 1899 nodes and 20296 edges.

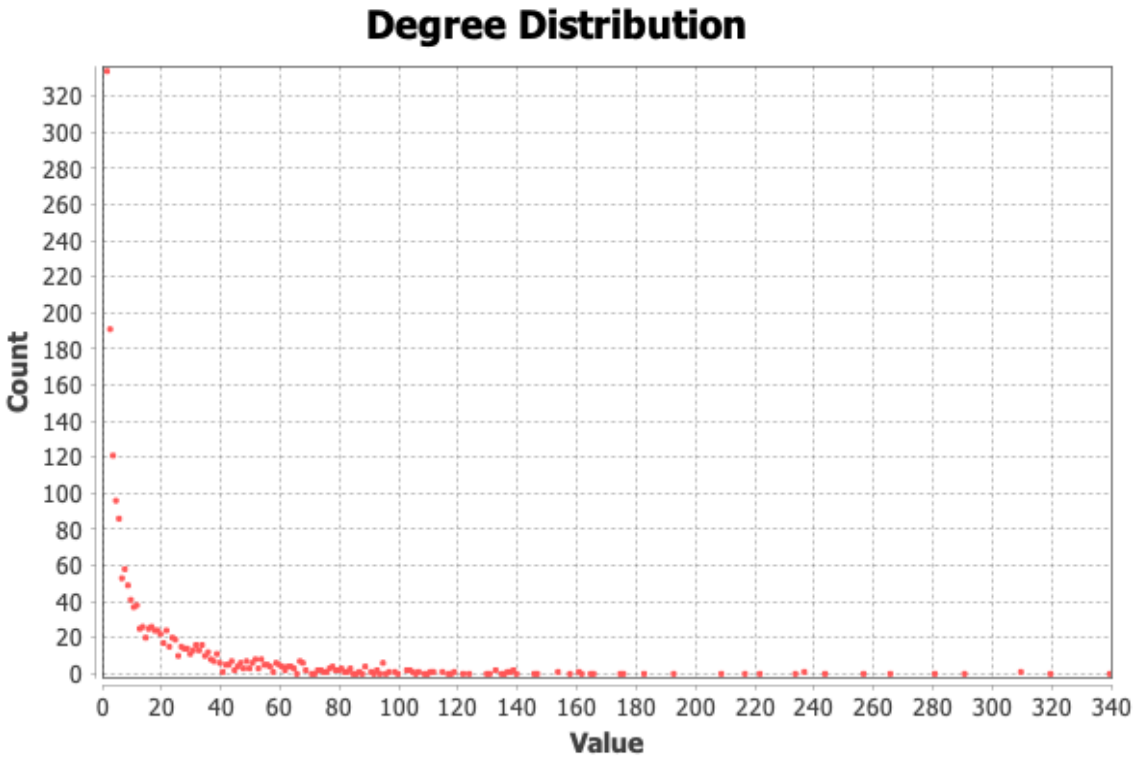
To visualize the graph OpenOrd layout was used with 15000 iterations. It highlights popular users.



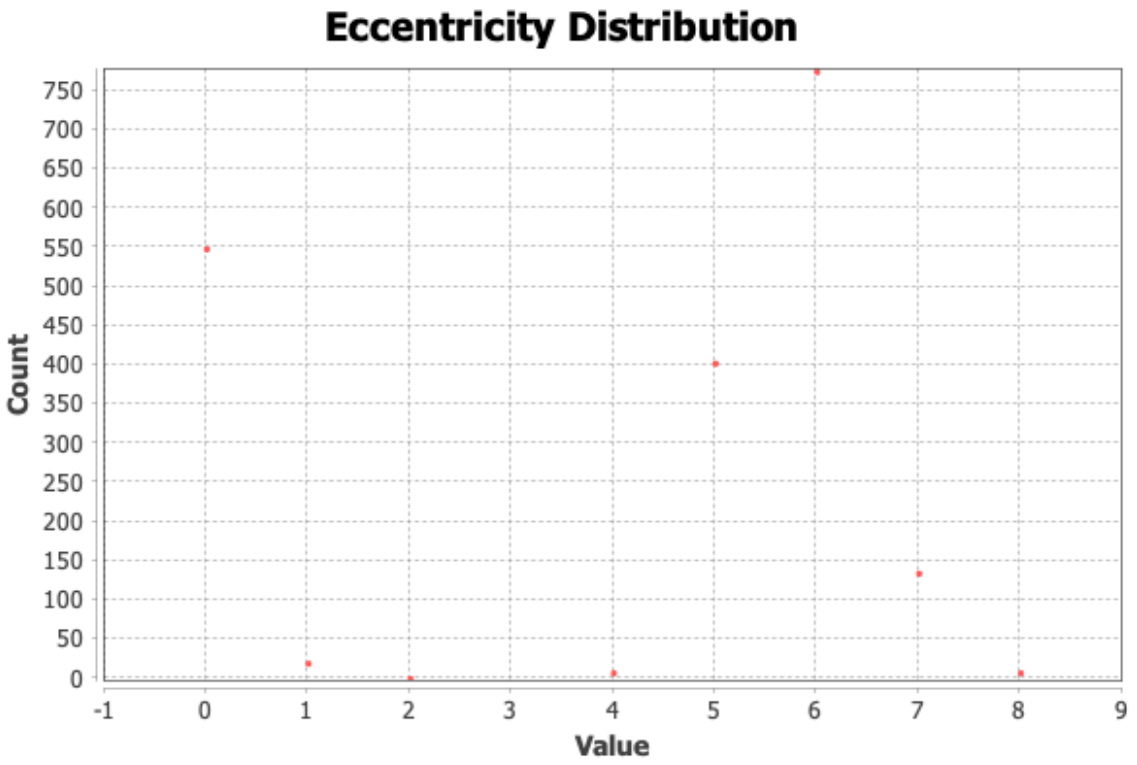
Another layout used for appropriate visualization is Fruchterman Reingold with base parameters. It shows that a lot of users sent or received only one or two messages.



