Social Network Analysis – Fall 2018
Assignment 1

INDEX

1. The Gephi & Extracting Datasets.

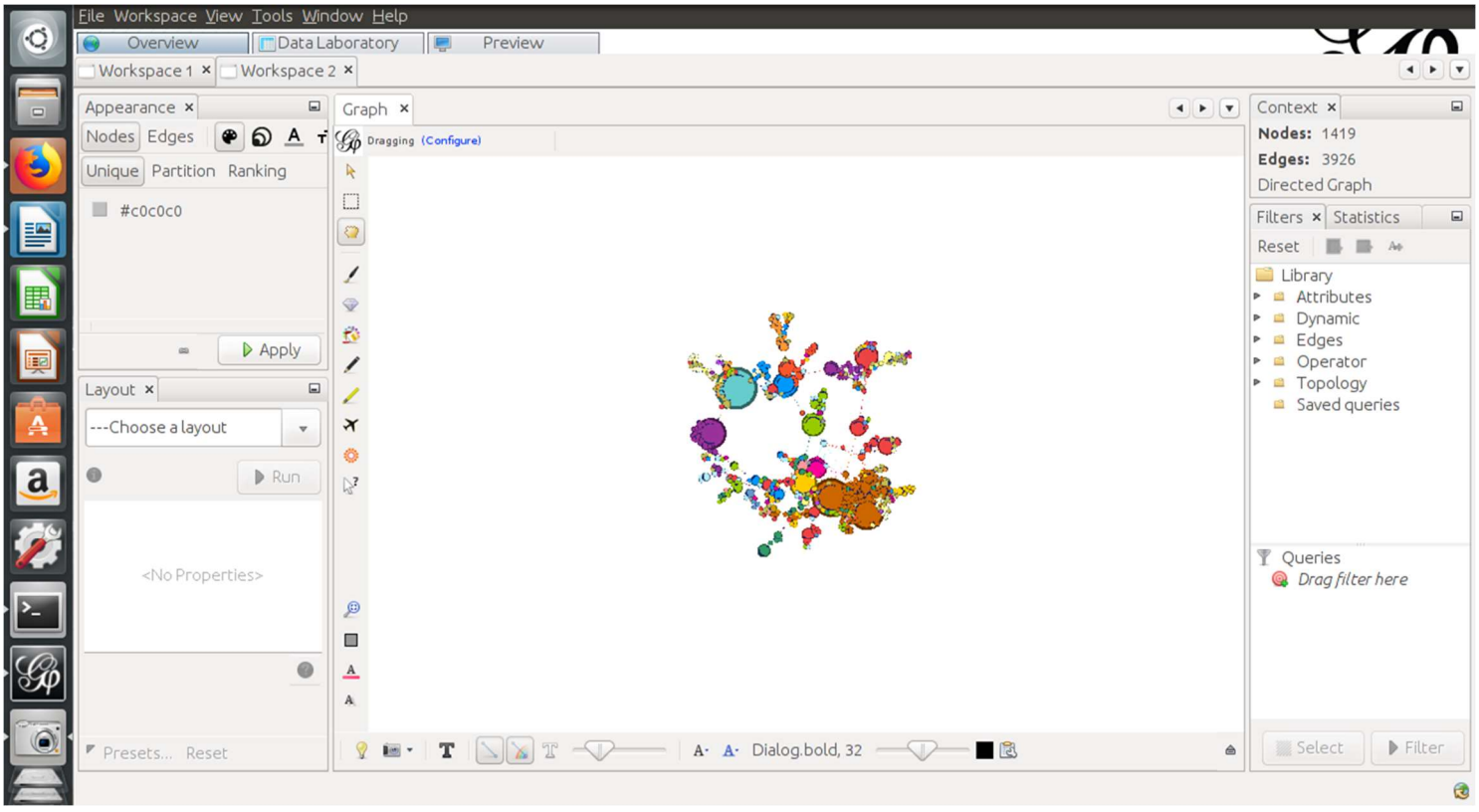
Dataset: Diseasome

Vertices :1419

Edges :3926

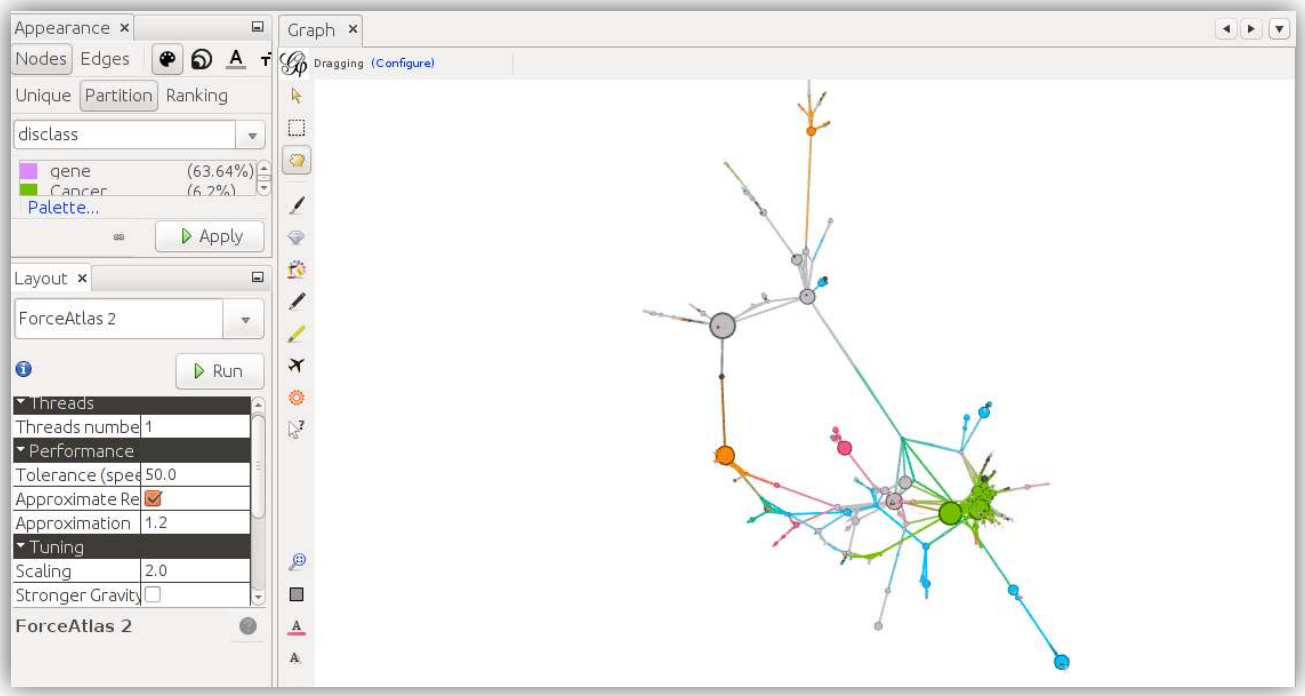
Diseasome: A network of disorders and disease genes linked by known disorder–gene associations, indicating the common genetic origin of many diseases. Genes associated with similar disorders show both higher likelihood of physical interactions between their products and higher expression profiling similarity for their transcripts, supporting the existence of distinct disease-specific functional modules.

Dataset Reference: The Human Disease Network, Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L (2007), Proc Natl Acad Sci USA 104:8685-8690

