

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

Rami Al-Naim

J4132c

Accepted by

Dr Petr Chunaev

St. Petersburg
2020

1 Goal

The use of the network analysis software Gephi.

2 Formulation of the problem

During this work Gephi package was used for visualizing and analyzing a network. The network itself was downloaded from [snap](#) database.

The chosen network should be visualized in two layouts and analyzed based on statistics calculated by Gephi.

3 Brief theoretical part

Gephi is an open-source network analysis and visualization software package written in Java on the NetBeans platform. The user interacts with the representation, manipulate the structures, shapes and colors to reveal hidden patterns. 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.

Basic graphs measures are the number of vertices $|V|$ and the number of edges $|E|$. There are some degree measures, such as:

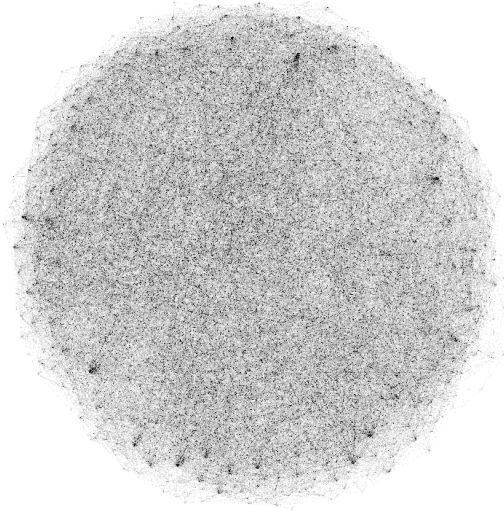
- $d(v)$ – degree of v , the number of edges for vertex v ;
- $d_{in}(v)$ – in-degree of v , the number of in-edges for vertex v ;
- $d_{out}(v)$ – out-degree of v , the number of out-edges for vertex v ;
- $d = \frac{1}{|V|} \sum d(v)$, $v \in V$ – average degree over all vertices.

4 Results

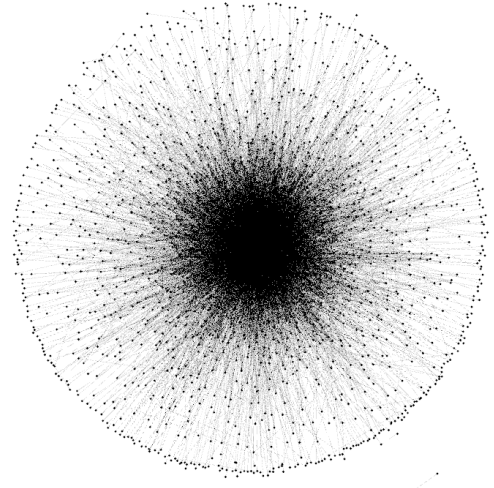
For this task the [CollegeMsg](#) directed dataset was used. It's edges (u, v, t) represent the private message sent from user u to user v at moment t .

The graph was loaded into Gephi and visualized using Force Atlas layout and Fruchterman Reingold layout (Fig. 1).

The distribution of degrees of the graph can be found in the Figure 2. The calculated statistics are present in the Table 1

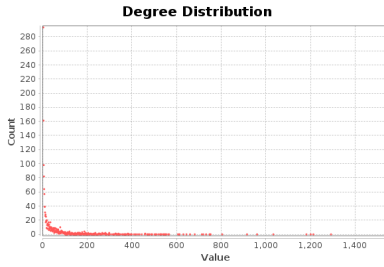


(a) Force Atlas layout

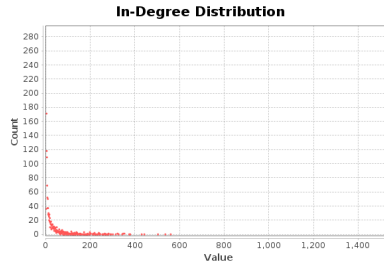


(b) Fruchterman Reingold layout

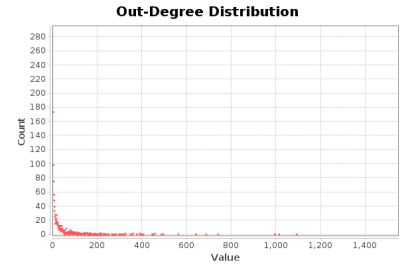
Figure 1: Graph in two layouts



(a) Degree of a graph



(b) In-degree of a graph



(c) Out-degree of the graph

Figure 2: Degree distributions of the graph

Statistic	Value
Average degree	31.509
Strong components	601
Weak components	4
Density	0.017
Average path length	3.1973

