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

Abizer Safdari

J4134c

Accepted by

Dr Petr Chunaev

St. Petersburg

2020

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 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

- Gephi is an open-source network analysis and visualization software package for all kinds of graphs and networks.

- **Gephi** is an interactive visualization and exploration platform for all kinds of networks and complex systems, dynamic and hierarchical graphs.

- Gephi is a tool for people that have to **explore and understand graphs**. Like Photoshop but for graphs, the user interacts with the representation, manipulate the structures, shapes and colors to reveal hidden properties. The goal is to help data analysts to make hypothesis, intuitively discover patterns, isolate structure singularities or faults during data sourcing. It is a complementary tool to traditional statistics, as visual thinking with interactive interfaces is now recognized to facilitate reasoning. This is a software for Exploratory Data Analysis, a paradigm appeared in the Visual Analytics field of research.

- Undirected graphs have edges that do not have a direction. The edges indicate a two-way relationship, in that each edge can be traversed in both directions

- Directed graphs have edges with direction. The edges indicate a one-way relationship, in that each edge can only be traversed in a single direction.

- The degree (or valency) of a vertex of a graph is the number of edges that are incident to the vertex, and in a multigraph, loops are counted twice.

- Eccentricity of the vertex is defined as the maximum distance of one vertex from anyother vertex. It can be thought of as how far a node is from the node most distant from it in the graph

- The diameter of graph is the maximum distance between the pair of vertices. It is the maximum eccentricity of any vertex in the graph.

- Average path length is a concept in network topology that is defined as the average number of steps along the shortest paths for all possible pairs of network nodes. It is a measure of the efficiency of information or mass transport on a network

- The density ρ of an graph is the ratio of number of edges and the number of possible edges.

