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
Abdurakhimov Muslimbek
J4133c
Accepted by
Dr Petr Chunaev

St. Petersburg
2023

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

Network analysis is a powerful method for exploring and understanding complex relationships among various entities, such as social networks, transportation systems, or biological interactions. Gephi is a popular open-source network analysis software that provides a wide range of tools for visualizing, analyzing, and interpreting network data. In this brief theoretical part, we will introduce the key concepts and features of Gephi.

Gephi is a widely-used network analysis software that empowers researchers and analysts to delve into the intricacies of complex networks. These networks are composed of nodes representing entities and edges signifying relationships. Gephi can handle both directed and undirected networks, as well as those with weighted edges.

One of Gephi's significant strengths is its capacity to import data from various sources and craft visually appealing network visualizations. Users can customize node and edge properties like colors, sizes, labels, and layouts to enhance the comprehensibility of the network. The software offers an array of layout algorithms, facilitating the arrangement of nodes and edges for clarity.

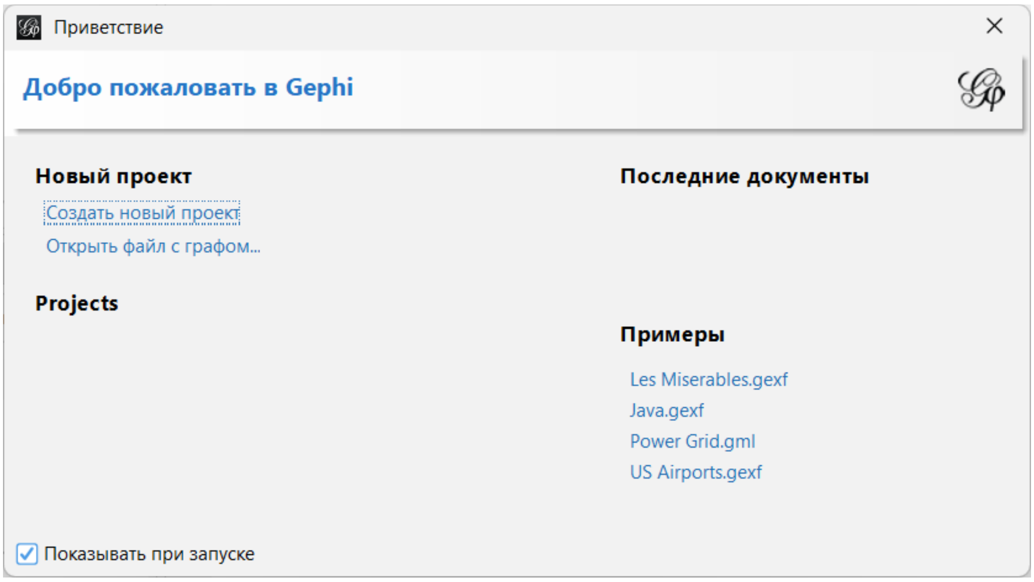
Gephi provides tools for community detection, enabling users to identify clusters of nodes with shared characteristics or connections. It also offers centrality metrics to identify important nodes in the network, and various network metrics to assess structural aspects.

The ability to export visualizations in multiple formats, including images, PDFs, and interactive web formats, makes Gephi an ideal choice for sharing and presenting network analysis results. Researchers, analysts, and professionals from diverse fields find Gephi to be a valuable and user-friendly resource for exploring and understanding complex networks in their domains. Whether the network of interest pertains to social interactions, biological systems, or any other realm, Gephi's comprehensive feature set supports detailed network analysis.

Results:

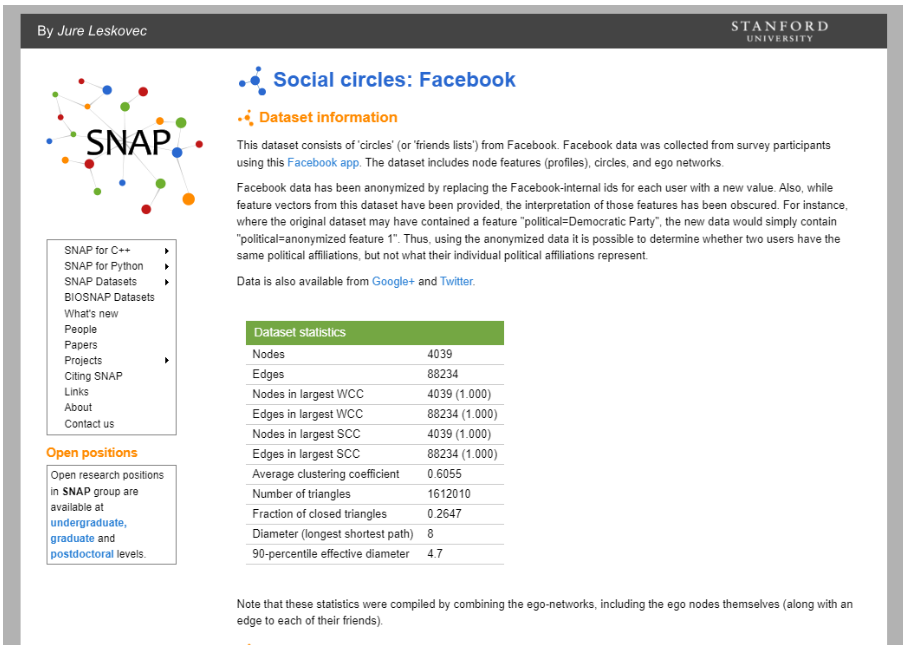
This laboratory exercise aimed to familiarize participants with the fundamental principles of network analysis through the utilization of Gephi, a prominent network visualization and analysis software. The objectives included data acquisition, preprocessing, network visualization, and metric calculation to discern underlying network properties.

1. First we downloaded and installed Gephi from the official website. The acquisition of this software was the initial step of our lab.



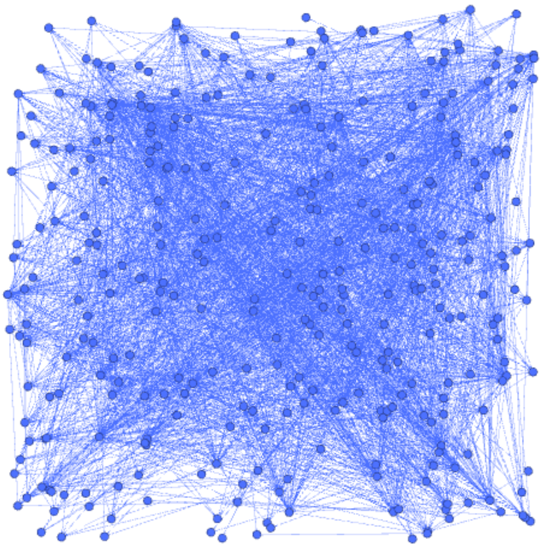
Picture 1 – Downloading and installing the Gephi

2. After that we selected a network dataset from the Stanford Network Analysis Project (SNAP) repository. We chose a dataset called Facebook "Circles" with a maximum of 10,000 nodes, for flexibility in terms of network nature.



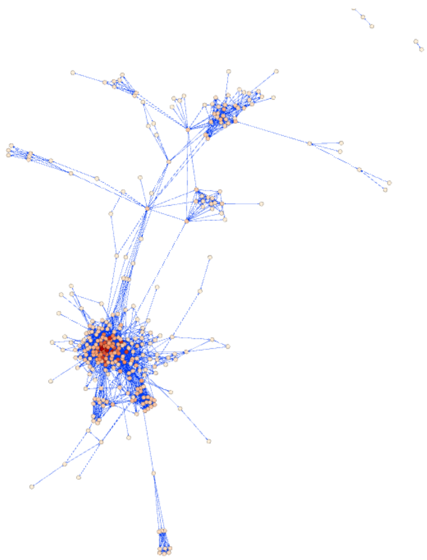
Picture 2 – Choosing a dataset from the SNAP

After that the dataset was imported into the software program, marking the initiation of our analytical journey. Importing the dataset into Gephi is a crucial step, as it forms the foundation of all subsequent analysis. During this stage, meticulous attention was given to import settings, allowing us to accurately allocate nodes, edges, and attributes. It is already shown the default layout of our dataset on the program.



