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
Chumakov Mike
J4132c
Accepted by
Dr Petr Chunaev

St. Petersburg
2022

Goal

The use of the network analysis software Gephi

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.

Brief theoretical part

Let us introduce the characteristics of the degrees of the vertices of the unweighted graph:

1) *D(V), degree V, is the number of incoming and outgoing edges of the vertex V*
2) *D_{in}(V), approach half-degree v, is the number of incoming edges of the vertex V*
3) *D_{out}(V), outcome half-degree v, is the number of outgoing edges of the vertex V*
4) *D = $\sum_{k=0}^{n} \binom{n}{k} x^k a^{n-k} D = \frac{1}{|v|} \sum_{v \in V} D(v)$-average degree of vertices.*

For a weighted graph, similar values can be entered, with the only disagreement is that they don't get into account the no of (incoming/outgoing) edges, but the sum of the weights on the (incoming/outgoing) edges.

Now we introduce the characteristics of the graph in terms of distances:

1) *dist (v, u) is the distance (length of the shortest path) between v and u (G connected)*
2) *eccentricity ∈(v) = max u∈V dist (v, u) - the greatest distance between v and other vertices;*
3) *radius r = min v∈V ∈(v) - the smallest eccentricity at all vertices;*
4) *diameter D = max v∈V∈(v) - the greatest eccentricity for all vertices, i.e. the greatest distance between a pair of vertices;*
5) *average path length*

Further, we recall that the density ρ of a graph is the quotient | E | and numbers possible edges with the same | V |, i.e. the number of edges in a complete graph with | V | vertices:

$$\rho = \frac{2|E|}{|V|(|V| - 1)}$$

The over and many other quantities create it possible to characterize the graph different points of view. This is extremely ful when analysis of graphs simulating complex real-world networks. One of the convenient tools for such analysis, and also for visualization of graphs is Gephi software.

Results

To analyze the graph in the program Gephi, a network was chosen: «Bitcoin OTC trust weighted signed network». This is who-trusts-whom network of people who trade using Bitcoin on a platform called Bitcoin OTC. Since Bitcoin users are anonymous, there is a need to maintain a record of users' reputation to prevent transactions with fraudulent and risky users. Members of Bitcoin OTC rate other members in a scale of -10 (total distrust) to +10 (total trust) in steps of 1. This is the first explicit weighted signed directed network available for research.

Using the program Gephi, the graph decomposition was obtained a graph layout «ForceAtlas 2» (see Figure 1). With parameters: Treads number = 7, Tolerance = 1, Approximation 1.2, Scaling 2.0, Gravity = 1.0

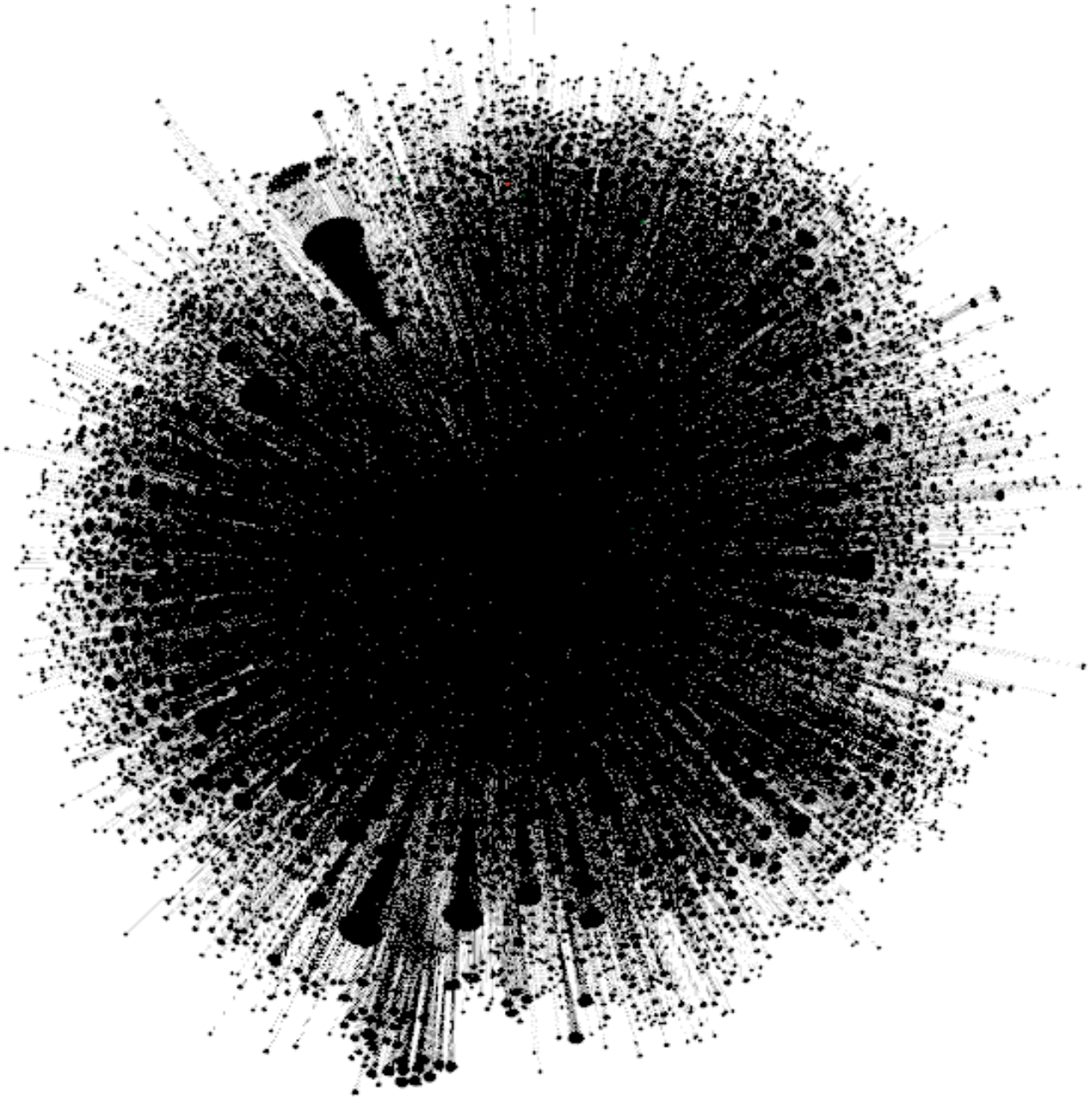

Figure 1: Graph layout «Force Atlas 2»

