

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  
Denis Zakharov,  
Maxim Shitilov  
Academic group J4132c  
Accepted by  
Dr Petr Chunaev

St. Petersburg  
2022

## **Goal**

The use of the network analysis software Gephi.

## **Formulation of the problem**

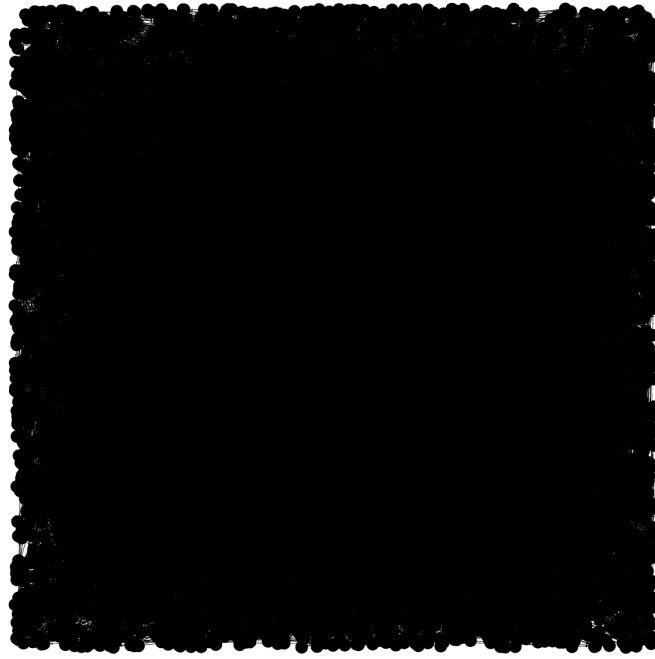
1. Download and install Gephi from <https://gephi.org/>.
2. Choose a network dataset from <https://snap.stanford.edu/data/> with number of nodes at most 10,000. You are free to choose the network nature and type (un/weighted, un/directed).
3. Change the format of the dataset for that accepted by Gephi(.csv,.xls,.edges, 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**

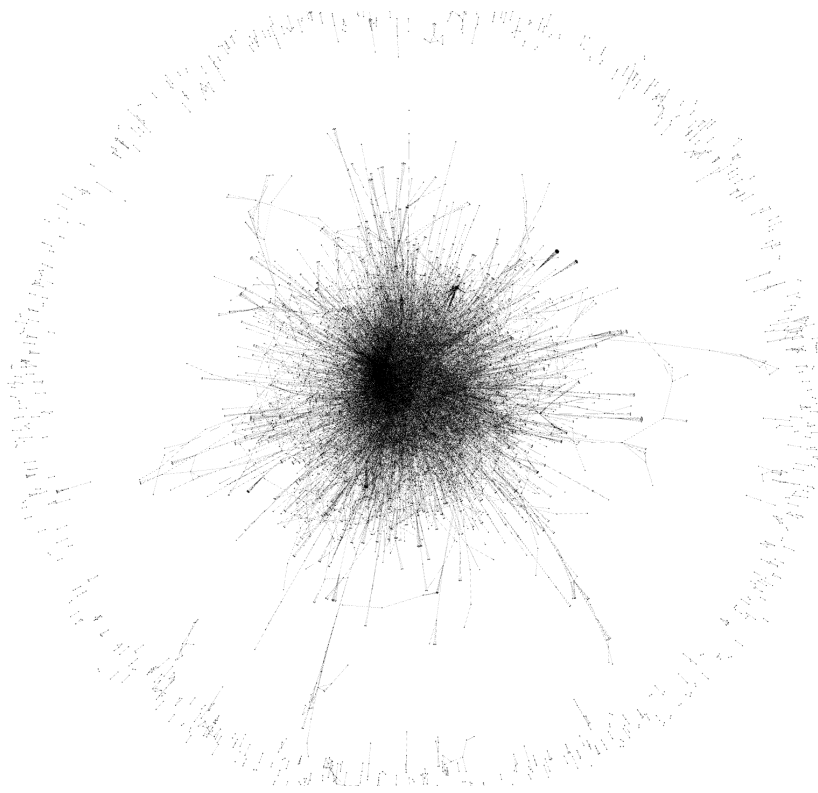
In the domain of mathematics and computer science, graph theory is the study of graphs that concerns with the relationship among edges and vertices. It is a popular subject having its applications in computer science, information technology, biosciences, mathematics, and linguistics to name a few.