Table 1: Graph statistics

Based on the aforementioned values (Table 1) of the graph degree the network was visualized with respect to those stats (Fig. 3). The degree of the node defined the size and color of it (from smaller values to greater values from blue color to red) and the timestamp of the edge defined it's color (older edges to closer to blue, newer to red).

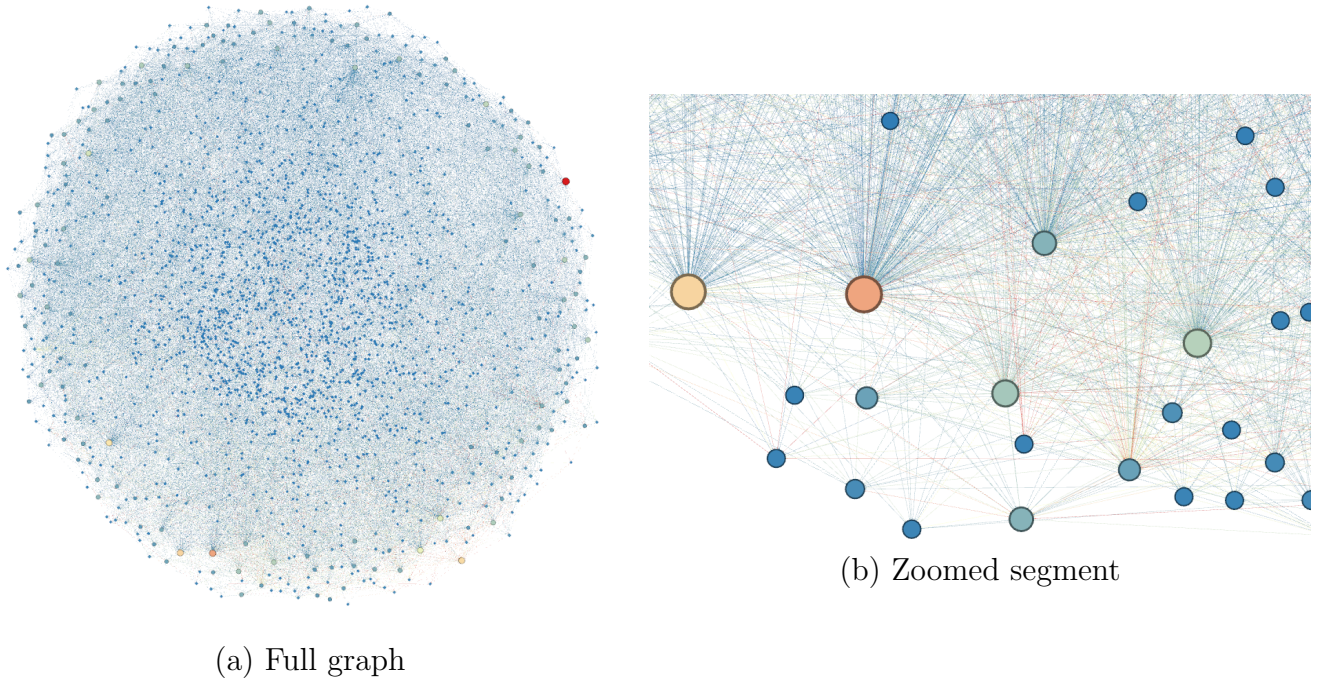


Figure 3: Visualized graph

5 Conclusions

During this work the network of messages on Facebook was analyzed with Gephi platform.

The basic statistics from Table 1 can help in analyzing the graph. For example, the average degree 31.509 indicates that most of the messages were send roughly inside groups of 30 people. Since the data was gathered from students, such hypothesis could be valid.

Appendix

Project file available here: [GitHub](#).