- Modularity compares the number of edges within clusters in graph with the expected number of edges in a random graph regardless of clusters.
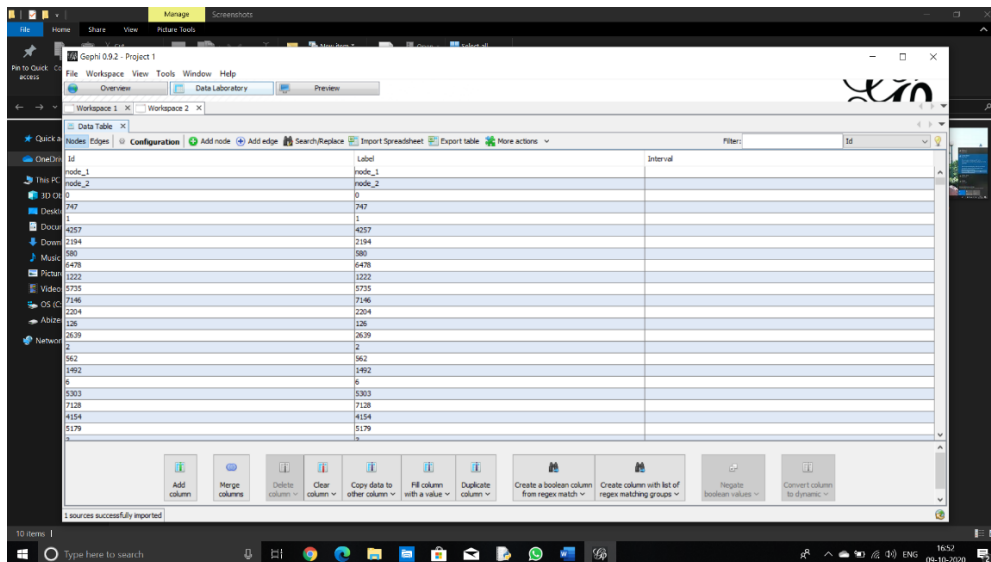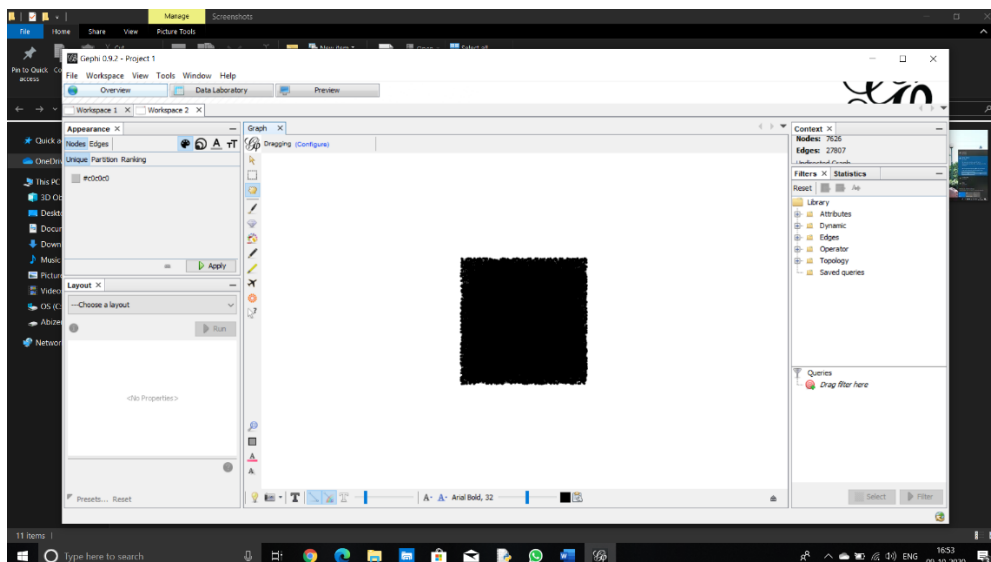
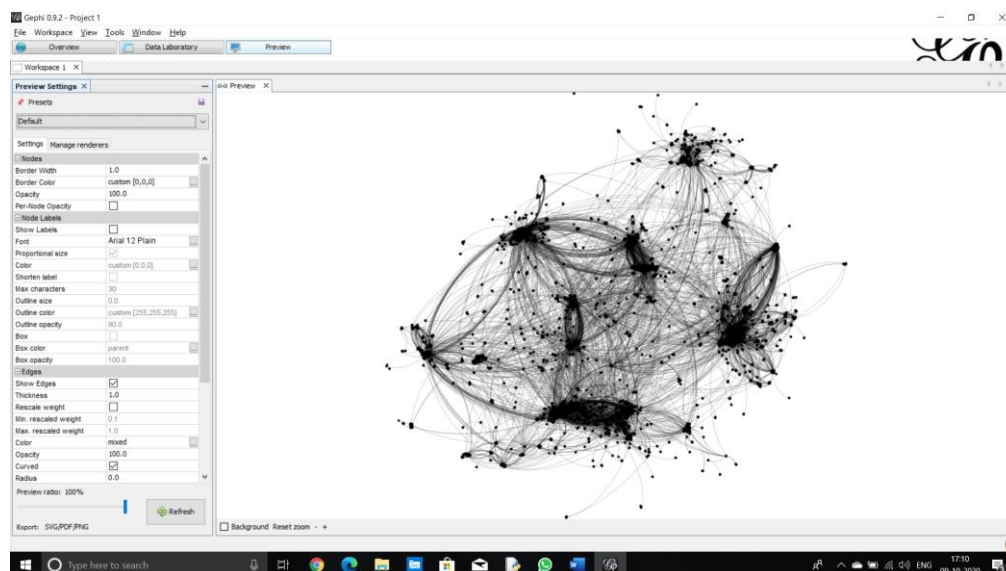# Implementation

