

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:

Putnikov Semyon

J4132c

Accepted by:

Dr Petr Chunaev

St. Petersburg

2020

1. Goal

The use of the network analysis software Gephi.

2. Problems and methods

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 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.

3. Results

Arxiv GR-QC (General Relativity and Quantum Cosmology) collaboration network is from the e-print arXiv and covers scientific collaborations between authors papers submitted to General Relativity and Quantum Cosmology category.

3.1 Make graph from data

Initial data was txt file with structure: “node node” on each row that means a edge of the graph. Preprocessing:

1. Rename header to Source, Target.
2. Replace all tabs with comma to make csv file.

Now we can open our data as graph in Gephi.

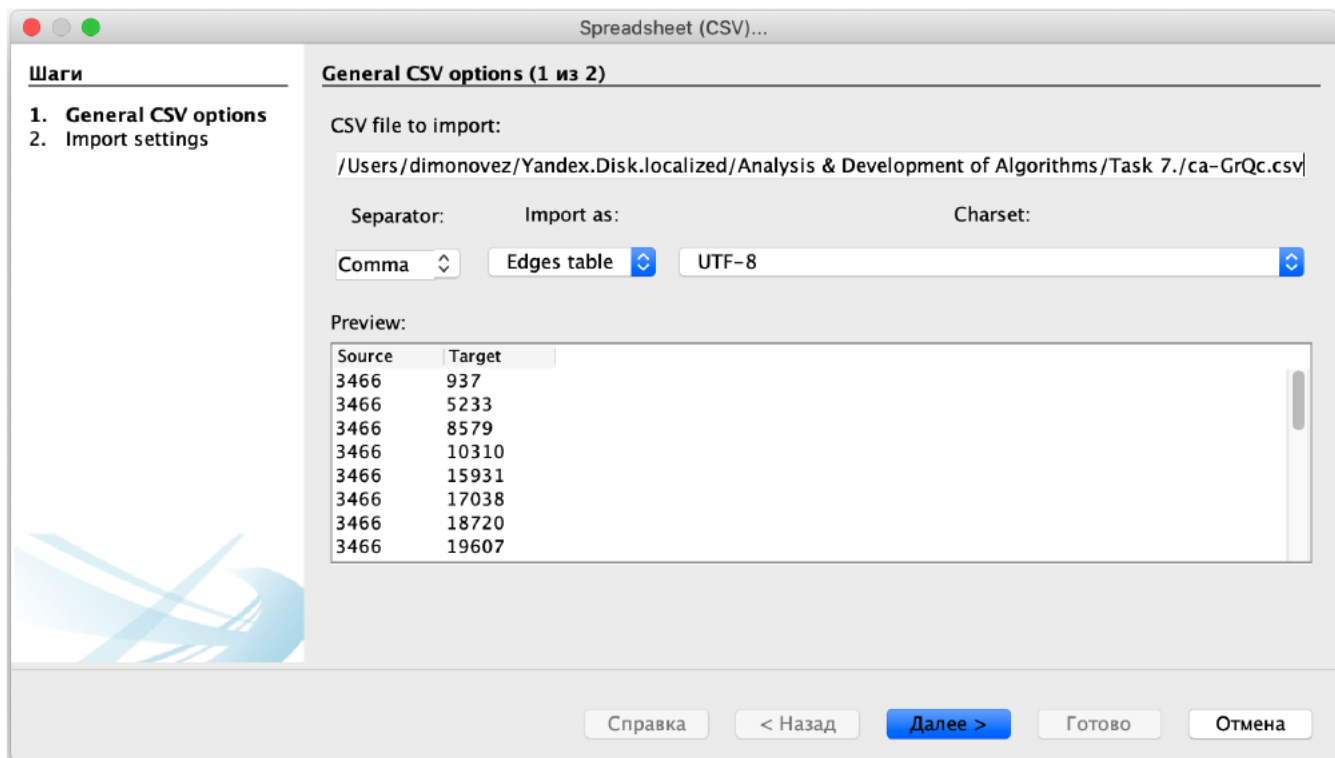


Figure 1: Import data.