Using the program Gephi, the graph decomposition was obtained a graph layout « Yifan Hu» (see Figure 2). Yifan Hu's properties: Optimal Distance: 100, Relative Strength 0.2, Initial Step size = 20, Step ratio 0.95, Convergence Threshold = 1.0E-4.

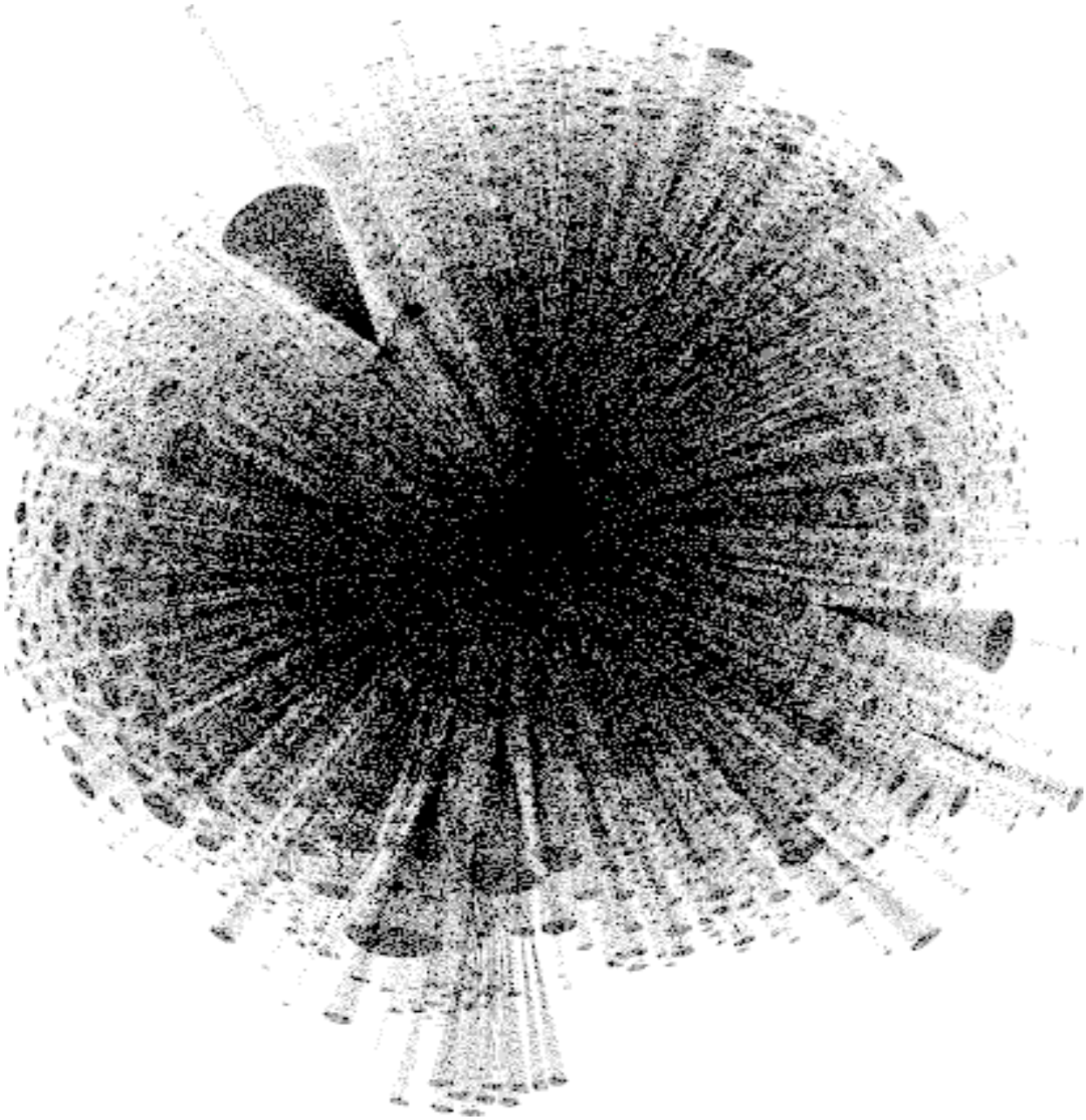

Figure 2: Graph layout «Yifan Hu »

The selected graph contains 5880 vertices and 35590 edges. Using the program Gephi, the following graph statistics were calculated: Average Degree = 1.95, Average Weighted Degree = 2.57, Network Diameter = 9, Modularity: 0.45, Modularity with resolution: 0 .45, Number of Communities: 41, Average Path Length: 3.7

Conclusions

In This laboratory used network analysis software. A graphic layout of a graph with 5880 vertices and 35590 edges of two different types «Force Athlas 2» and «Yifan Hu» was obtained. The basic statistics of the graph were calculated. In practice, we managed to make sure that it is a very powerful tool for network analysis and visualization.