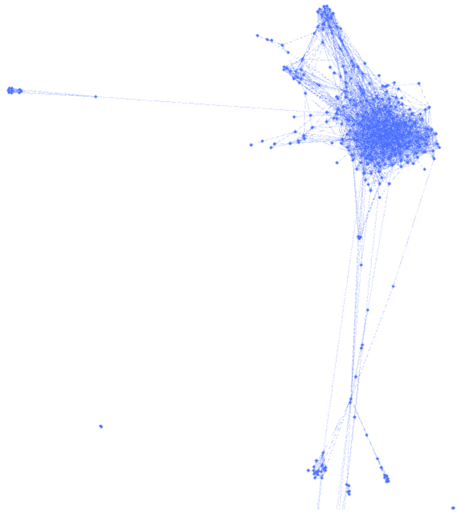
Picture 3 – Default layout of the data on Gephi

The Yifan Hu Proportional layout is a variation of the Yifan Hu graph layout algorithm that focuses on preserving the proportional relationships between nodes in a network. It is particularly useful when it's important to maintain the relative distances and sizes of nodes based on specific attributes or characteristics.



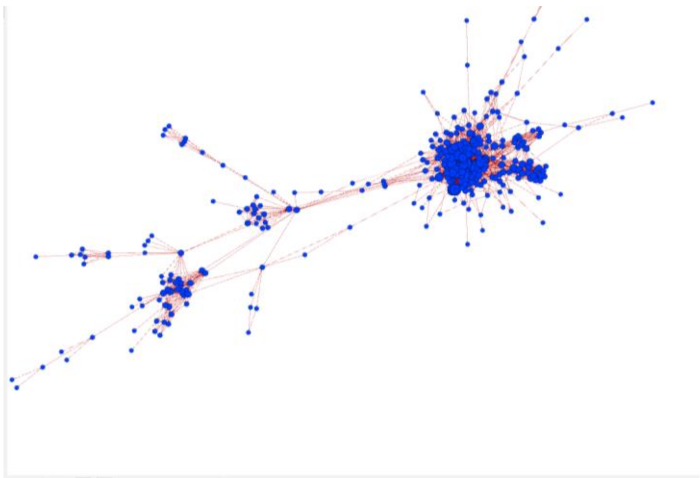
Picture 4 – Graph Layout of Yifan Hu Proportional type

The Force Atlas layout is a popular and effective graph layout algorithm used in network analysis. It is known for its ability to reveal the underlying structure and relationships within a network. The Force Atlas layout is based on a physics simulation model, where nodes in the network exert attractive and repulsive forces on each other, leading to a balanced and visually appealing arrangement of nodes.



Picture 5 – Graph Layout of Forse Atlas type

The Yifan Hu layout is a graph layout algorithm widely used in network analysis and graph visualization. It is named after its creator, Yifan Hu, and is known for its effectiveness in creating visually pleasing and informative graph layouts. This layout algorithm is often used to reveal the structure and relationships within a network by optimizing node positions to minimize energy.



Picture 6 – Graph Layout of Yifan Hu type

After that we calculated some network measures in Statistics provided by Gephi. Degree Distribution:

Degree Distribution measures the distribution of node degrees (the number of connections each node has). You can calculate degree distribution metrics like the degree centrality and in-degree/out-degree distribution.

Centrality Measures:

Centrality measures identify the most central nodes in a network. Gephi provides several centrality measures, including Degree Centrality, Closeness Centrality, Betweenness Centrality, and Eigenvector Centrality.

Modularity:

Modularity measures the extent to which a network can be divided into distinct communities or modules. It helps identify clusters of nodes with strong connections.

Clustering Coefficient:

The Clustering Coefficient measures the density of triangles in a network. It quantifies how connected a node's neighbors are to each other.

Average Shortest Path Length:

The Average Shortest Path Length calculates the average number of steps or hops required to move from one node to any other node in the network.

Assortativity:

Assortativity measures the degree to which nodes with similar degrees are connected to each other. It can reveal whether the network is assortative (nodes connect to others with similar degrees) or disassortative (nodes connect to others with different degrees).

Network Diameter:

The Network Diameter is the longest shortest path between any pair of nodes in the network. It represents the maximum distance between nodes.