In each of the graph layouts we can see that graph have many connectivity components. That means that authors in each component are closely related and are not published with anyone else.

3.2 Statistics

- $|V| = 5242$
- $|E| = 14496$
- Average degree: 5.528

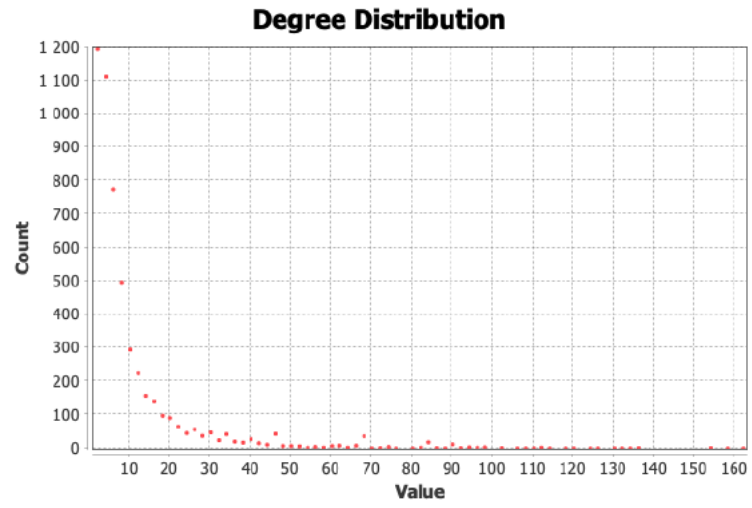


Figure 2: Degree Distribution.

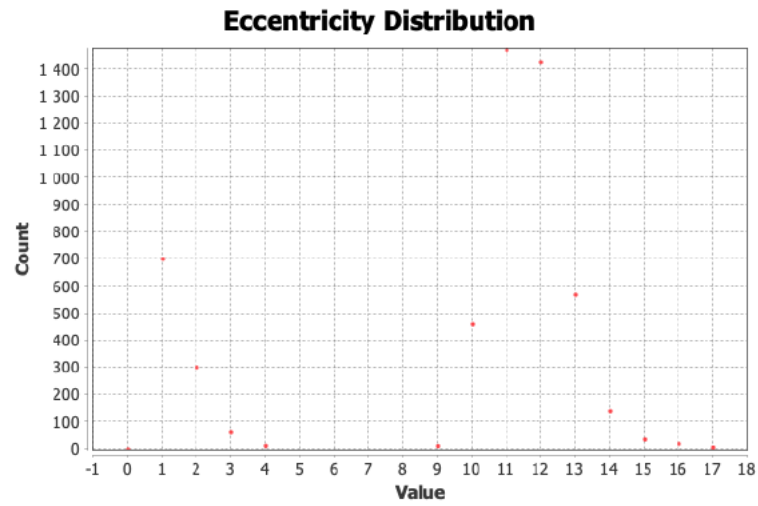


Figure 3: Eccentricity Distribution.

- Diameter: 17
- Average path length: 6.05
- Radius: 1
- Density of the graph: 0.001 (Sparse)
- Modularity: 0.861
- Modularity with resolution: 0.861
- Number of Communities: 389