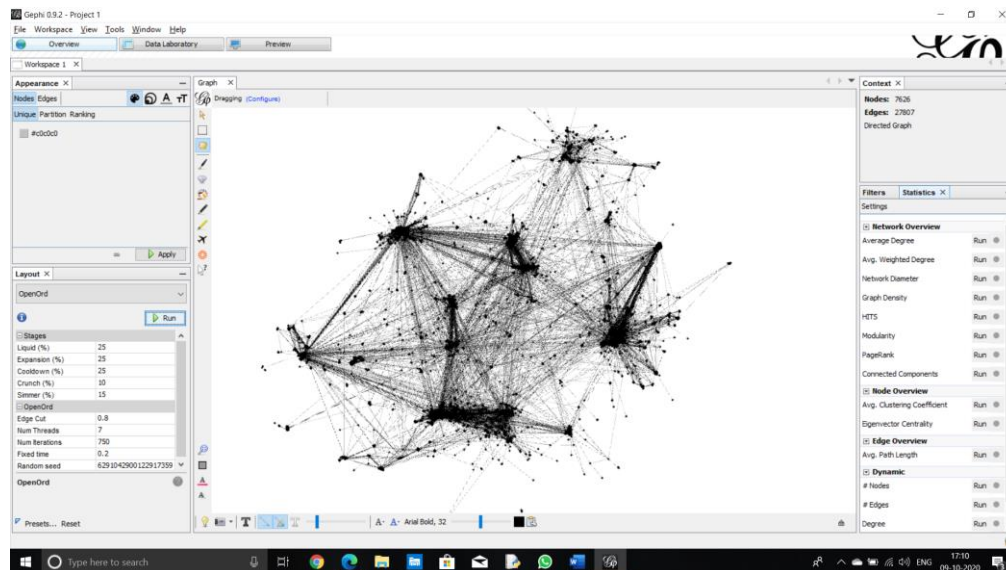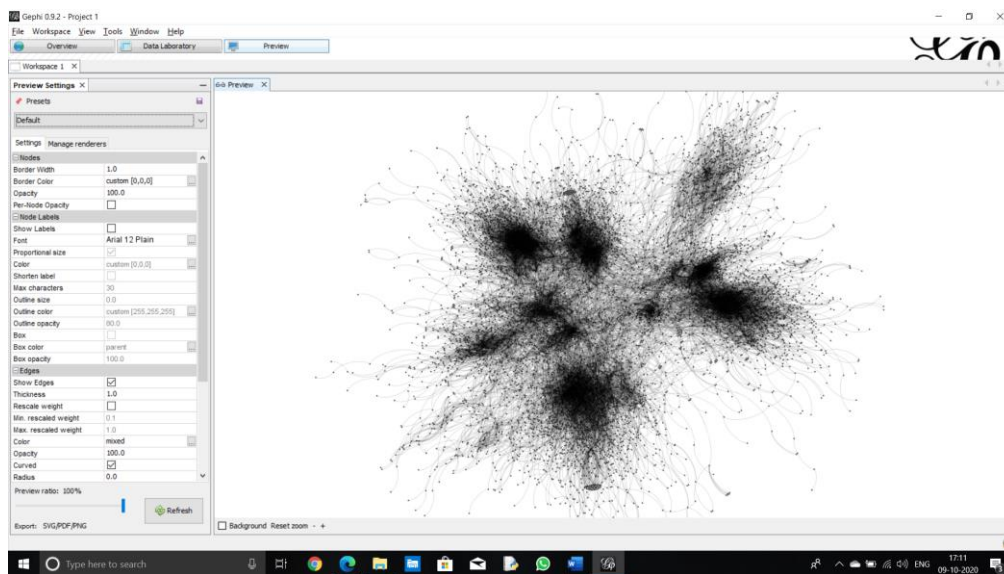