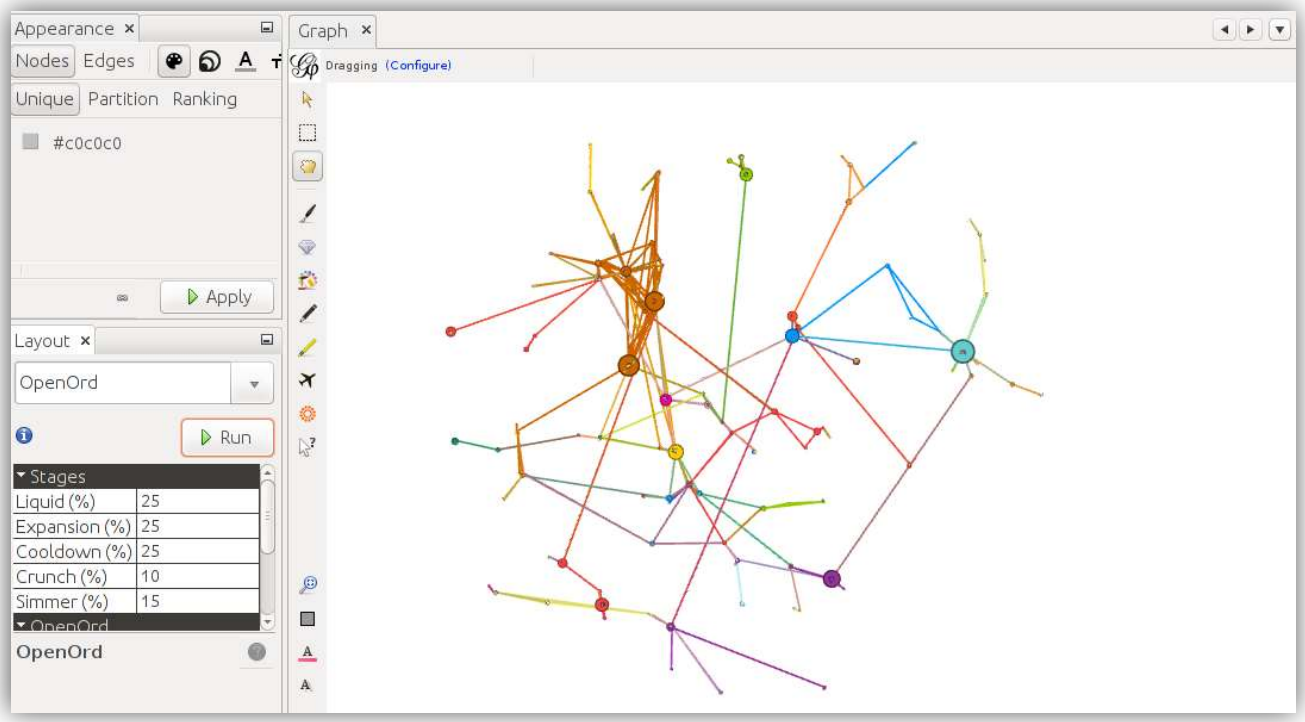


2. Setting Different Layouts.

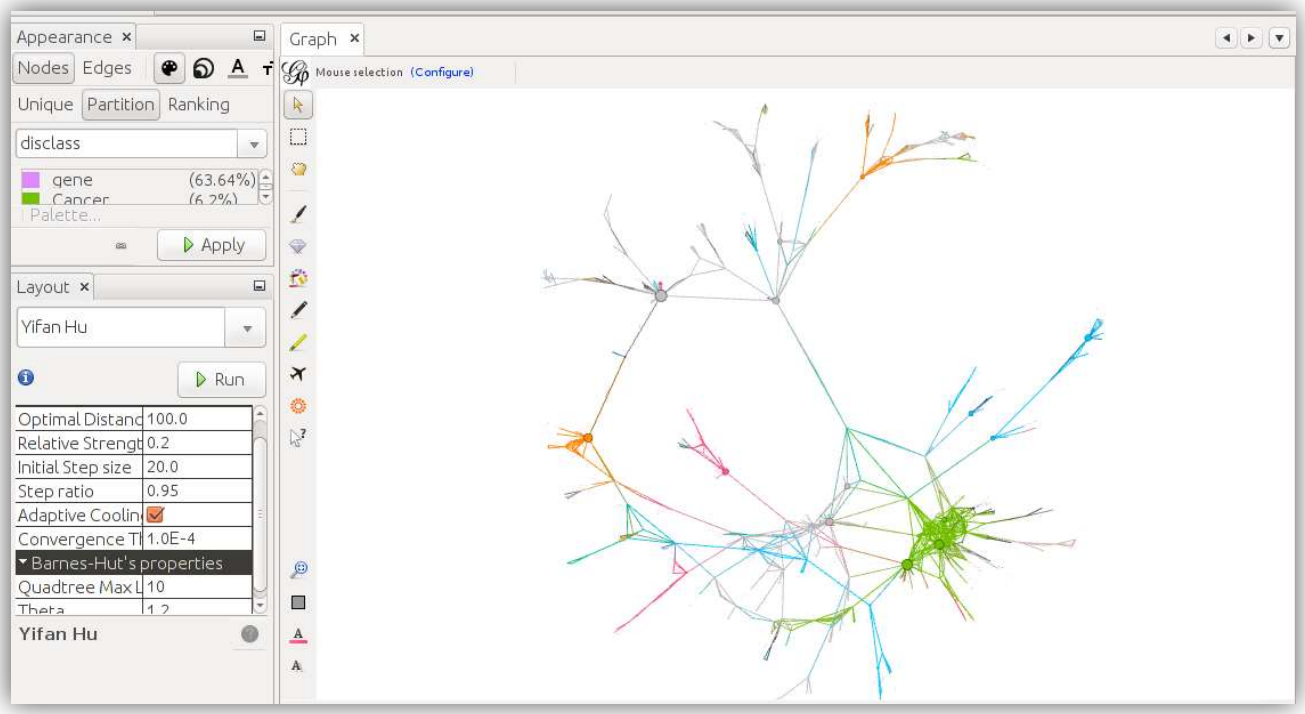
i) ForceAtlas 2