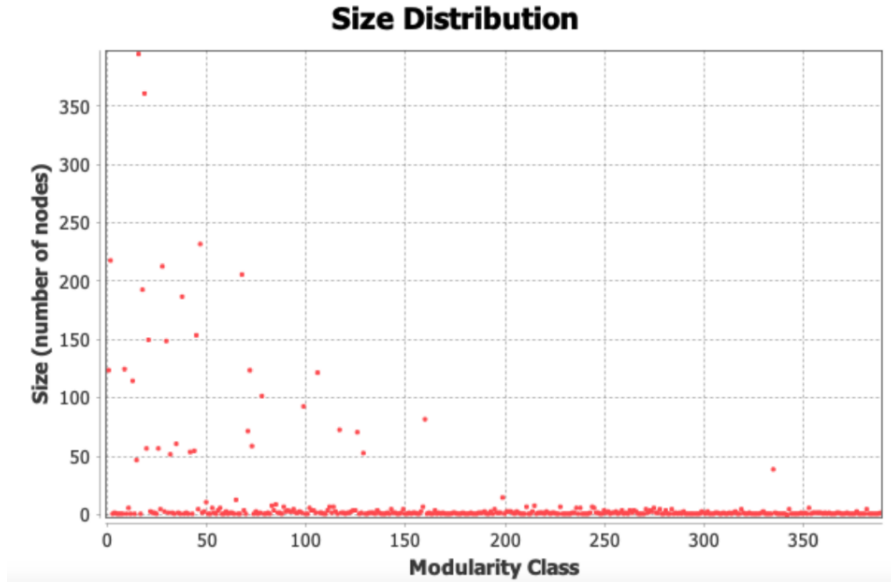


Figure 4: Size Distribution.

Статистика по графу	
Средняя степень	5,528
Средняя взвешенная степень	5,528
Диаметр графа	17
Плотность графа	0,001
HITS	
Модулярность	0,861
PageRank	
Связные компоненты	355
Статистика по узлу	
Средний коэффициент кластеризации	0,687
Eigenvector Centrality	
Статистика по ребру	
Средняя длина пути	6,049

Figure 5: Statistics.

4. Conclusion

General Relativity and Quantum Cosmology collaboration graph have sparse structure with 355 components. Authors in each component are closely related and are not published with anyone else. Authors have an average of 6 (5.528) people with whom they published together. Average path from one author to another 6.05 people.

