FEDERAL STATE AUTONOMOUS EDUCATIONAL INSTITUTION
OF HIGHER EDUCATION
ITMO UNIVERSITY

Report

on the practical task 7

“Algorithms on graphs. Tools for network analysis”

Performed by

Ilya Lyalinov

Academic group J4134c

Accepted by

Dr Petr Chunaev

St. Petersburg

2021

Goal

The use of the network analysis software Gephi

Problem formulation

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.

Results

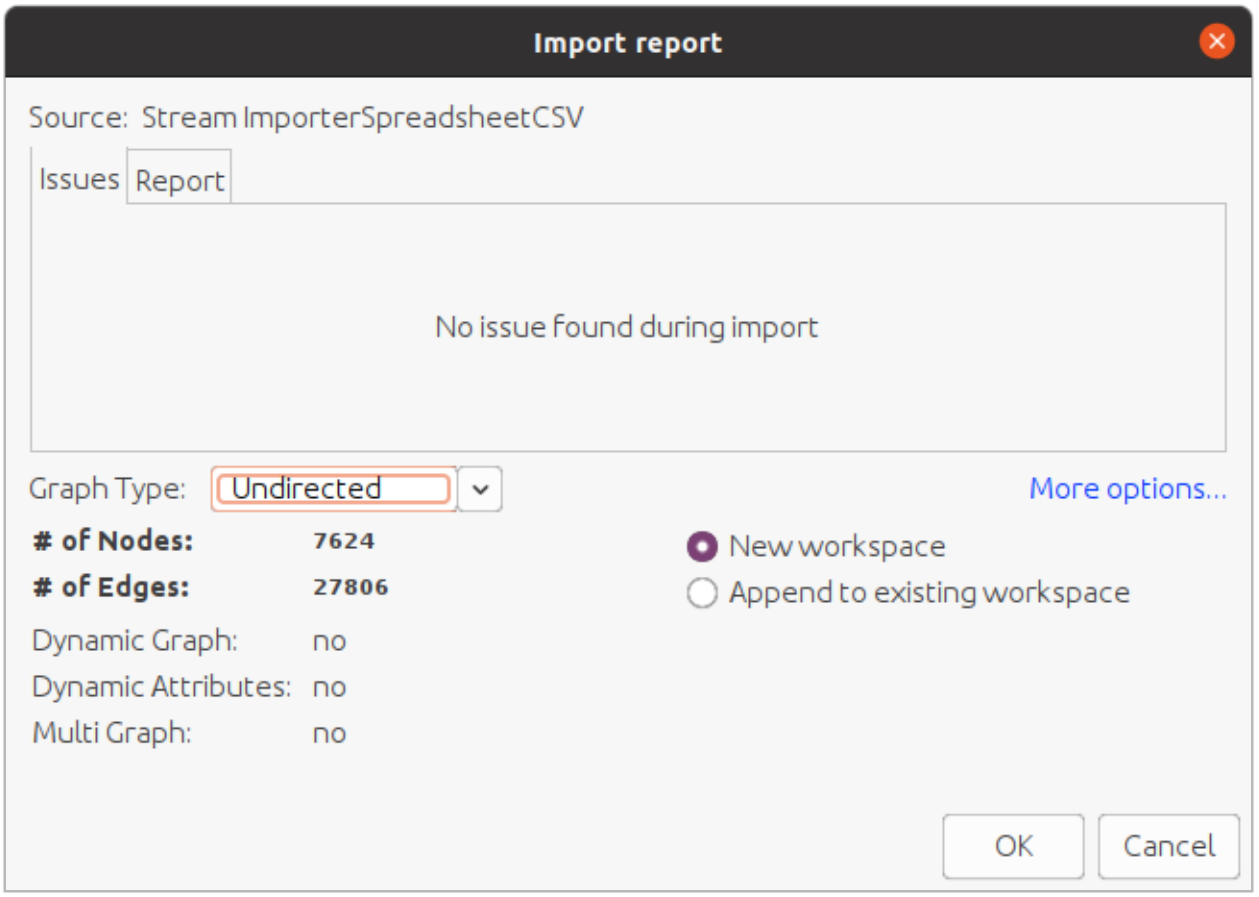
About the dataset:

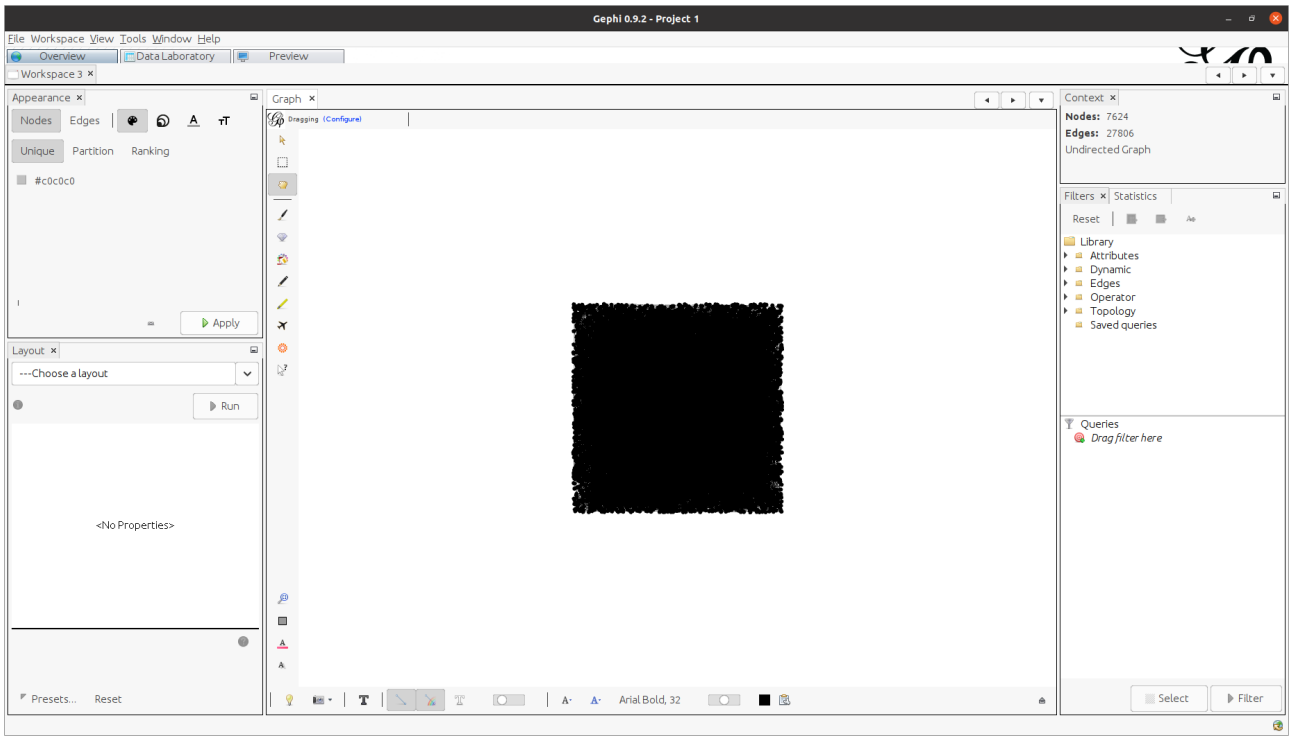
https://snap.stanford.edu/data/feather-lastfm-social.html
A social network of LastFM users which was collected from the public API in March 2020. Nodes are LastFM users from Asian countries and edges are mutual follower relationships between them. The vertex features are extracted based on the artists liked by the users. The task related to the graph is multinomial node classification - one has to predict the location of users. This target feature was derived from the country field for each user.