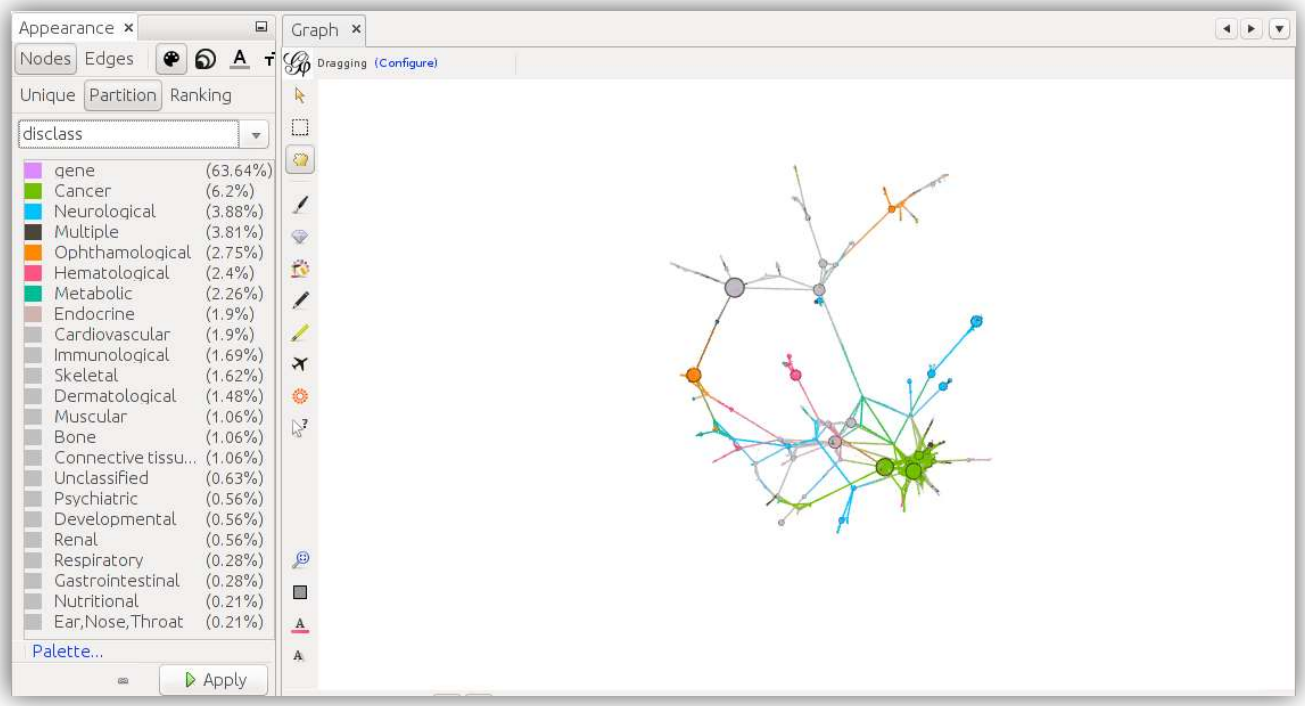
ii) Openord



iii) Yifan Hu

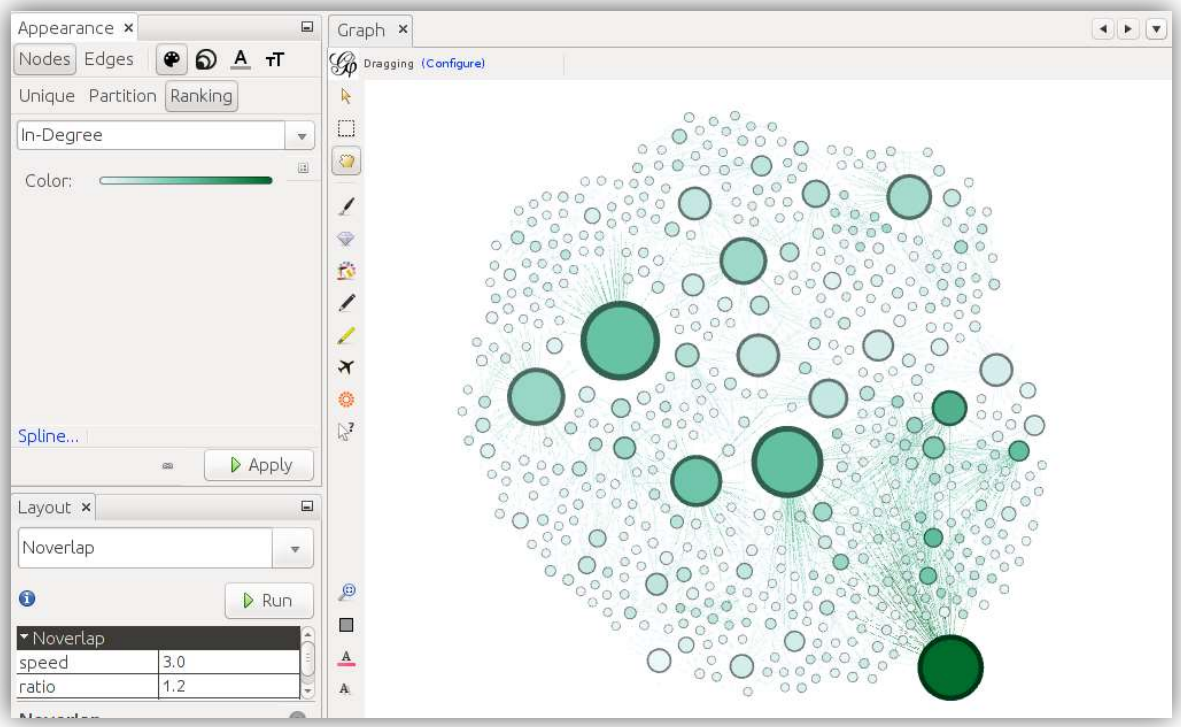


iv) ForceAtlas (with Partition by class)

