Report

on the practical task No. 7

Algorithms on graphs. Tools for network analysis

Performed by
Carlos Andres Daza Rachen

Academic group
Accepted by
Dr Petr Chunaev

St. Petersburg
2022

Goal

The use of the network analysis software Gephi

Formulation of the problem

Choose a dataset, process it in Gephi, obtain different graph layouts, calculate network measurements.

Brief theoretical part

Gephi is an open source software for graph and network analysis. It uses a 3D render engine to display large networks in real-time and to speed up the exploration. A flexible and multi-task architecture brings new possibilities to work with complex data sets and produce valuable visual results. It provides easy and broad access to network data and allows for spatializing, filtering, navigating, manipulating and clustering[1].

Social network analysis (SNA) is the process of investigating social structures through the use of networks and graph theory. It characterizes networked structures in terms of nodes (individual actors, people, or things within the network) and the ties, edges, or links (relationships or interactions) that connect them[2].

Implementation

Selected dataset: Gnutella peer-to-peer network, August 4 2002, A sequence of snapshots of the Gnutella peer-to-peer file sharing network from August 2002. There are total of 9 snapshots of Gnutella network collected in August 2002. Nodes represent hosts in the Gnutella network topology and edges represent connections between the Gnutella hosts.
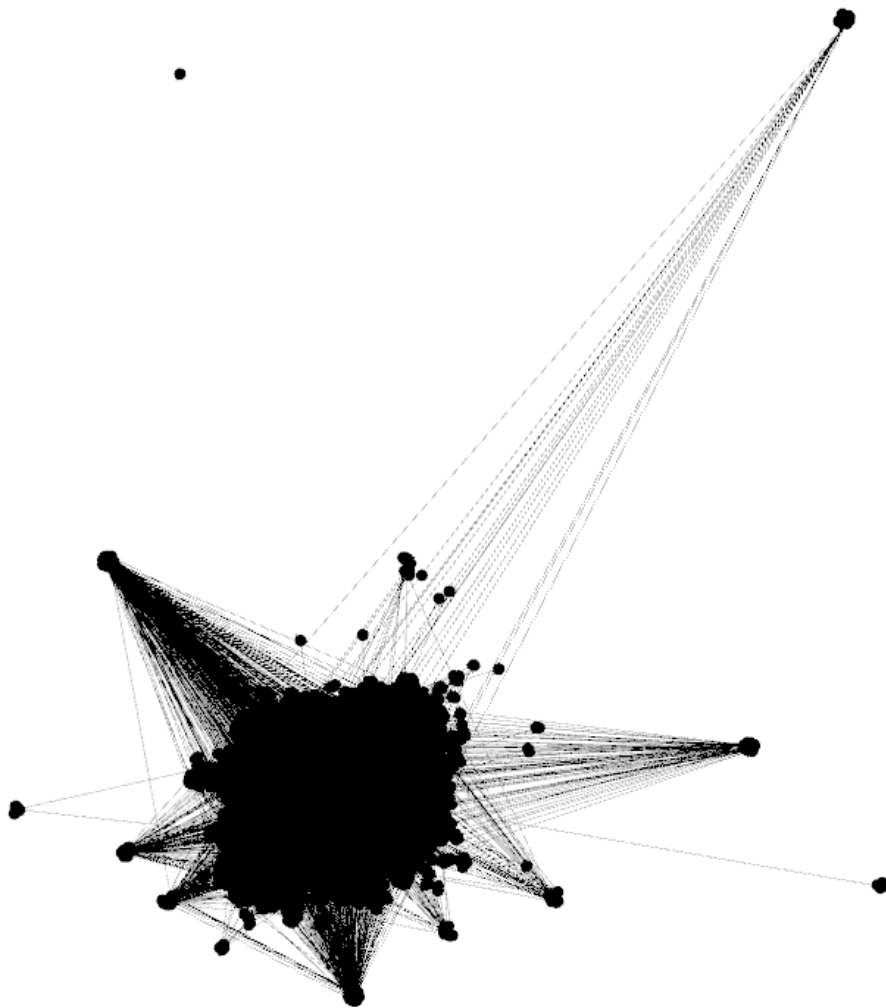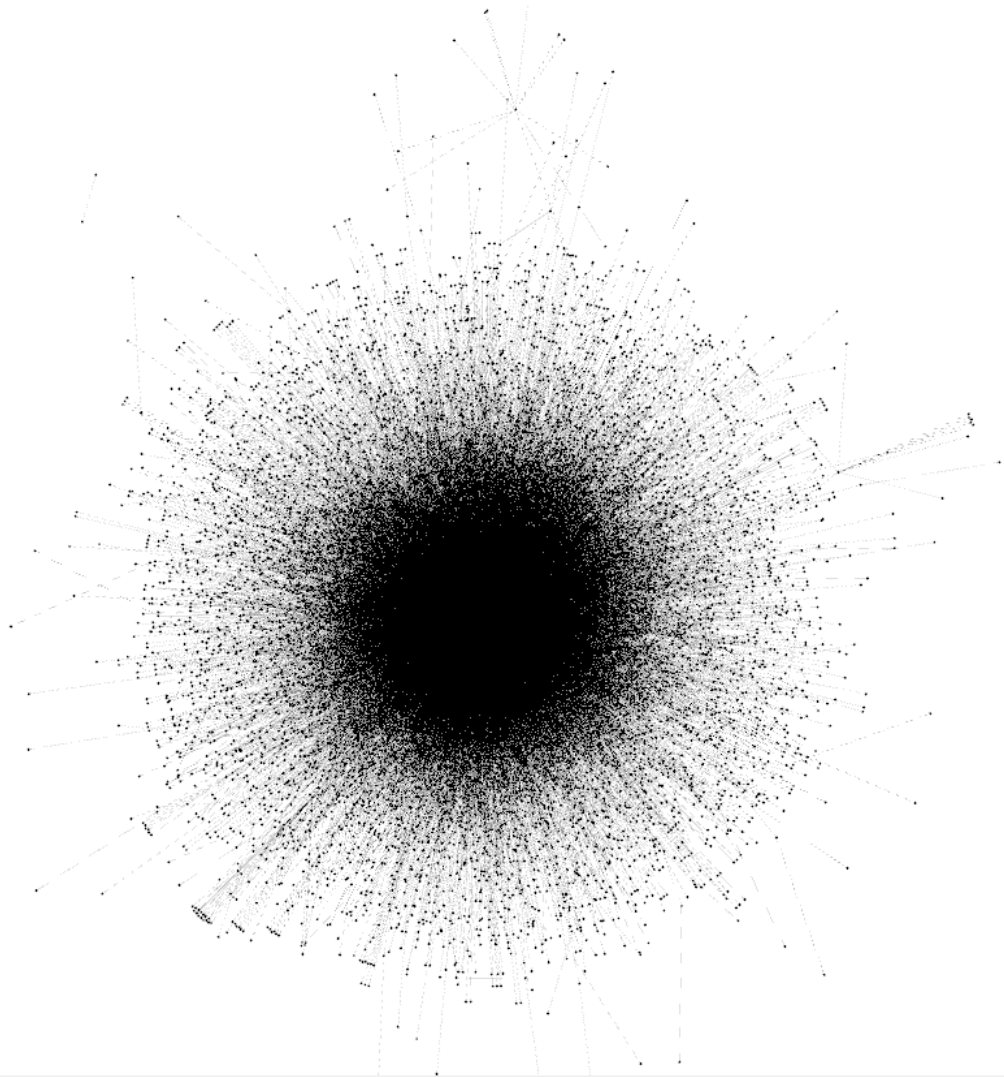
1.1 Open dataset
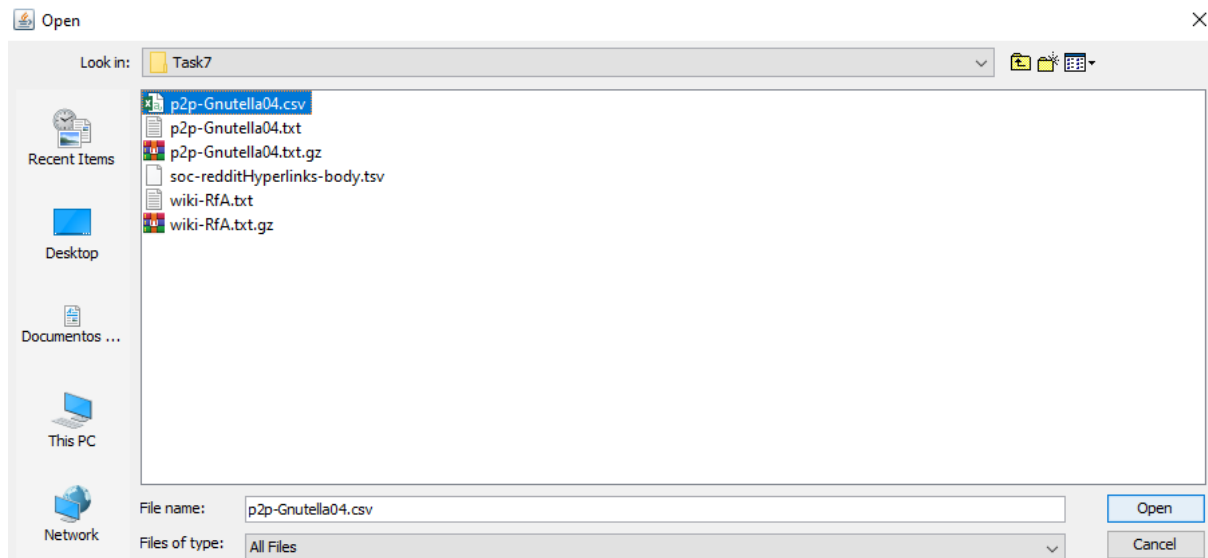
1.2 Overview
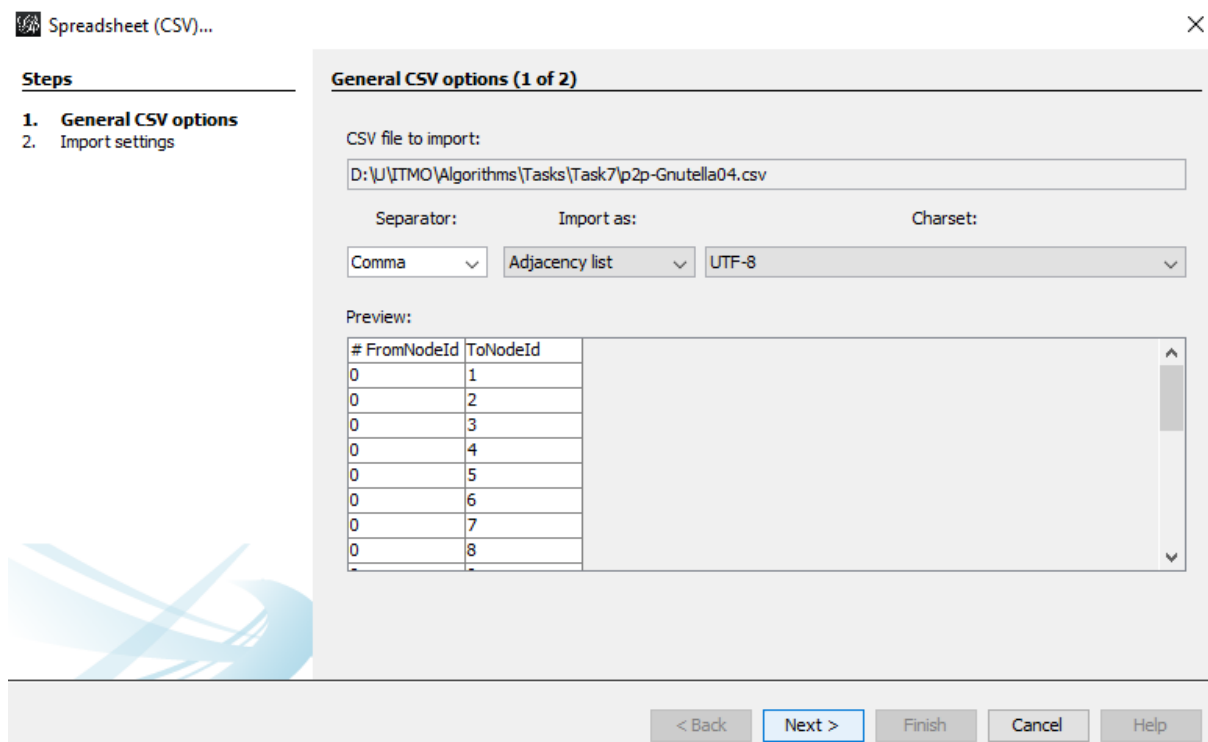
1.3 Initial grapfh