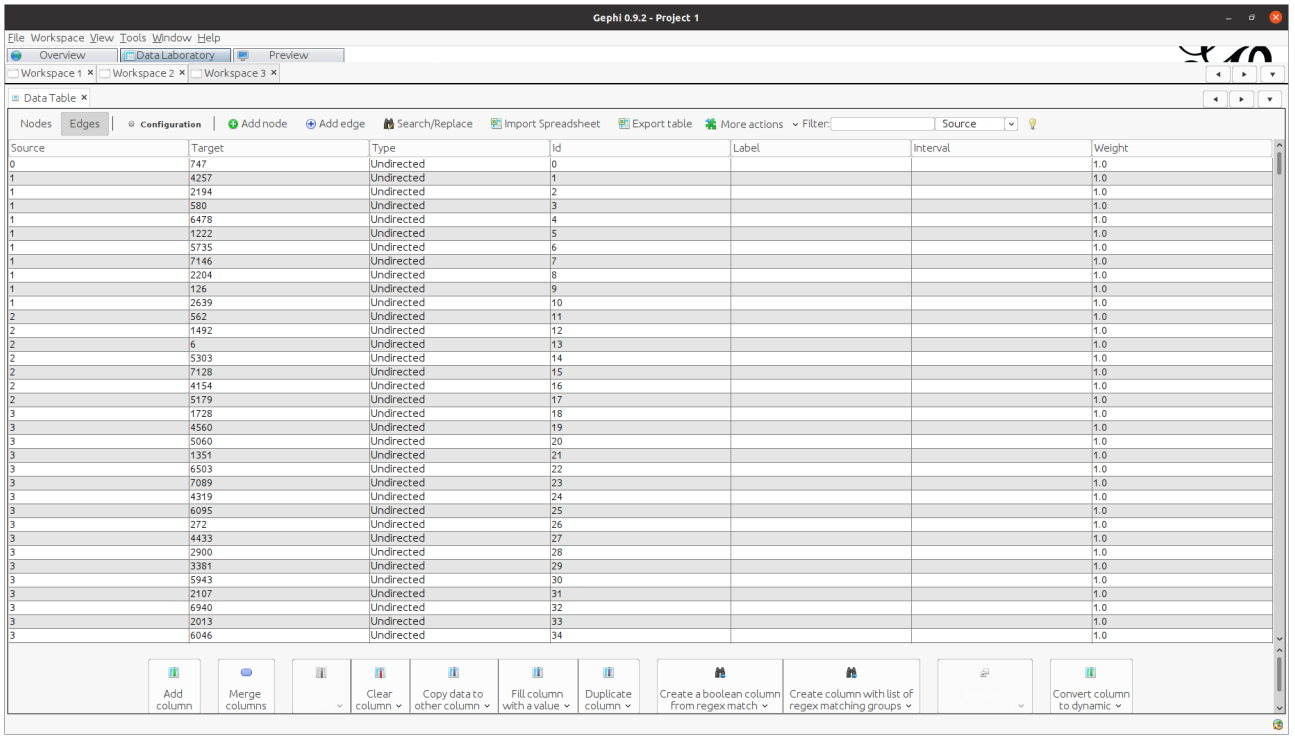
Statistics (evaluated by the dataset creators):

Nodes 7,624
Edges 27,806
Density 0.0009
Transitivity 0.1787

1) load the graph to gephi:

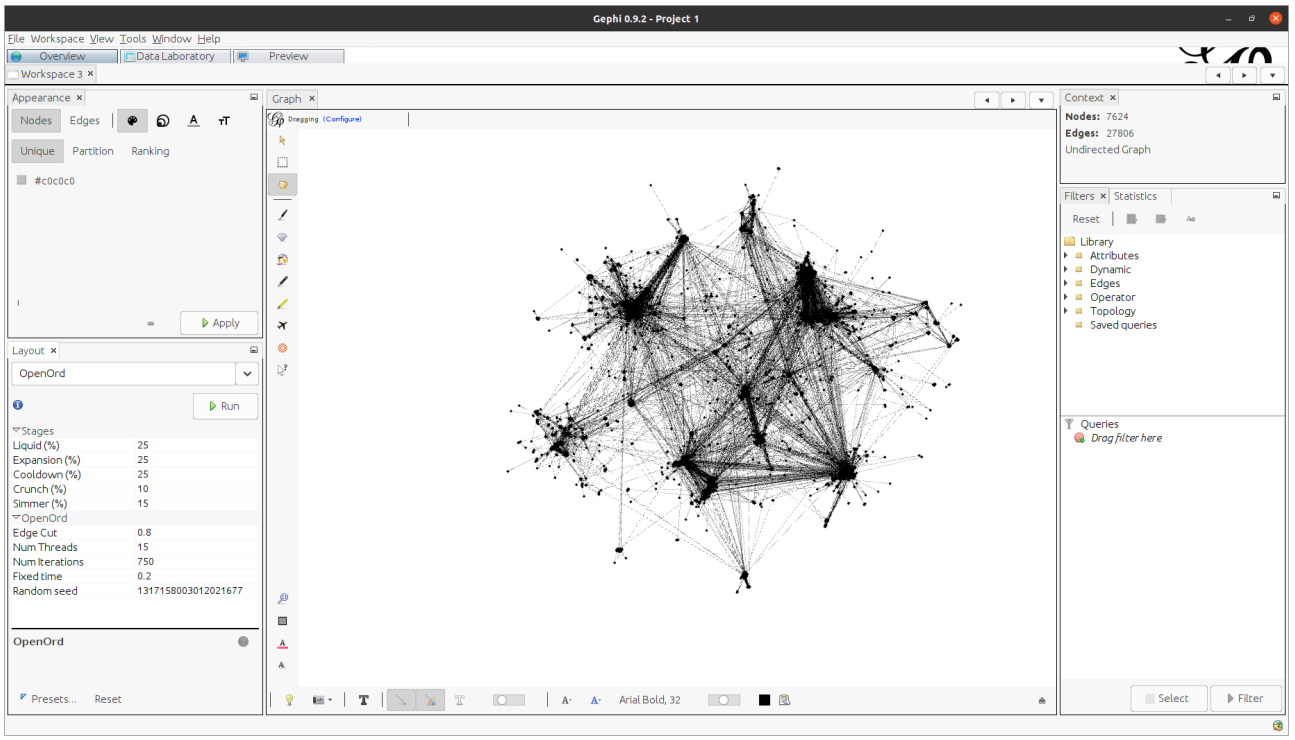


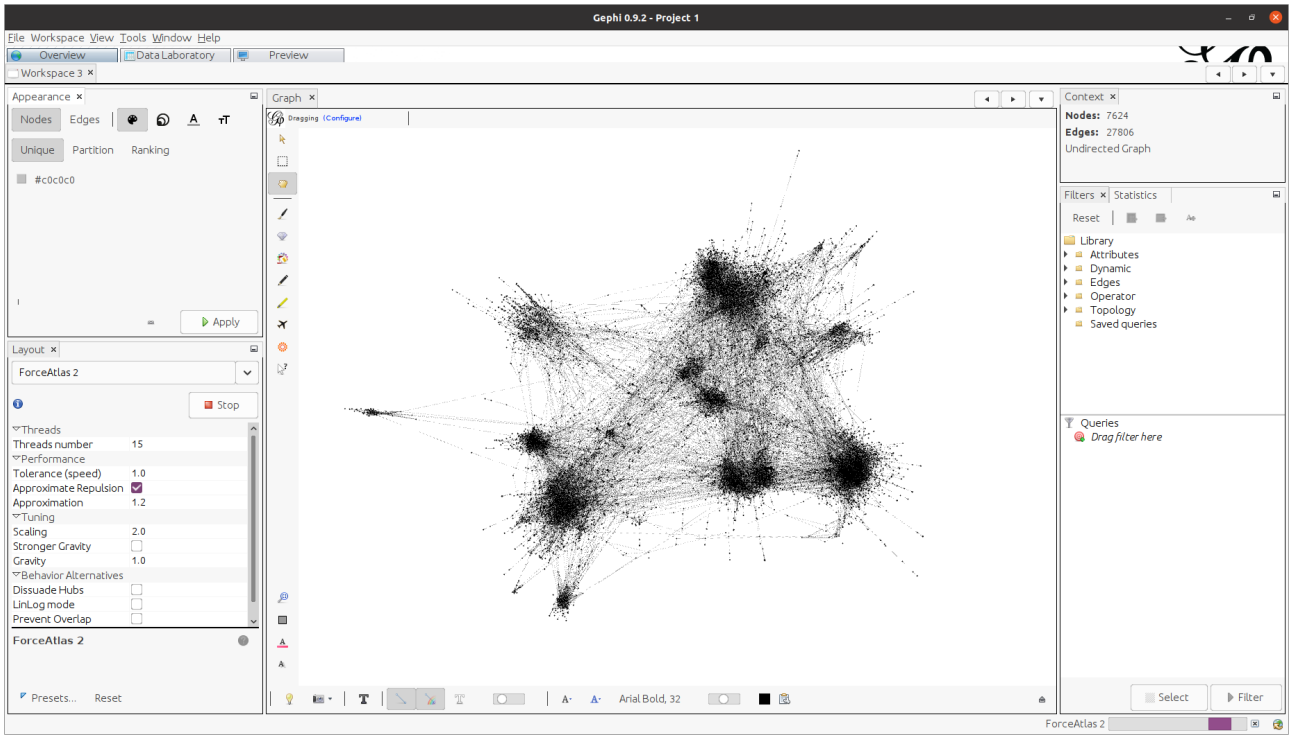




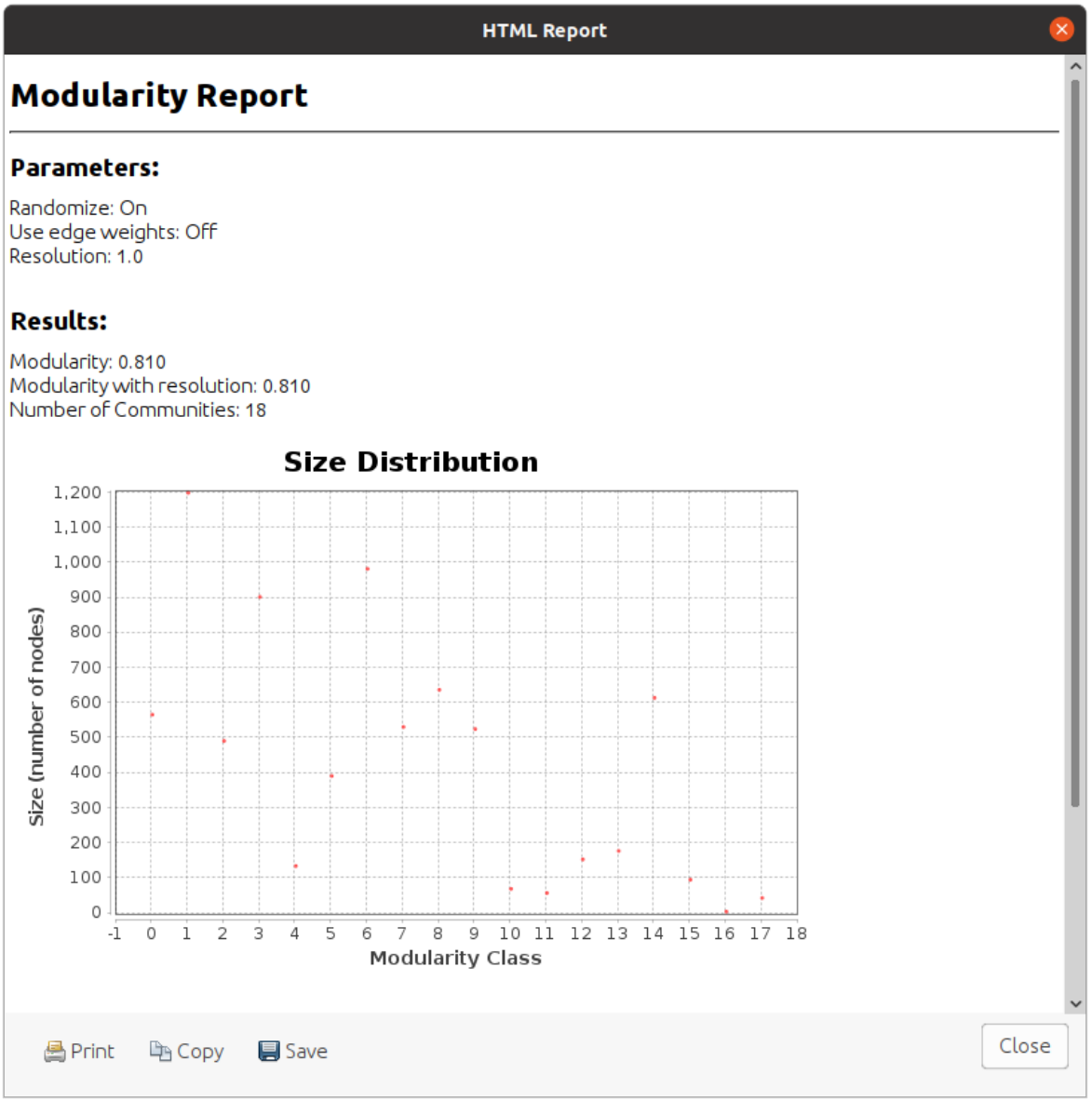
Source	Target	Type	Id	Label	Interval	Weight
0	747	Undirected	0			1.0
1	4257	Undirected	1			1.0
1	2194	Undirected	2			1.0
1	580	Undirected	3			1.0
1	6478	Undirected	4			1.0
1	1222	Undirected	5			1.0
1	5735	Undirected	6			1.0
1	7146	Undirected	7			1.0
1	2204	Undirected	8			1.0
1	126	Undirected	9			1.0
1	2639	Undirected	10			1.0
2	562	Undirected	11			1.0
2	1492	Undirected	12			1.0
2	6	Undirected	13			1.0
2	5303	Undirected	14			1.0
2	7128	Undirected	15			1.0
2	4154	Undirected	16			1.0
2	5179	Undirected	17			1.0
3	1728	Undirected	18			1.0
3	4560	Undirected	19			1.0
3	5060	Undirected	20			1.0
3	1351	Undirected	21			1.0
3	6503	Undirected	22			1.0
3	7089	Undirected	23			1.0
3	4319	Undirected	24			1.0
3	6095	Undirected	25			1.0
3	272	Undirected	26			1.0
3	4433	Undirected	27			1.0
3	2900	Undirected	28			1.0
3	3381	Undirected	29			1.0
3	5943	Undirected	30			1.0
3	2107	Undirected	31			1.0
3	6940	Undirected	32			1.0
3	2013	Undirected	33			1.0
3	6046	Undirected	34			1.0