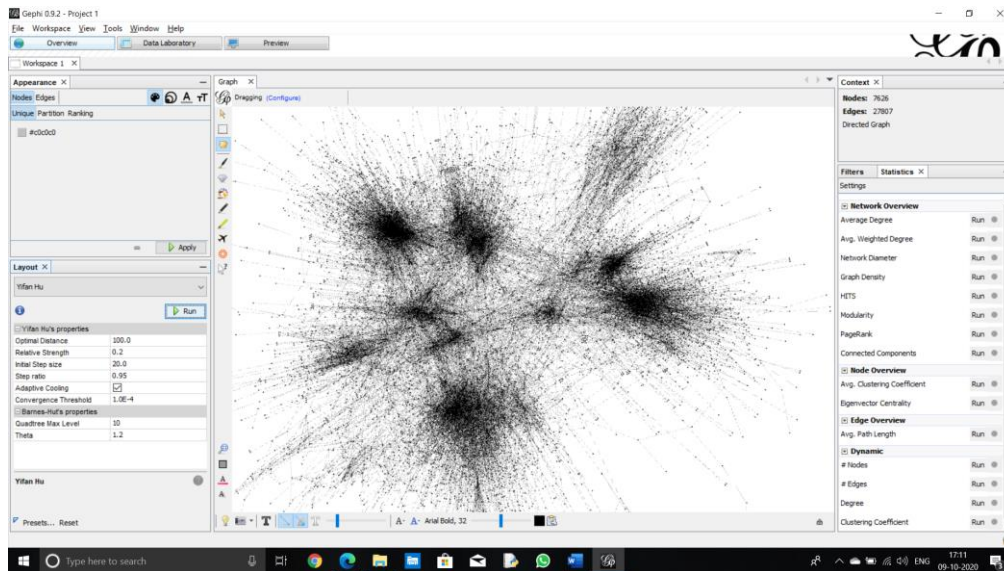## Import data
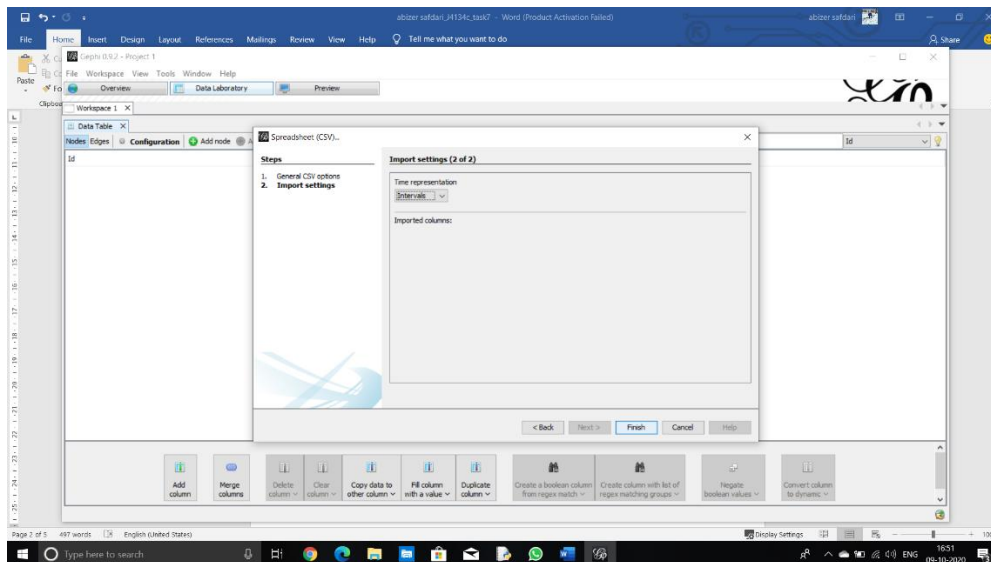