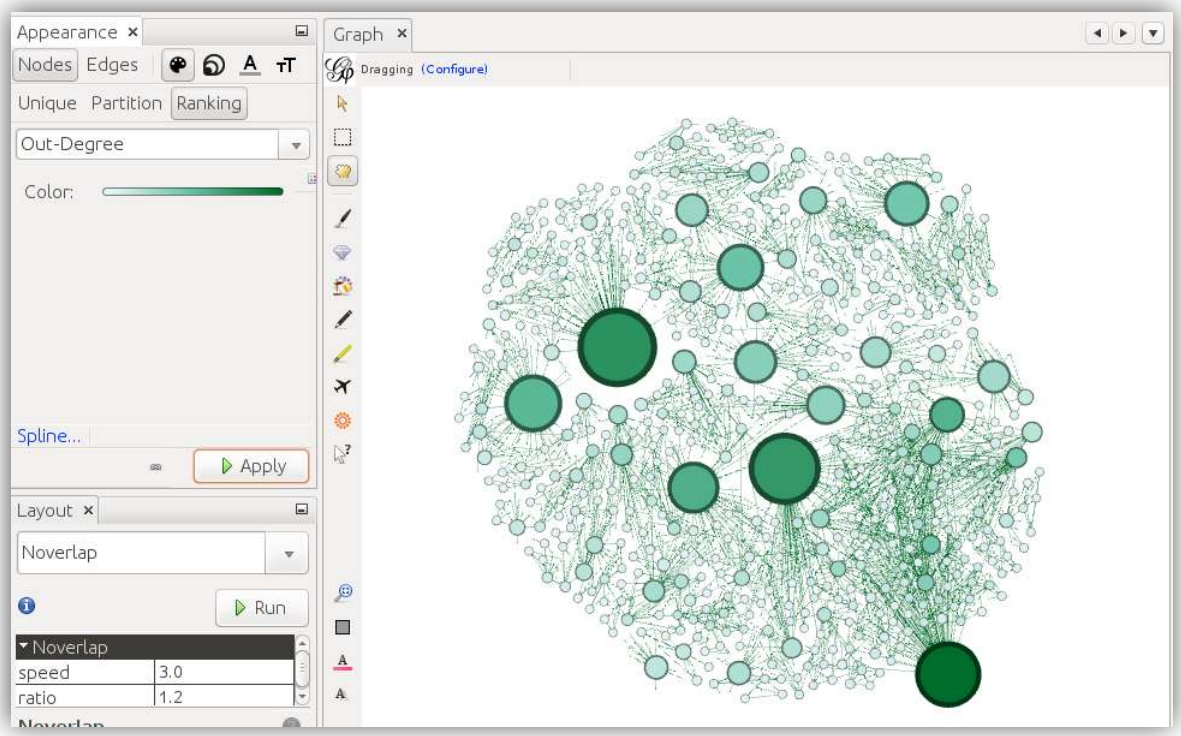


3. Distribution & Visualizing nodes by their Sizes

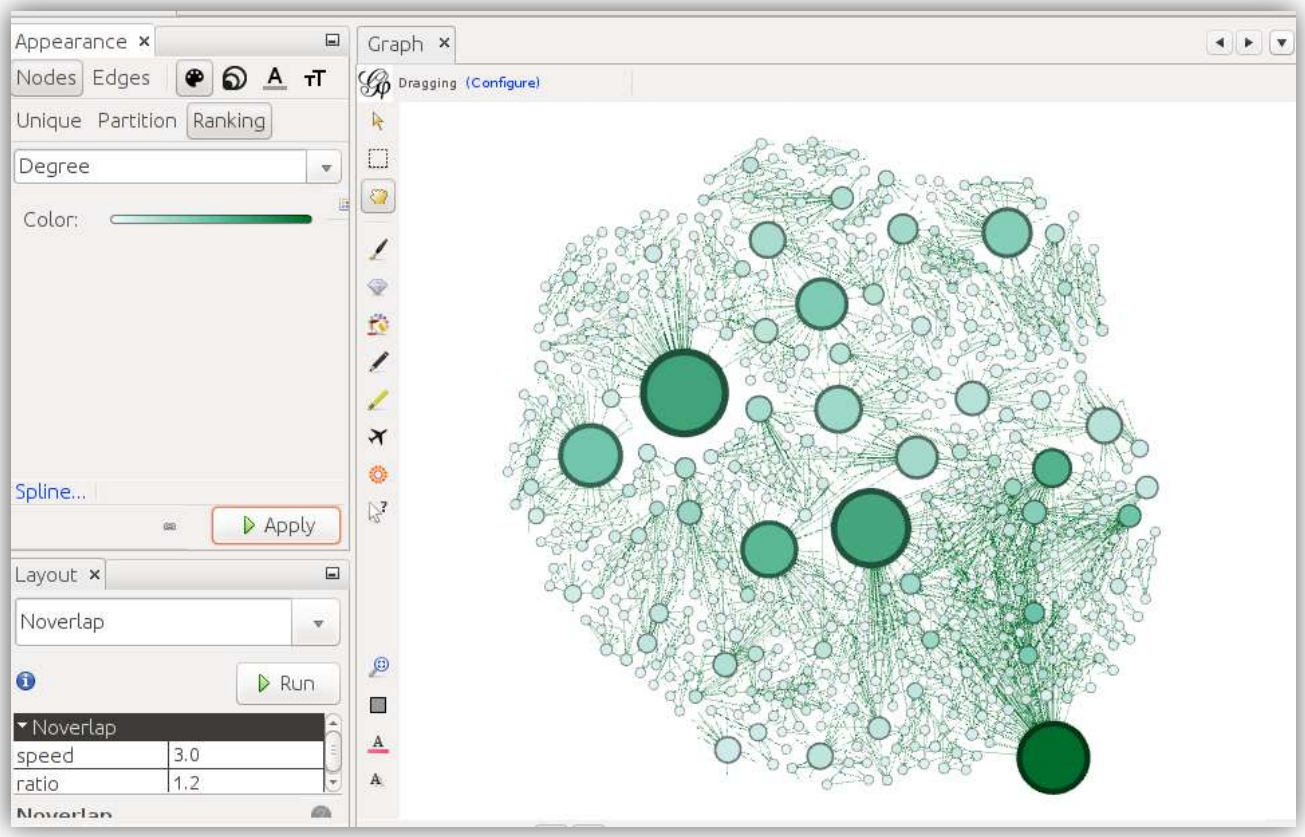
i) **Indegree** (Colour is also used alongside to enhance visualization & egdes are also darkened accordingly)



ii) **OutDegree** (Colour is also used alongside to enhance visualization & egdes are also darkened accordingly)