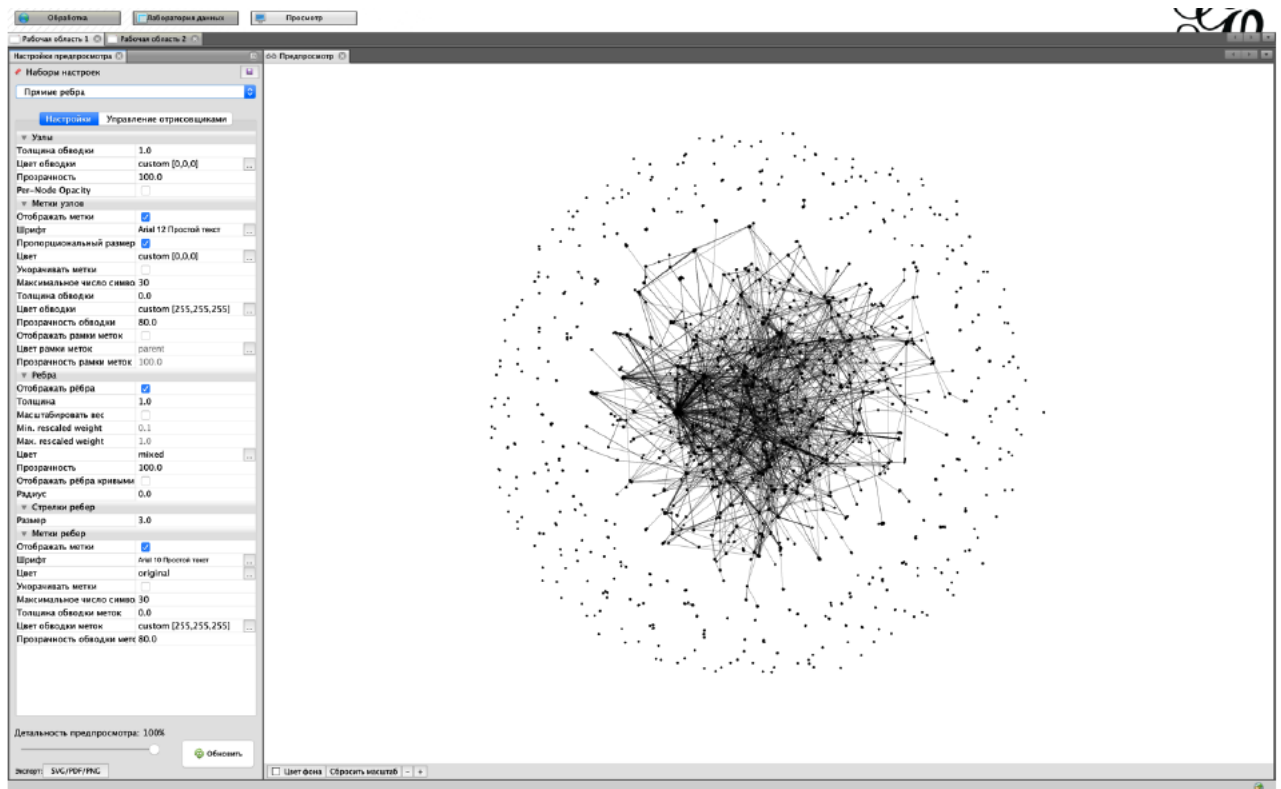


Figure 6: OpenOrd layout.

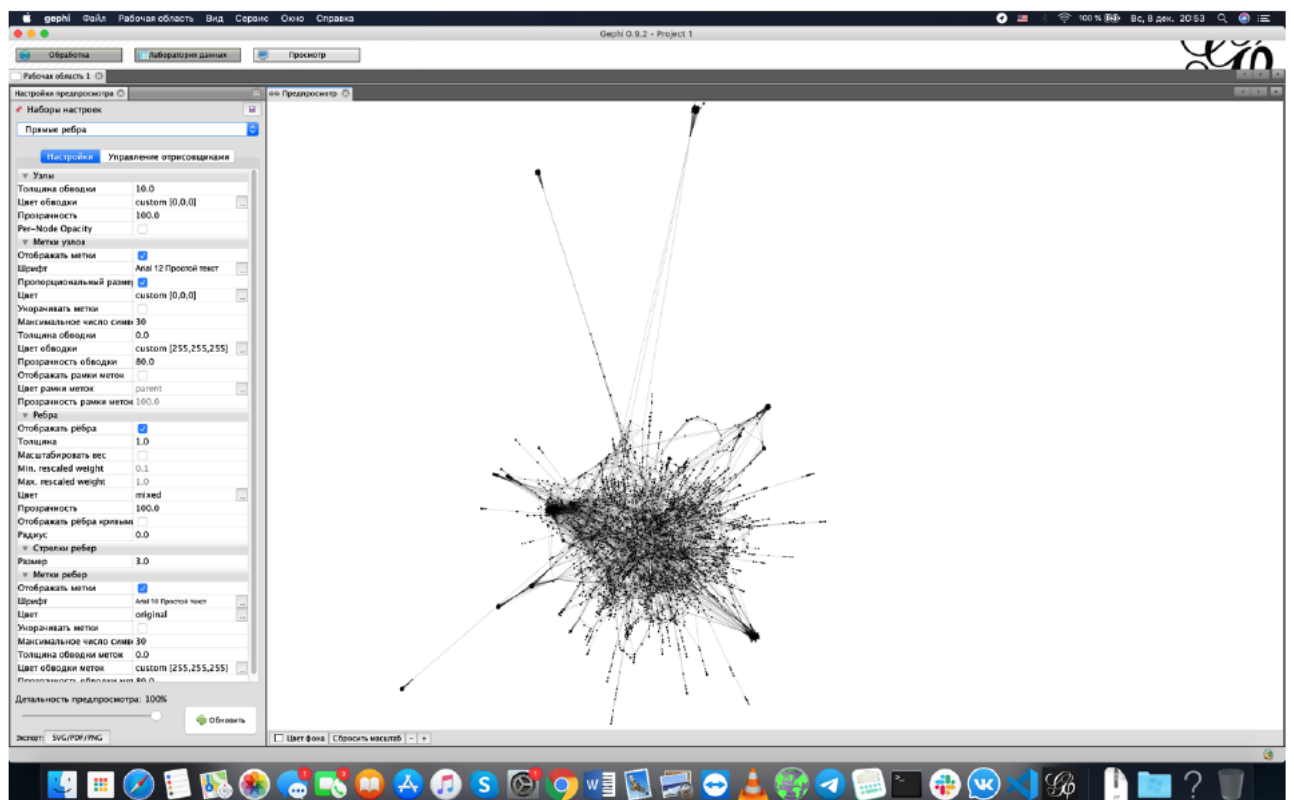


Figure 7: ForceAtlas2 layout.