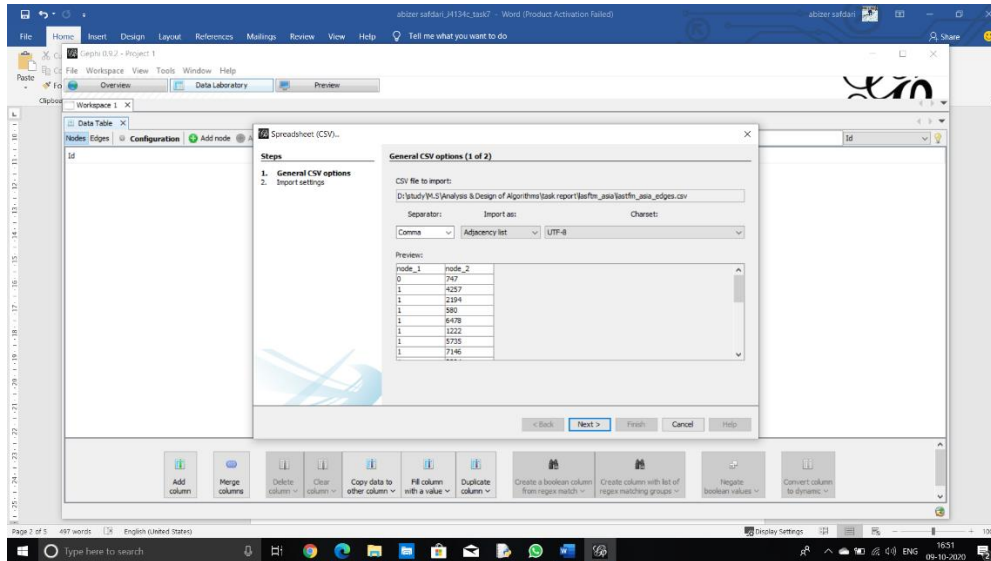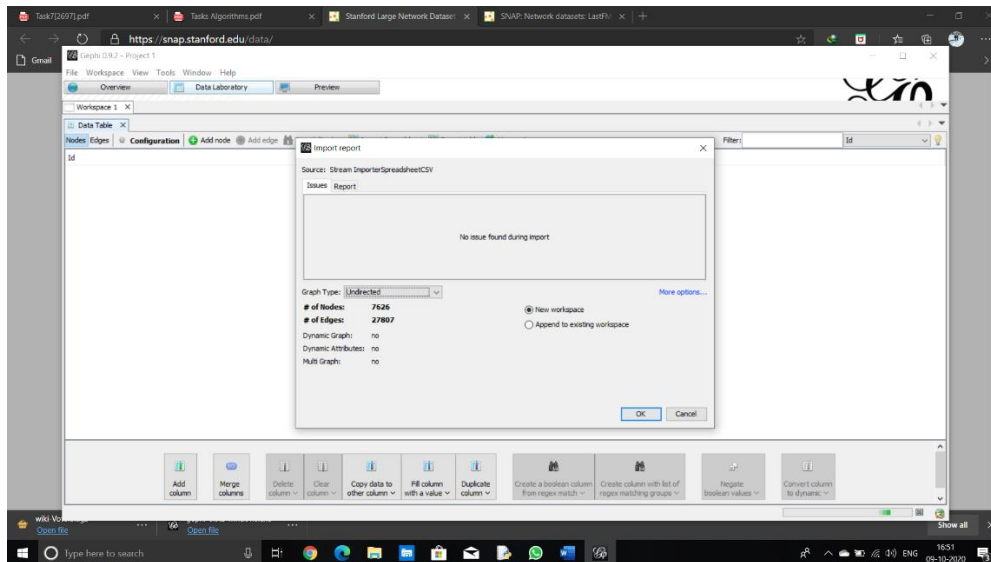
Imported graph