Gephi is a visualization application developed in the Java language. It is mainly used for visualizing, manipulating, and exploring networks and graphs from raw edge and node graph data. It is a free and open-source application. It is built on the top of the Netbeans Platform and uses OpenGL for its visualization engine. It runs on Windows, macOS, and Linux. It is an excellent tool for data analysts and data science enthusiasts to explore and understand graphs. It is similar to Photoshop but deals with graph data. The user interacts with the representation, manipulates the structures, shapes, and colors to reveal hidden patterns. The primary goal is to enable the user to make a hypothesis, discover hidden patterns, isolate structure singularities and defects during data sourcing.

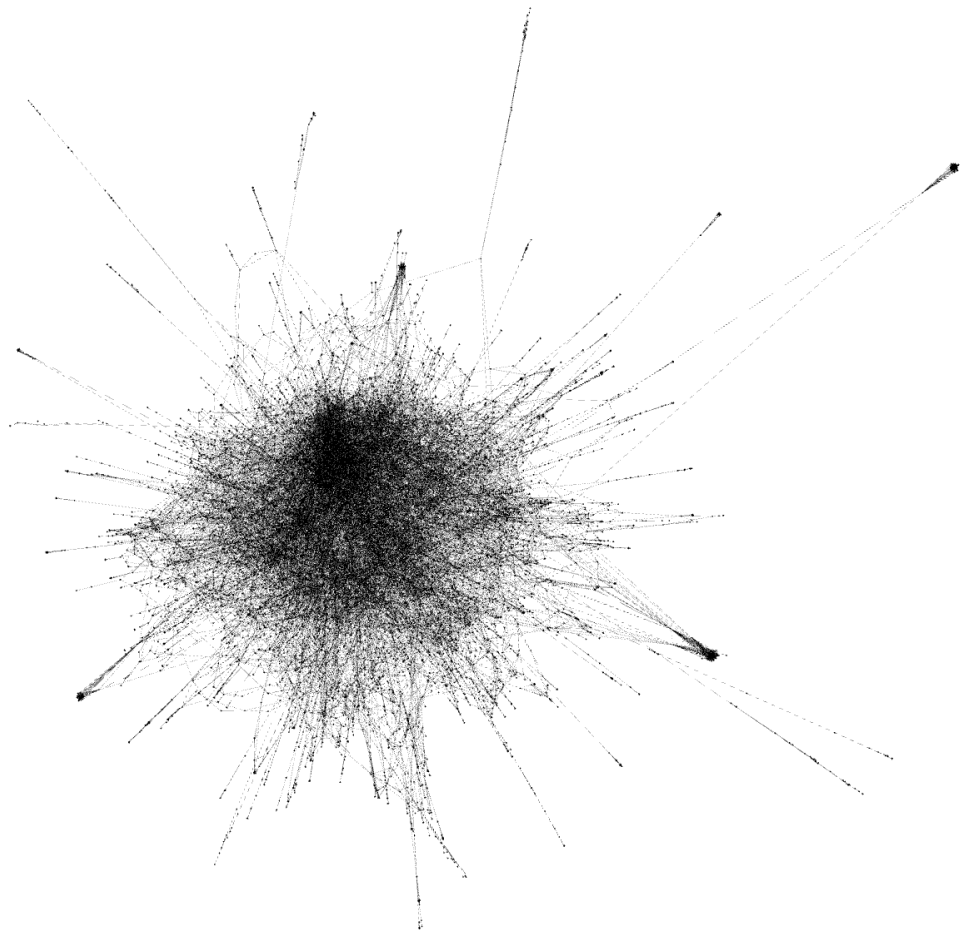
## **Results**



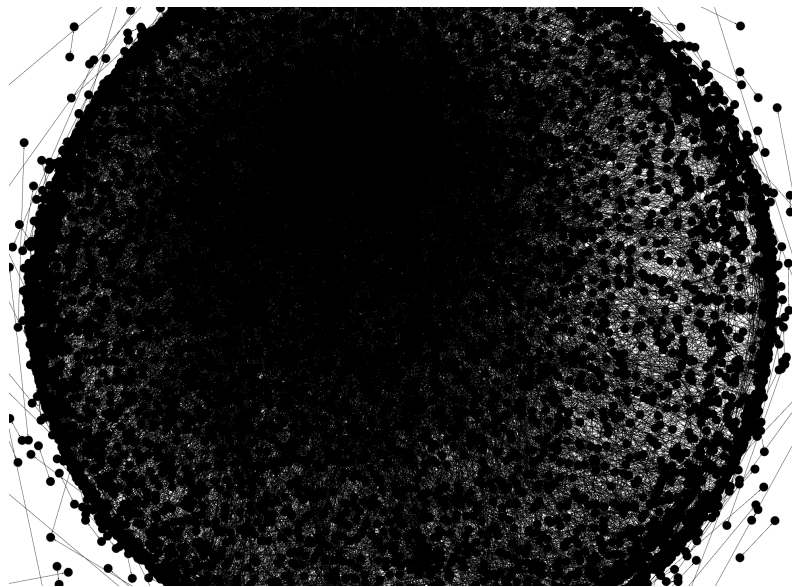
Pic. 1 — Visualization of graph before processing