2) 2 graph layouts:





3) different graph metrics:



18 clusters were detected.

HTML Report

Graph Density Report

Parameters:

Network Interpretation: undirected

Results:

Density: 0.001

Print Copy Save Close

$0.001 \approx 0$ => the graph is sparse

HTML Report

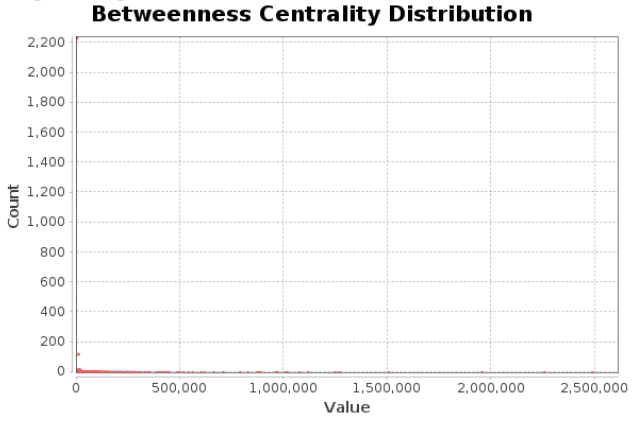
Graph Distance Report

Parameters:

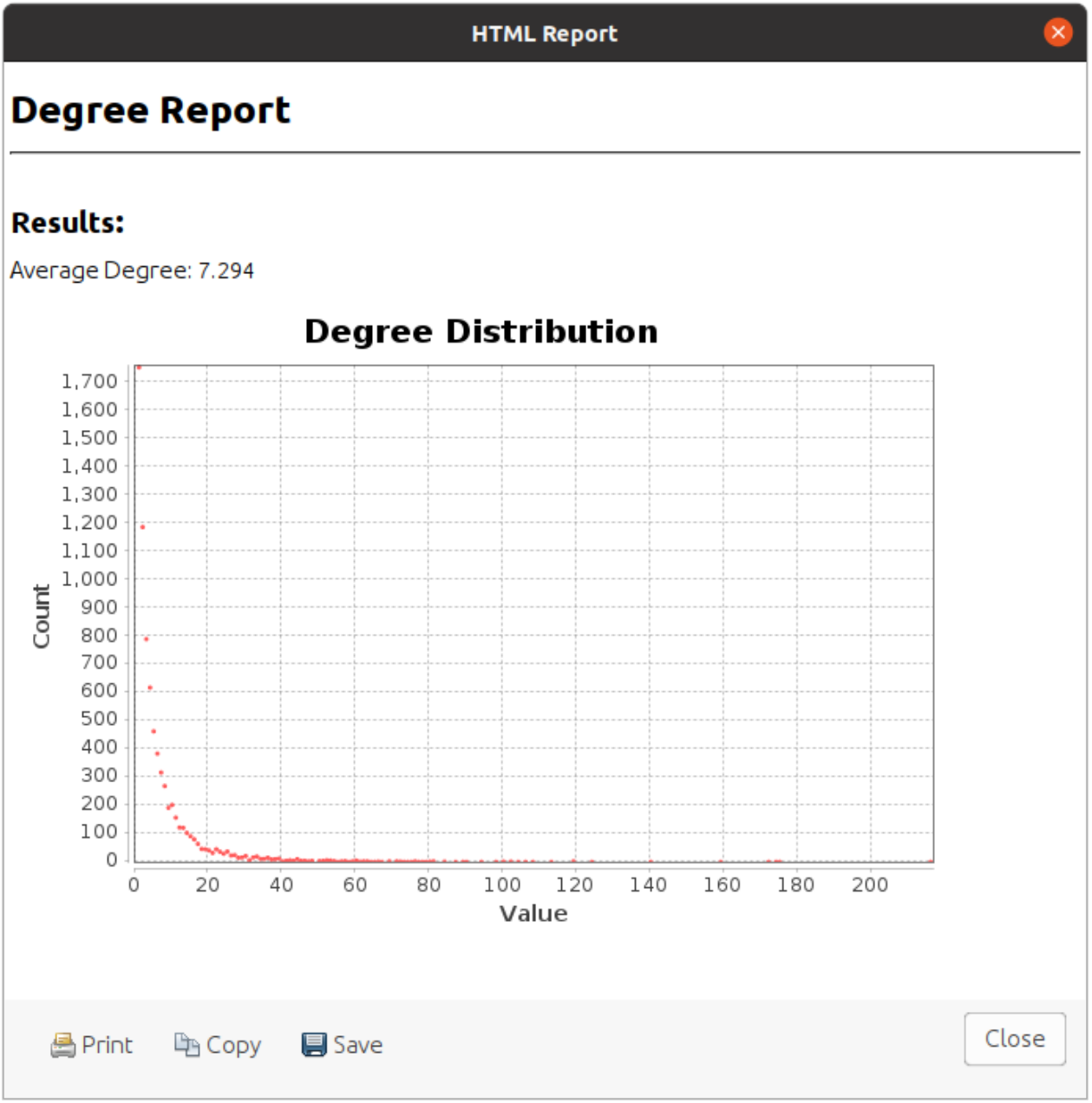
Network Interpretation: undirected

Results:

Diameter: 15
Radius: 8
Average Path length: 5.232237268915701



Print Copy Save Close



The graph contains only 1 connected component.

Conclusion

We have downloaded and installed gephi -- a tool for network analysis. We have chosen a network, successfully loaded it to gephi and checked the correctness of it. We have evaluated its modularity, distance, degree and density metrics in gephi. We have analyzed the results. We visualized the network using 2 different layouts.