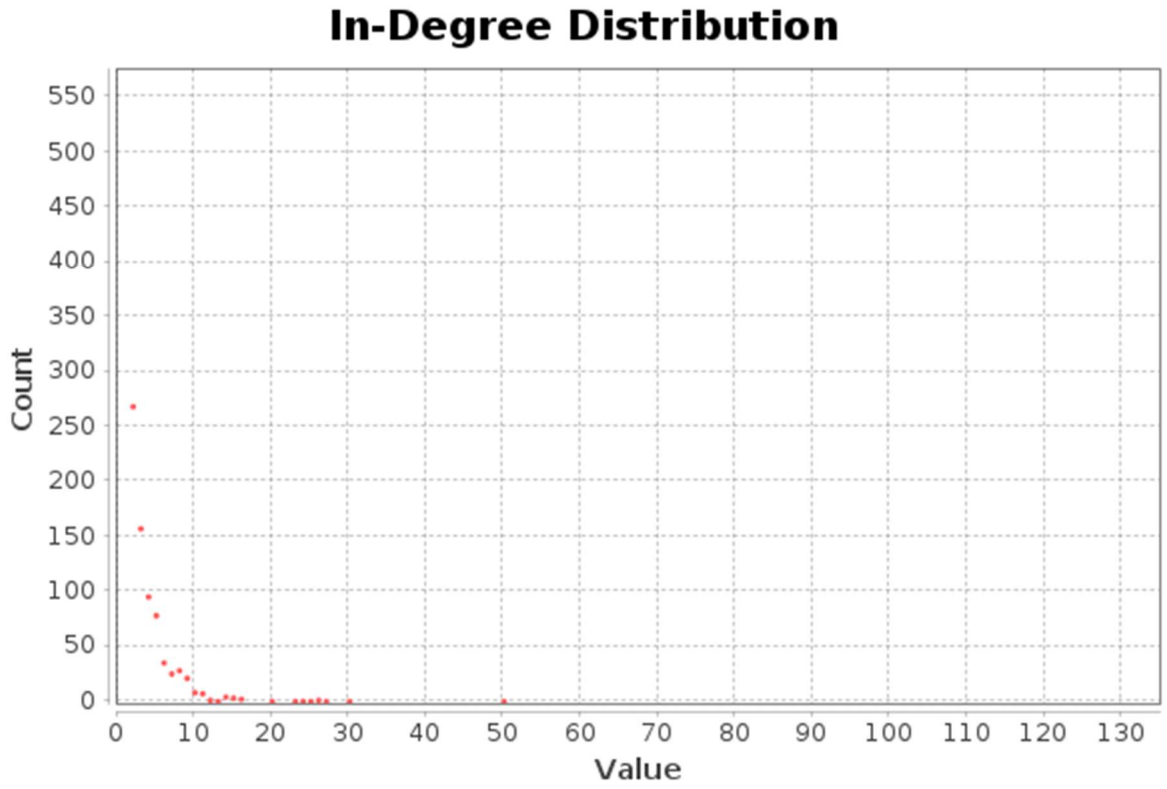


iii) **Total Degree** (Colour is also used alongside to enhance visualization & egdes are also darkened accordingly)