## 1.4 Openrd Layout

1.5 Yifan Hu Layout

Statisitcs

According snap

Dataset statistics	
Nodes	10876
Edges	39994
Nodes in largest WCC	10876 (1.000)
Edges in largest WCC	39994 (1.000)
Nodes in largest SCC	4317 (0.397)
Edges in largest SCC	18742 (0.469)
Average clustering coefficient	0.0062
Number of triangles	934
Fraction of closed triangles	0.001807
Diameter (longest shortest path)	9
90-percentile effective diameter	5.4

According Gephi

Nodes: 10878
Edges: 39995
Directed Graph

Filters	Statistics ✕		—
Settings			

| ▽ **Network Overview**			
Average Degree	3.677	Run	ⓘ
Avg. Weighted Degree	3.677	Run	ⓘ
Network Diameter	26	Run	ⓘ
Graph Density	0	Run	ⓘ
HITS		Run	ⓘ
Modularity	0.358	Run	ⓘ
PageRank		Run	ⓘ
Connected Components	2	Run	ⓘ
▽ **Node Overview**			
Avg. Clustering Coefficient	0.003	Run	ⓘ
Eigenvector Centrality		Run	ⓘ
▽ **Edge Overview**			
Avg. Path Length	6.771	Run	ⓘ

Average Degree: 3.677

Degree Distribution
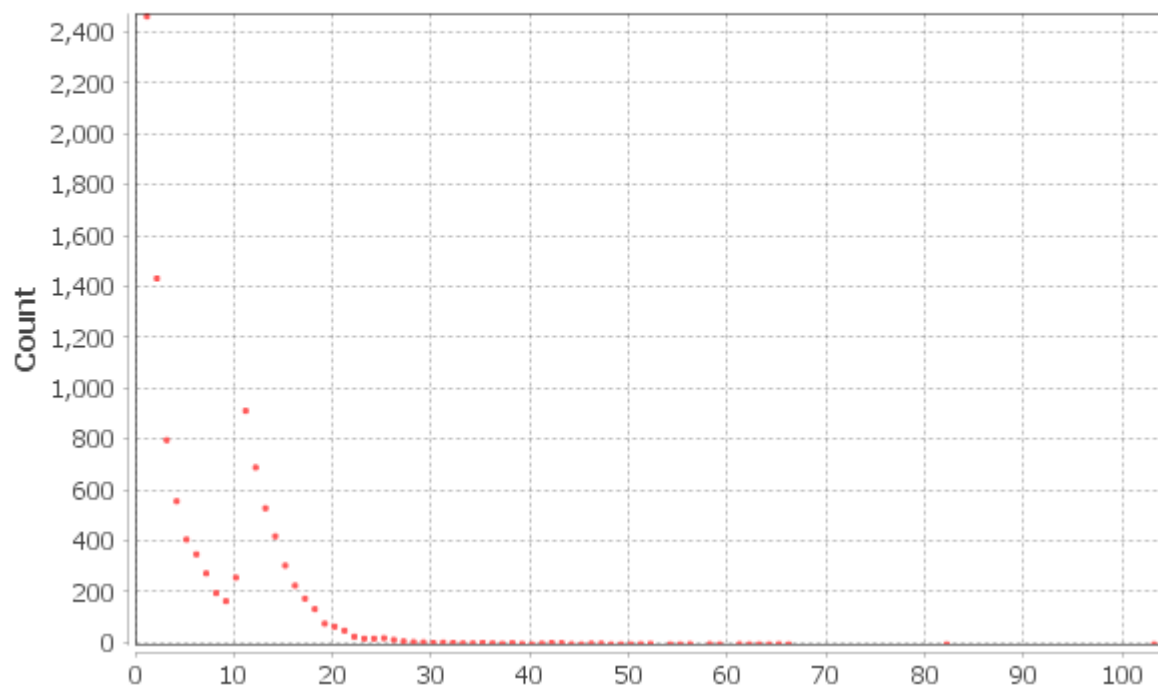
Average Weighted Degree: 3.677
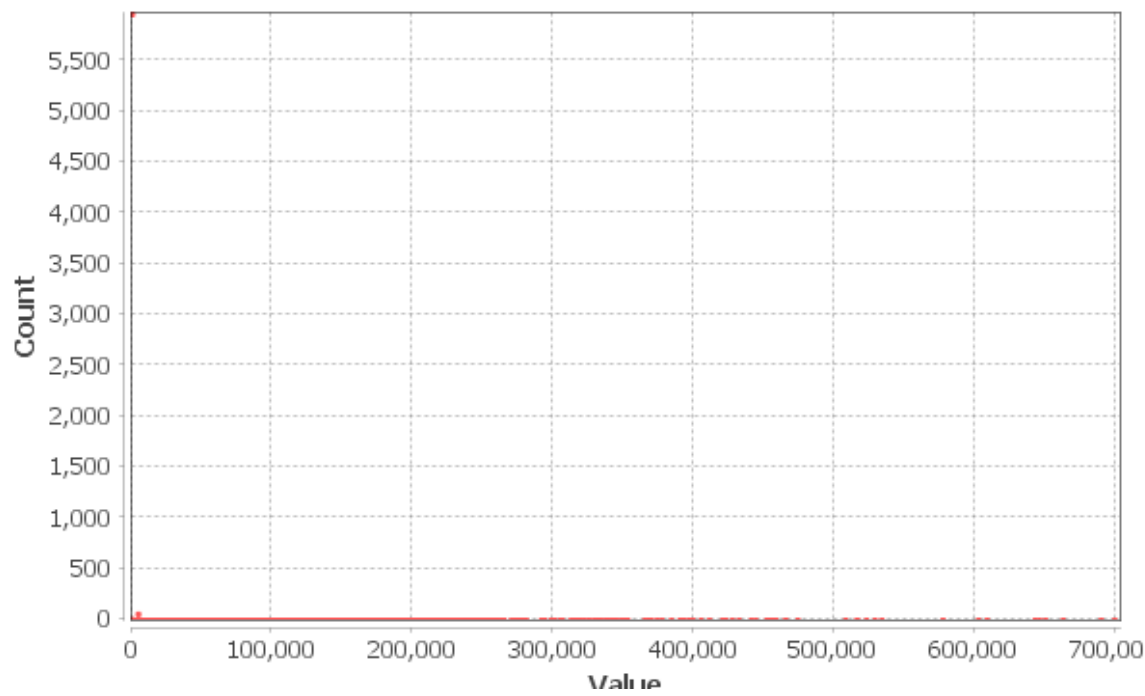
Degree Distribution
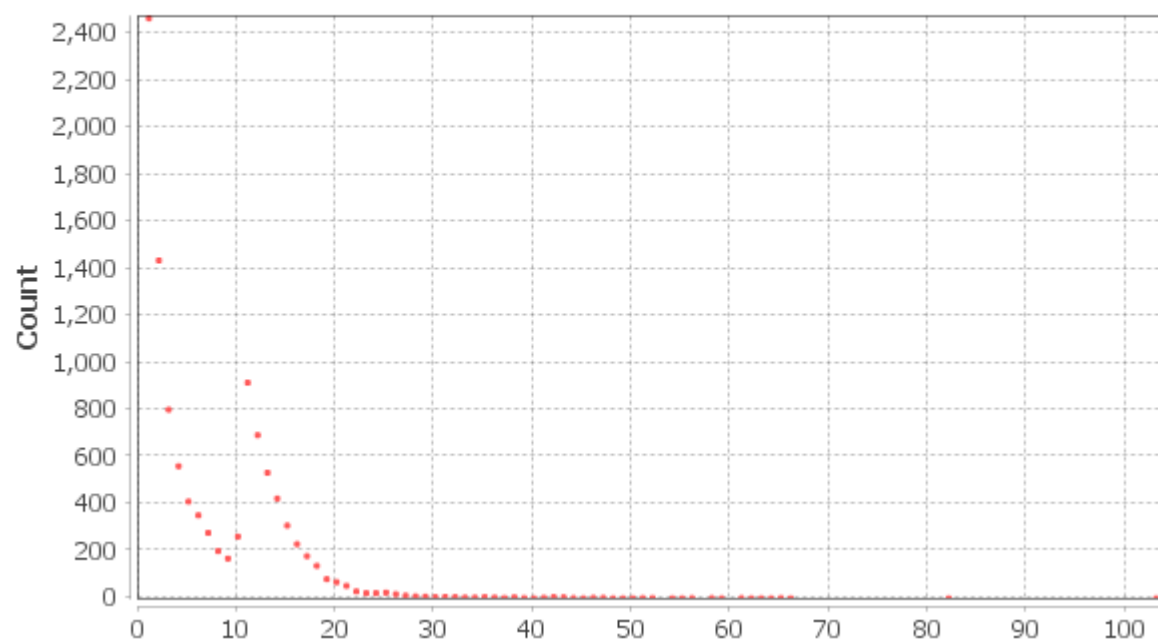
Parameters:

Network Interpretation: directed

Results:

Diameter: 26
Radius: 0
Average Path length: 6.7705442867953565
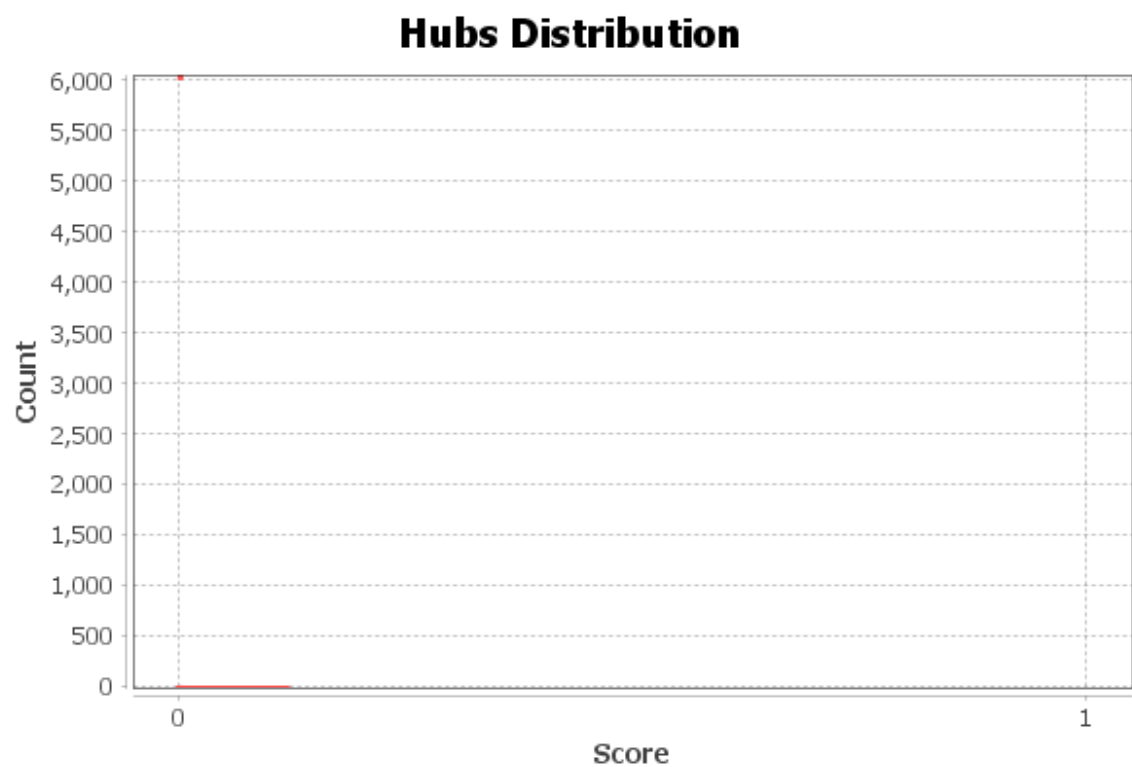
Betweenness Centrality Distribution

HITS Metric Report

Parameters:

E = 1.0E-4

Results:

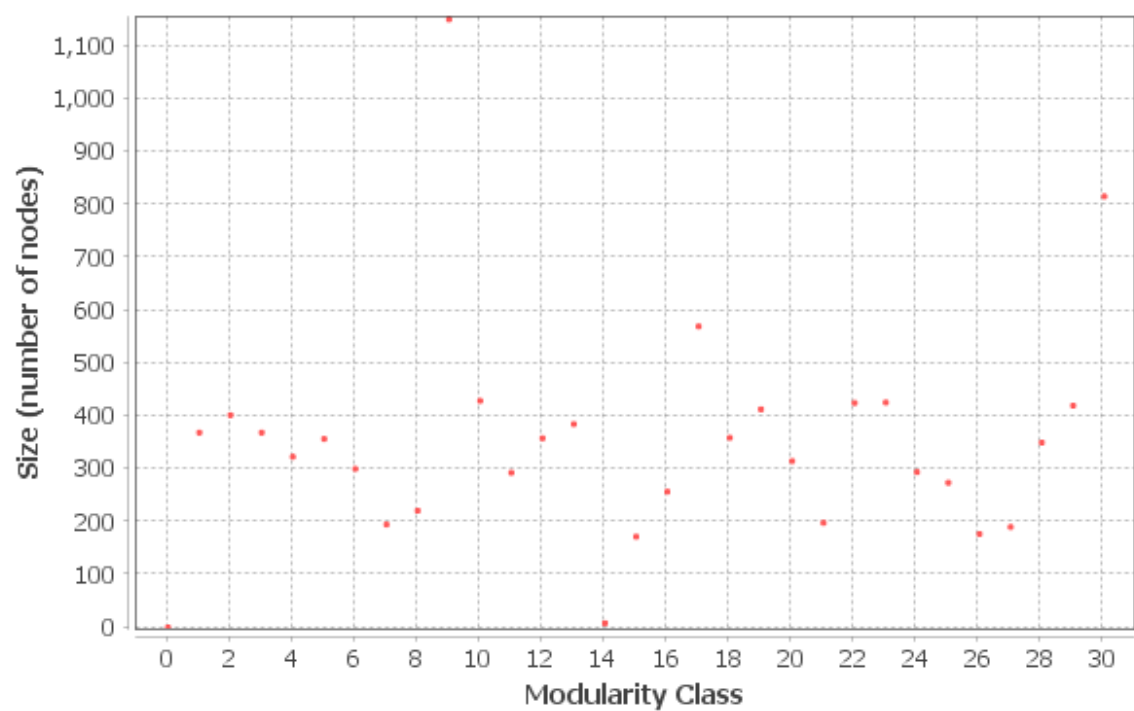### Hubs Distribution

Modularity Report

Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

Modularity: 0.358
Modularity with resolution: 0.358
Number of Communities: 31

Size Distribution

PageRank Report

Parameters:

Epsilon = 0.001
Probability = 0.85

Results:

PageRank Distribution

Clustering Coefficient Metric Report

Parameters:

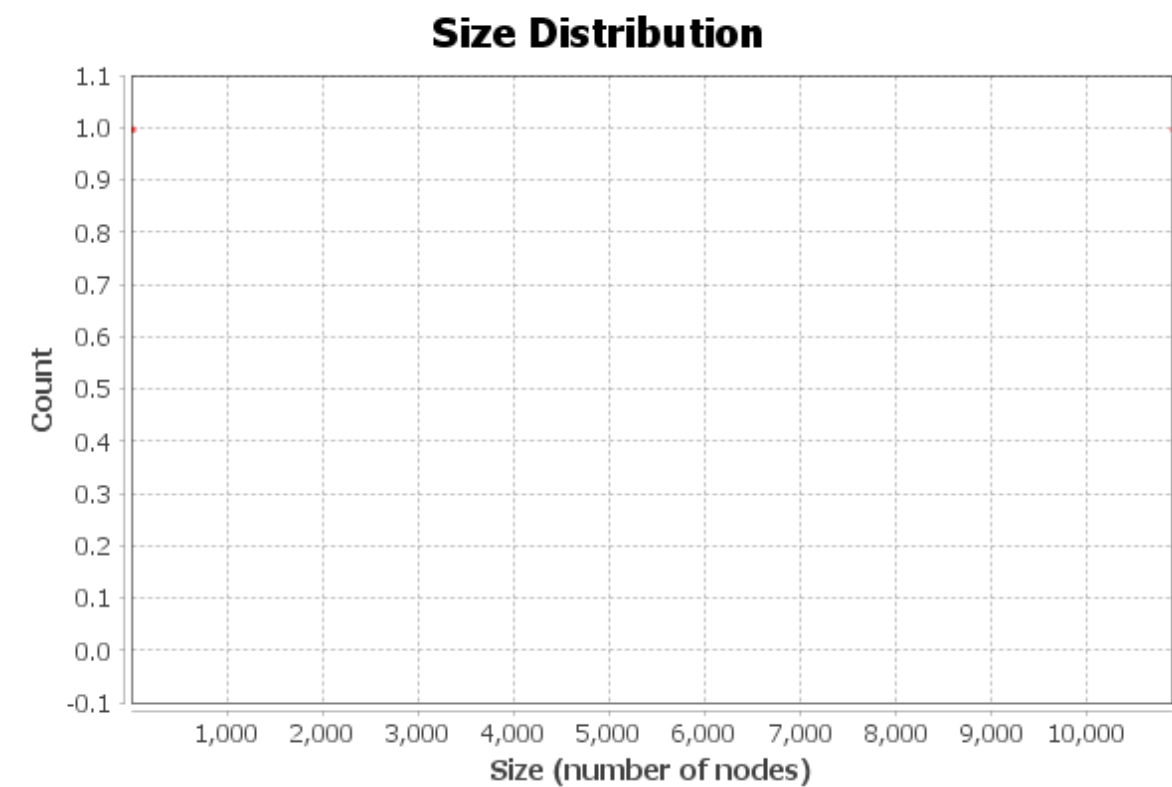Network Interpretation: directed

Results:

Average Clustering Coefficient: 0.003
The Average Clustering Coefficient is the mean value of individual coefficients.