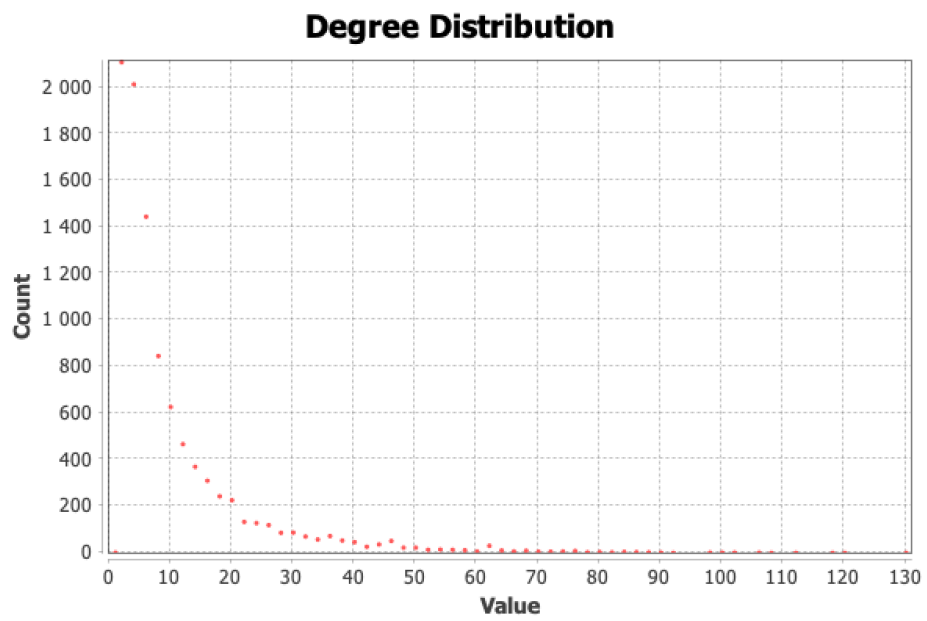
Pic. 2 — Yifan Hu layout



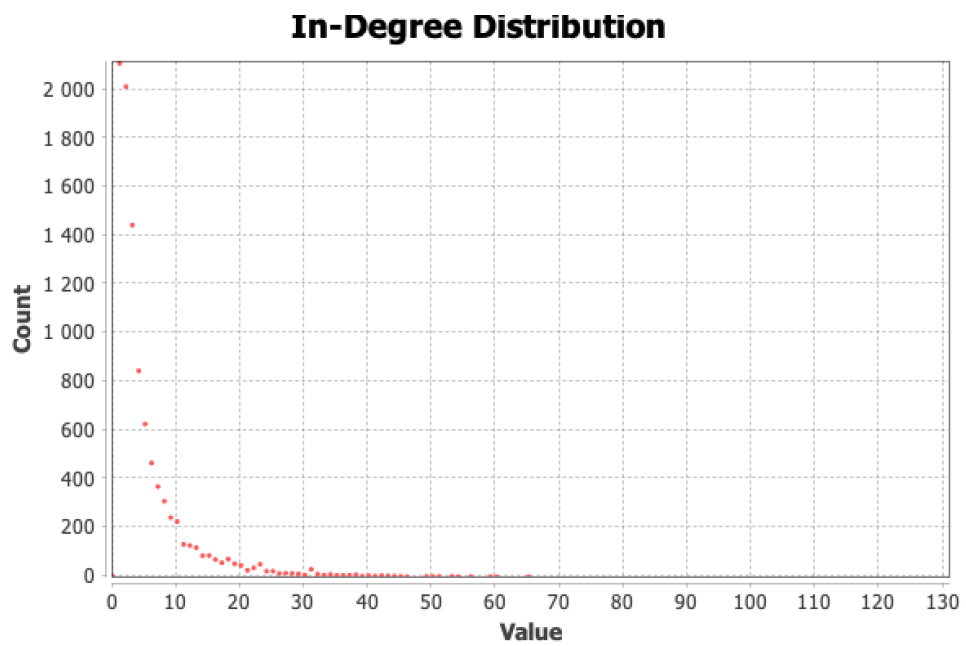
Pic. 3 — Force Atlas layout



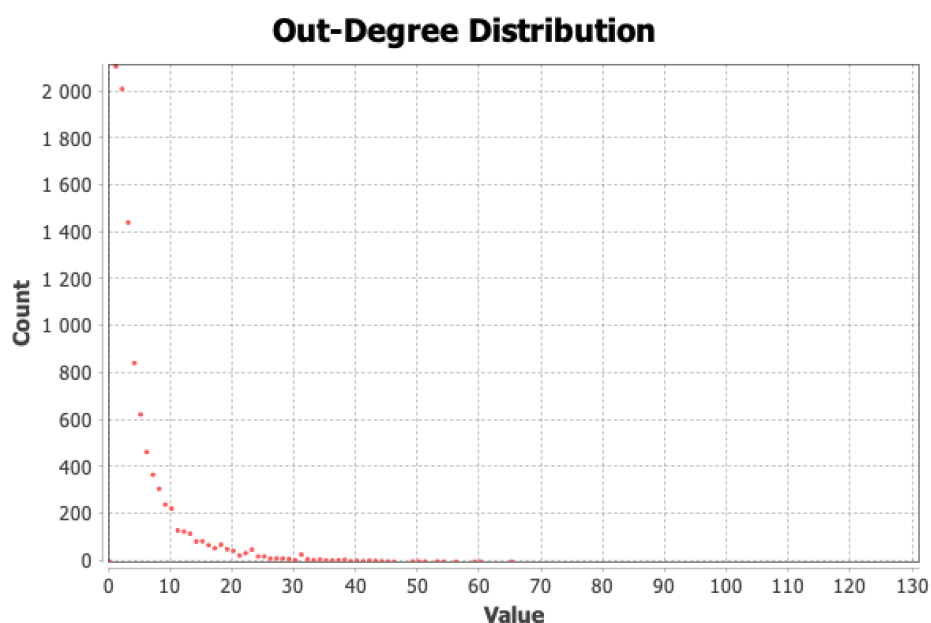
Pic. 4 — Fruchterman Reingold layout



Pic. 5 — Degree Report (Degree Distribution)



Pic. 6 — Degree Report (In-Degree Distribution)



Pic. 7 — Degree Report (Out-Degree Distribution)

Available network measures in Statistics provided by Gephi (Table 1):

Table 1 — Network measures

Average Degree	5,261
Average Path length	5.945216398512966
Density	0,001
Description Length	195920,915
Number of Communities	943

## Conclusions

As a result of the work, an acquaintance was made with Delphi software tools for network analysis. As a task, the analysis of the Arxiv collaboration network on the theory of high energy physics was carried out. Visualization and analysis of network data made it possible to find out that cooperation has close ties. However, there are also small, independent groups of cooperation networks. In addition, the calculated statistical parameters, such as average degree, average path length, density, description length, number of communities.

## Questions for self-monitoring from Russian educational materials

1. What is the density of any complete graph?

For **undirected** simple graphs, the graph density is defined as  $D = 2|E|/|V|(|V| - 1)$

While for **directed** simple graphs, the graph density is defined as  $D = |E|/|V|(|V| - 1)$ ,

In this case, the maximum number of edges is  $|V|(|V| - 1)/2$ .

2. Define the degree of the vertex, radius, diameter, average path and eccentricity of the graph.
  - The diameter of a connected graph is the maximum possible distance between its two vertices.
  - The center of a graph is a vertex such that the maximum distance between it and any other vertex is the smallest of all possible; this distance is called the radius of the graph.
  - A path in a graph is a sequence of vertices in which each vertex is connected to the next by an edge.
  - The eccentricity of a graph vertex is the distance to the vertex that is as far away from it as possible. For a graph for which the weight of its edges is not defined, the distance is defined as the number of edges.