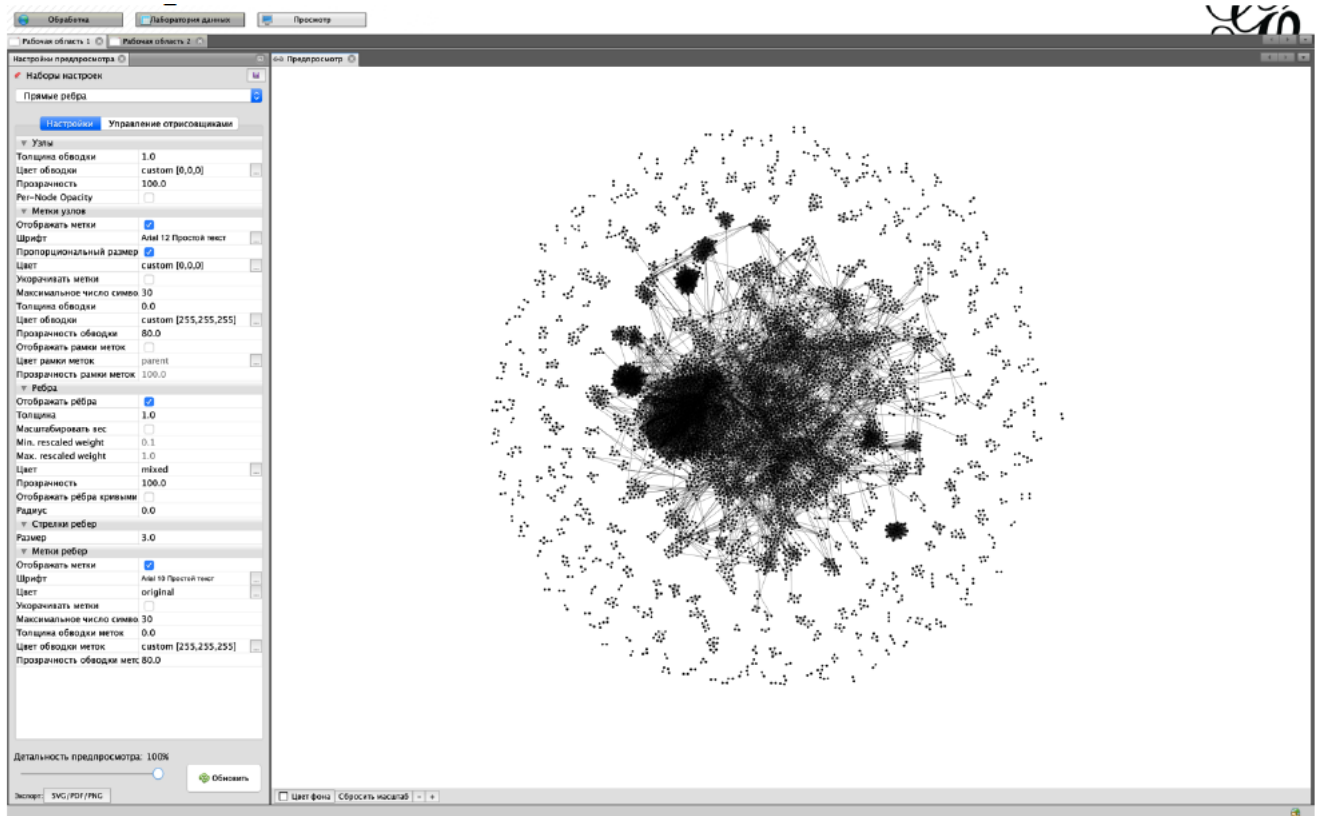


Figure 8: Noverlap layout.