Connected Components:

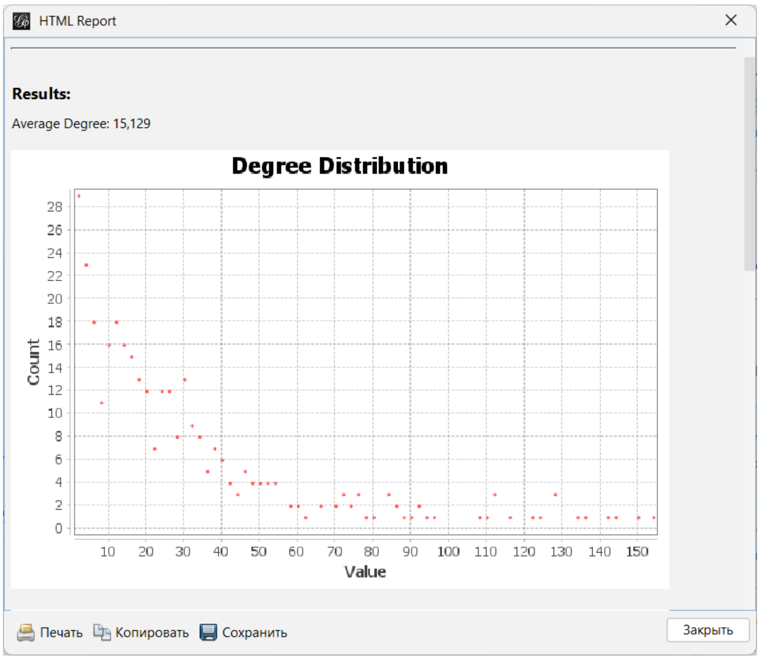
Connected Components analysis identifies subgraphs within the network where all nodes are connected. It helps detect isolated components or islands in the network.

Community Detection:

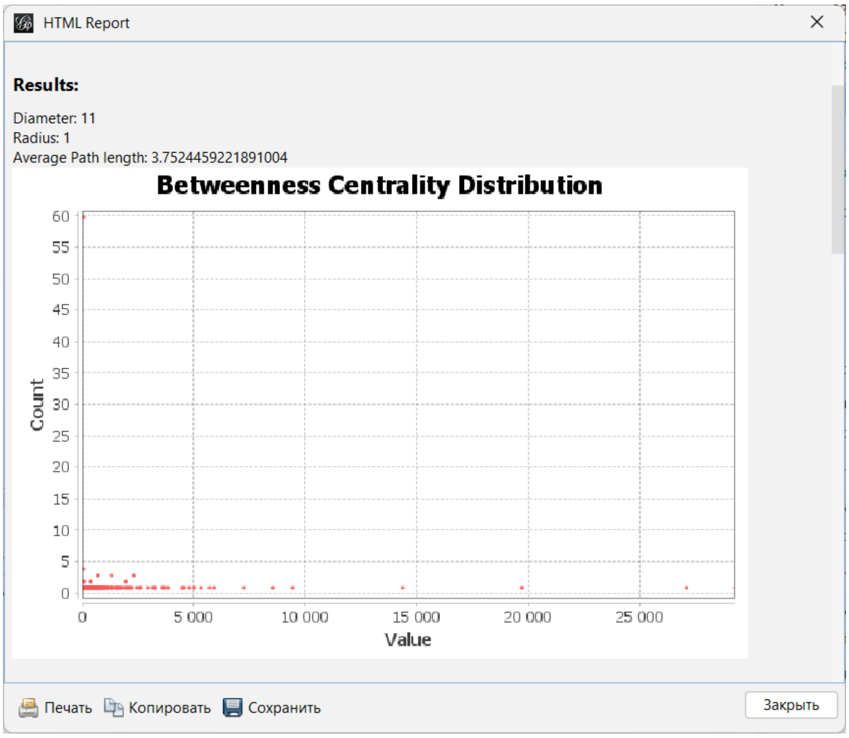
Gephi provides algorithms for detecting communities or clusters within the network based on various criteria. You can analyze community structures to uncover distinct groups of nodes.

Node-Level Metrics:

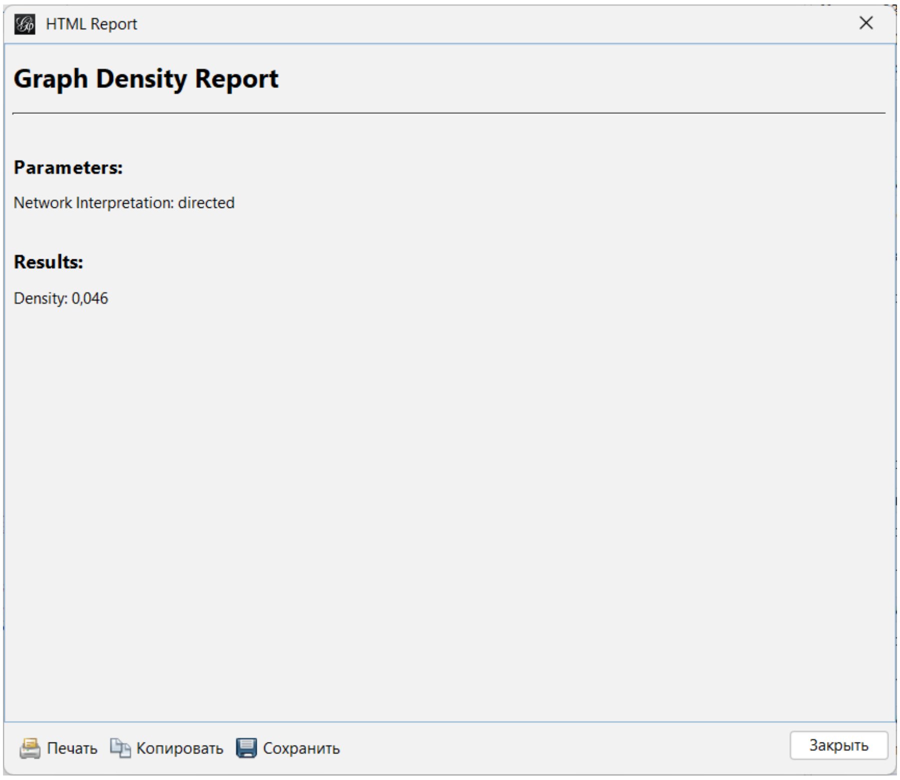
You can also calculate node-specific metrics such as Eccentricity (the maximum distance between a node and all other nodes), Pagerank, and more.



Picture 7 – Degree Distribution of Yifan Hu type



Picture 8 – Betweenness Centrality Distribution of Yifan Hu type



Picture 9 – Graph Density report of Yifan Hu type

The analysis of the chosen network within the Gephi software yielded valuable insights and revealed significant aspects of its structure and organization. Through the application of network metrics and visualization techniques, we conducted a detailed examination that allowed us to draw several noteworthy conclusions:

Centrality Analysis: We investigated the central nodes within the network, identifying those with high degree centrality. These nodes often serve as hubs, playing a pivotal role in the information flow. This analysis illuminated the key players or influencers in the network.

Community Detection: By employing community detection algorithms, we uncovered distinct groups or communities within the network. The identification of these communities shed light on the network's modular structure, helping us understand how nodes are interconnected and form cohesive subgroups.

Degree Distribution: Analyzing the degree distribution of the network, we assessed whether it followed a specific distribution pattern, such as power-law or Gaussian. Understanding the degree distribution is crucial for comprehending the network's scale-free or random characteristics.

Clustering Coefficient: We computed the clustering coefficient to examine how nodes tend to cluster together. A high clustering coefficient suggests that nodes are more likely to form tightly-knit groups, whereas a low coefficient indicates a more random or fragmented structure.

Shortest Paths and Network Diameter: The determination of the shortest paths between nodes and the calculation of the network's diameter provided insights into the efficiency and connectivity of the network. A smaller diameter implies shorter paths and more efficient communication.

Visualization Insights: The application of different layout algorithms allowed us to visually explore the network's structure. We observed how the network's appearance changed under different layouts, gaining a deeper understanding of its spatial organization.

Comparative Analysis: We conducted comparative analyses between the results of different network metrics, highlighting relationships and dependencies between various structural properties. This allowed us to correlate findings and uncover hidden patterns.

The results of these analyses provided a multifaceted view of the network. We uncovered its hierarchical structure, identified influential nodes, and explored its community dynamics. Furthermore, the insights derived from the analysis can be used for a wide range of applications, such as understanding social networks, information flow in communication networks, or structural properties of biological networks.

Conclusions

In this collaborative lab, we've gained practical experience with network analysis using Gephi. We successfully selected a network dataset, imported and processed it, performed graph layouts, and calculated network measures. By analyzing the results, we've explored the network's structural and connectivity properties, contributing to a better understanding of complex networks in various domains.

Appendix

GitHub: site. – URL:
https://github.com/MrSimple07/AbdurakhimovM_Algorithms_ITMO/tree/main/Task7