Parameters:

Network Interpretation: directed

Results:

Number of Weakly Connected Components: 2
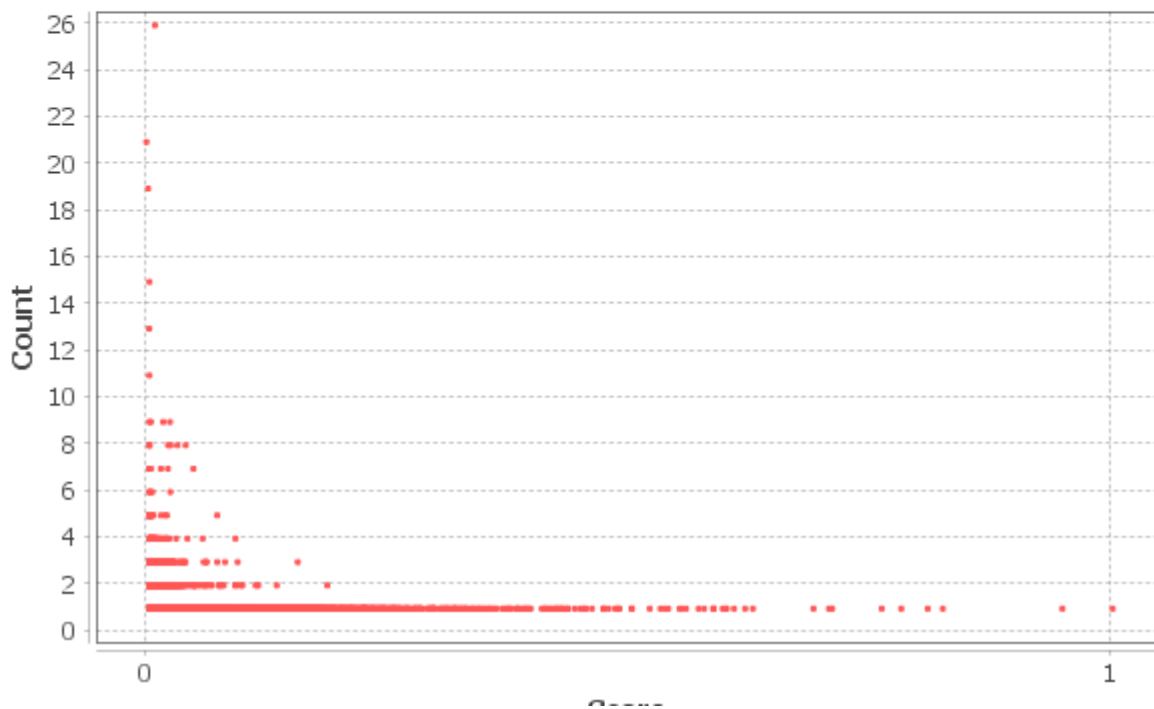Number of Strongly Connected Components: 6562

Size Distribution

Size Distribution plot. X-axis: Size (number of nodes), ranging from 1,000 to 10,000. Y-axis: Count, ranging from -0.1 to 1.1.

Eigenvector Centrality Report

Parameters:

Network Interpretation: directed
Number of iterations: 100
Sum change: 0.4855992486448611

Results:

Eigenvector Centrality Distribution

Analysis

- Accordig to the average degree = 3.67, that stablish that for every node acting like seed there are in average 3 or more leeches
- The major distance between two peers is 26 according to he network diameter
- There is zero density so that there are no steps in the middle of existing connections
- The modularity is less than 0.4, so there is not a strong cohesion in the network
- According the average path lenght the medium distance between to peers is 6.771

Conclusions

- Using Gephi we can have an easy and accurate view of the relationships in graphs and their characteristics
- Knowing the meaning of statistics characteristics related to a graph we can understand the information relaying on the system described
- Openrd has a good work on processing large datasets and offers quite good way of sisualization, his time performance was less than Yifan Hu
- Yifan Hu has a very good time execution and works very well on large datasets

Bibliography

[1] Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.
[2]Otte, Evelien; Rousseau, Ronald (December 2002). "Social network analysis: a powerful strategy, also for the information sciences"