Result

This picture isn't informative and readable. Thus, we should use some clustering algorithm to see more information.

OpenOrd

yiFan Hu

Statistics

Nodes: 7626
Edges: 27807
Directed Graph

Filters	Statistics ×	

Settings

▼ Network Overview

Average Degree	3.646	Run ⑦
Avg. Weighted Degree	3.646	Run ⑦
Network Diameter	15	Run ⑦
Graph Density	0	Run ⑦
HITS		Run ⑦
Modularity	0.812	Run ⑦
PageRank		Run ⑦
Connected Components	2	Run ⑦

▼ Node Overview

Avg. Clustering Coefficient	0.11	Run ⑦
Eigenvector Centrality		Run ⑦

▼ Edge Overview

Avg. Path Length	5.232	Run ⑦

▼ Dynamic

Degree Report

Average Degree: 3.646

Degree Distribution
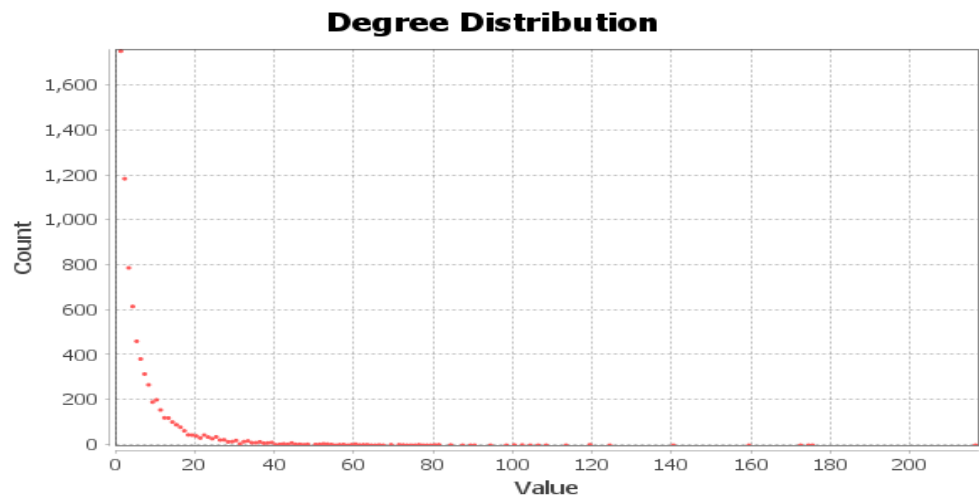
Graph Distance Report

Parameters:

Network Interpretation: directed

Results:

Diameter: 16
Radius: 0
Average Path length: 4.830866521205553

Betweenness Centrality Distribution
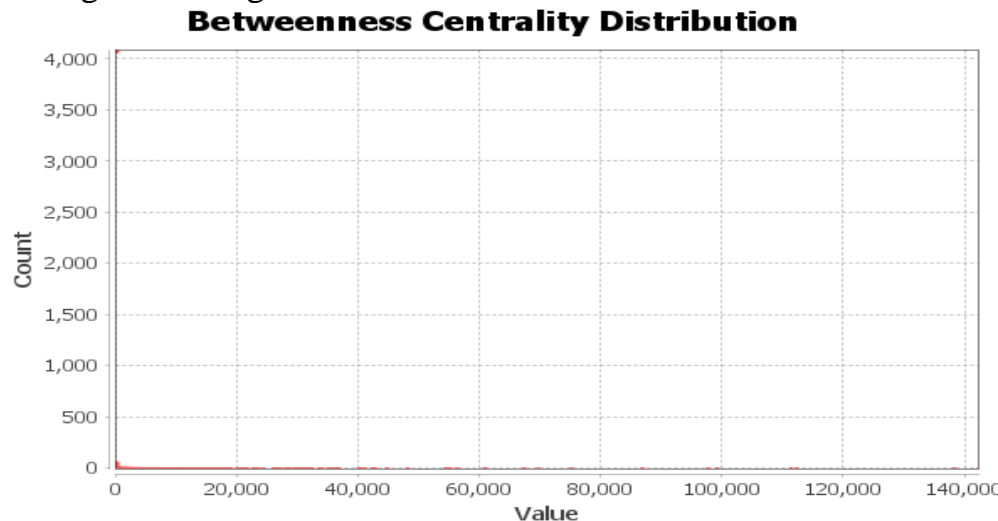
Connected Components Report

Results:

Number of Weakly Connected Components: 2
Number of Strongly Connected Components: 7626

Conclusions

- *In* social network of LastFM, most of the users of the nonnected with ech other
- We see that the graph has few weakly components and lots of stronger components, this is due to the fact that most people follow there friend, there a are very few who don't follow any one.
- The graph has small density (0.002), so, our graph is sparse. However, the diameter and average path length aren't big
- We can see that the average no. of people every person follow is around 3.646.
-

Appendix

Data: https://snap.stanford.edu/data/feather-lastfm-social.html

B. Rozemberczki and R. Sarkar. Characteristic Functions on Graphs: Birds of a Feather, from Statistical Descriptors to Parametric Models. 2020.