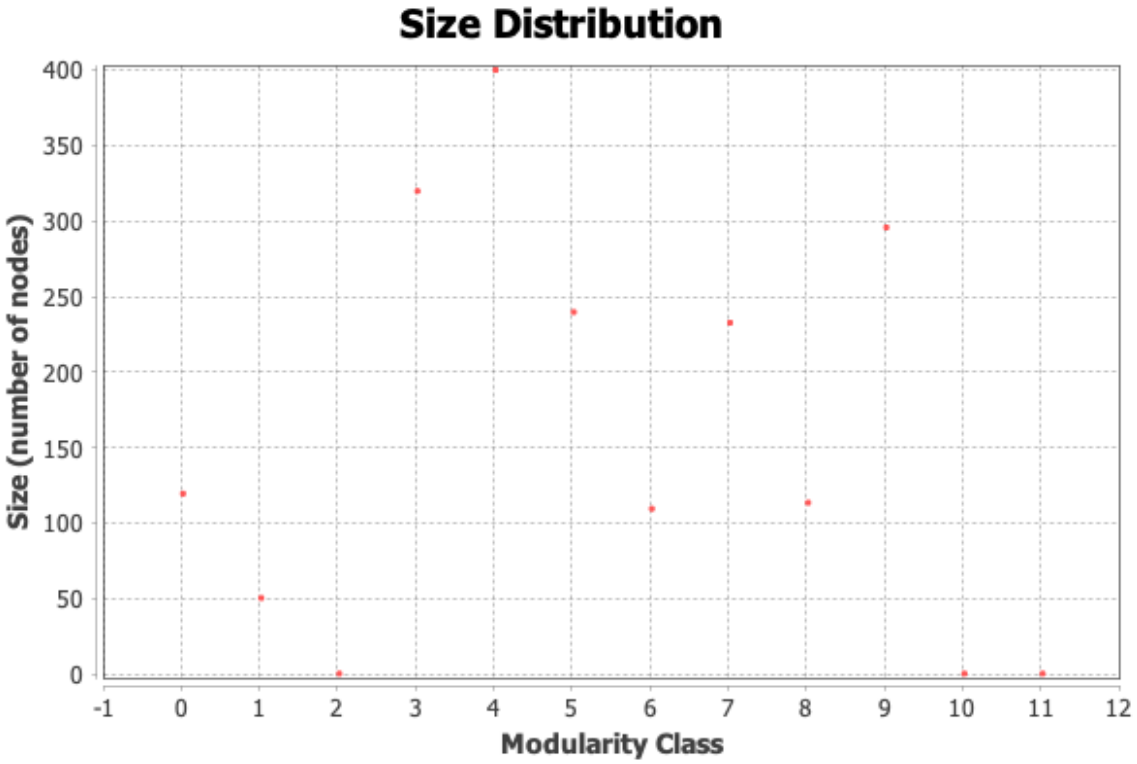
The average degree is 10.69, and the maximum degree is 340.



The diameter of the graph is 8. This means that every user knows any other in less than 7 handshakes. The average path length is 3.2. The density is 0.006.



The modularity of the graph is 0.25, and the number of communities is 12. Most communities have from 3 to 9 users.



The graph has 4 weakly connected components and 601 strongly connected components.

Conclusions

The goal of this study was to use the network analysis software Gephi. College messaging temporal network was analyzed. Results show that most users sent less than 11 messages, graph has 12 communities and 4 weakly connected components. Results of the HITS algorithm and PageRank stats were not presented in the report because these stats do not make sense for this dataset.

Appendix

Source code and full reports can be found at
https://github.com/T1MAX/itmo_algorithms/tree/main/task_7