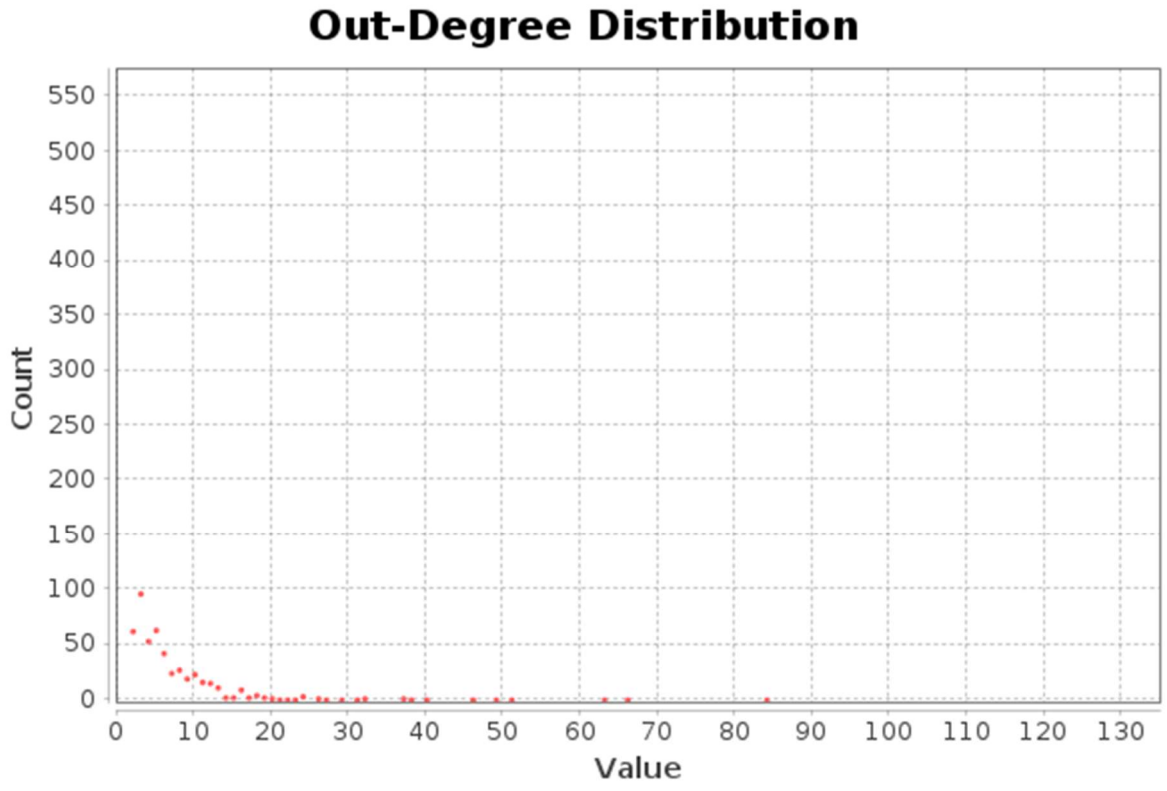


Degree Distribution

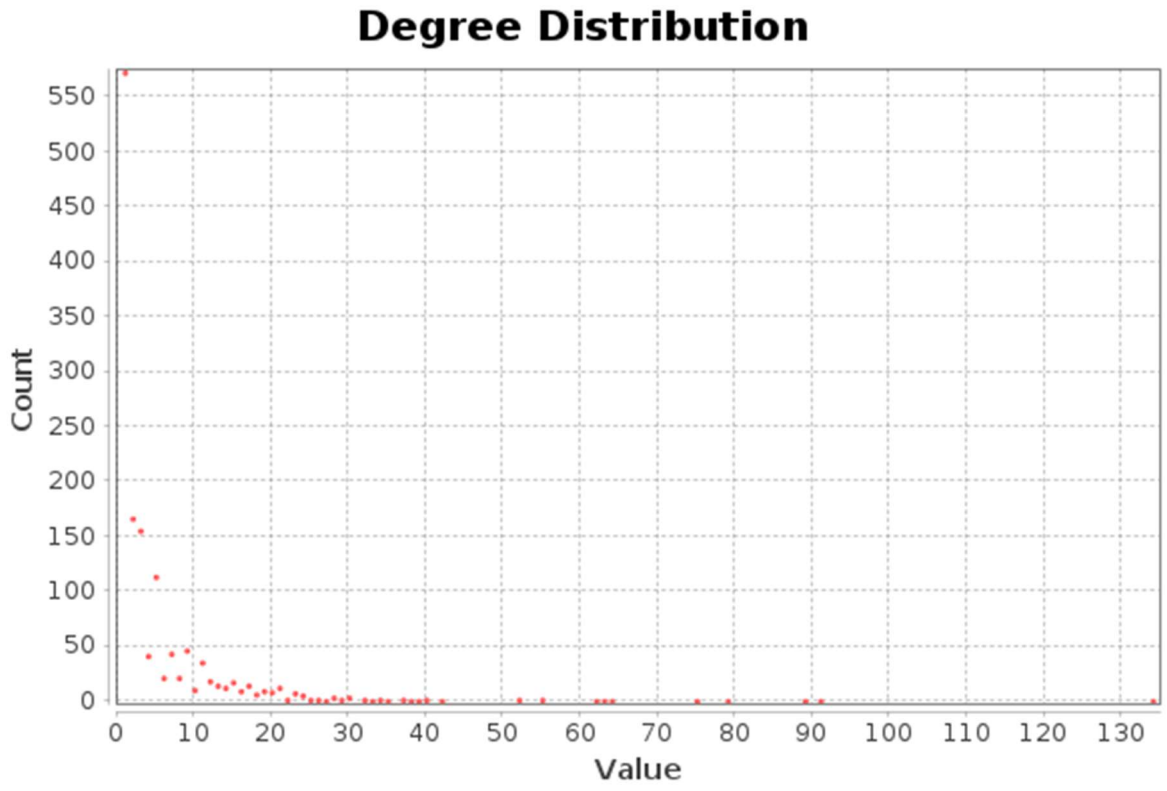
i) Indegree



ii) OutDegree



iii) Total Degree



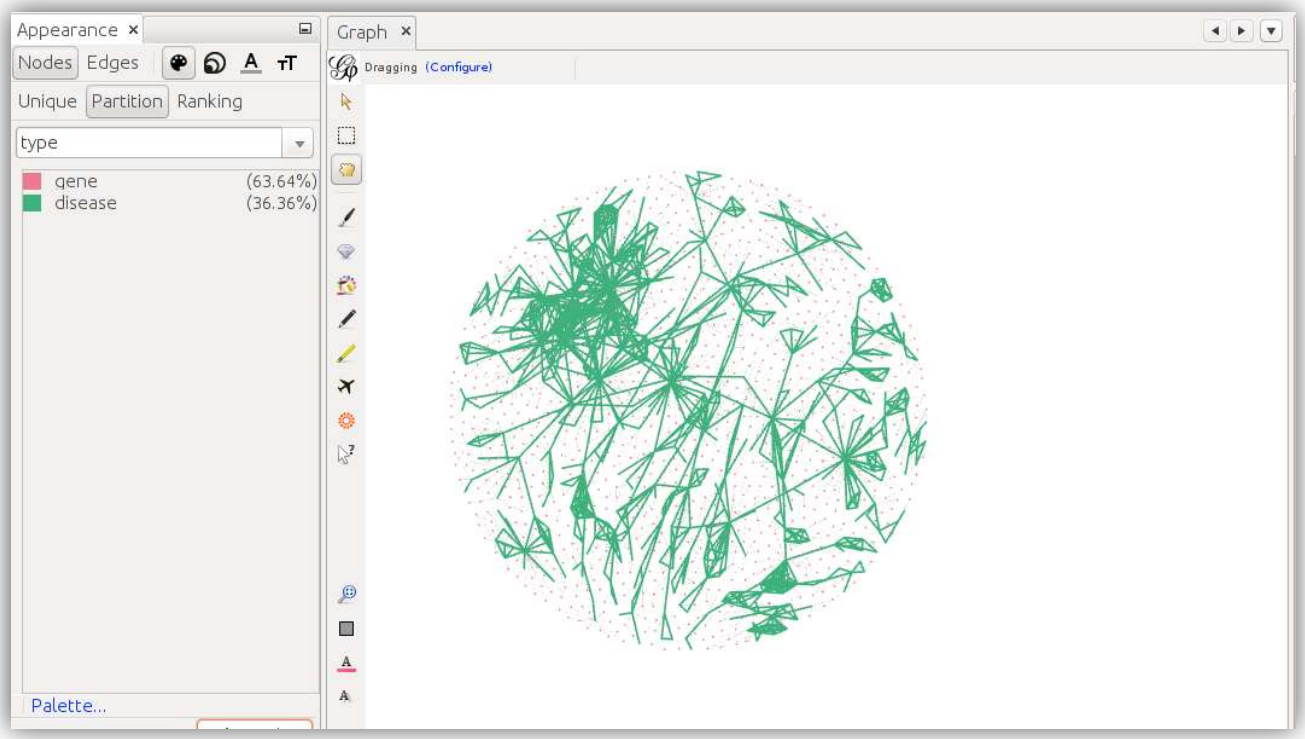
Results:

Average Degree: 2.767

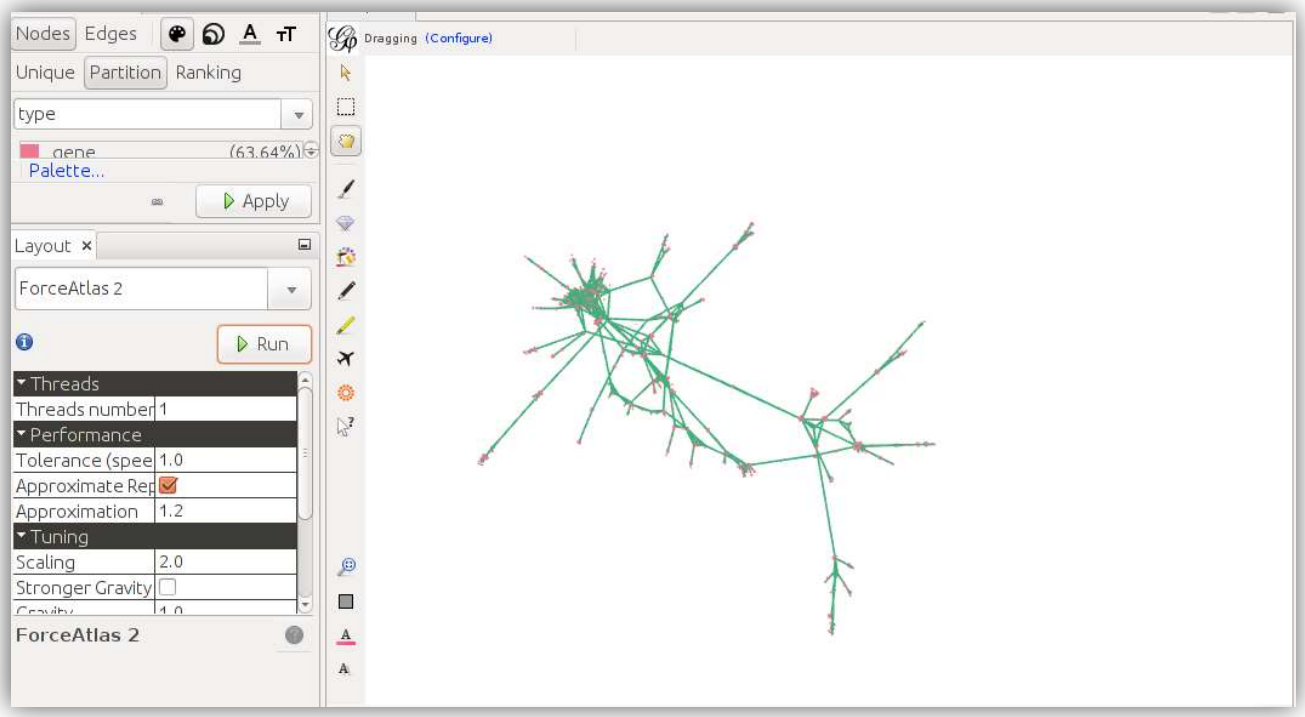
4. Identifying Communities in the Network.

A) Assigning colour to similar community

i) Fruchterman Reingold layout visualization of community



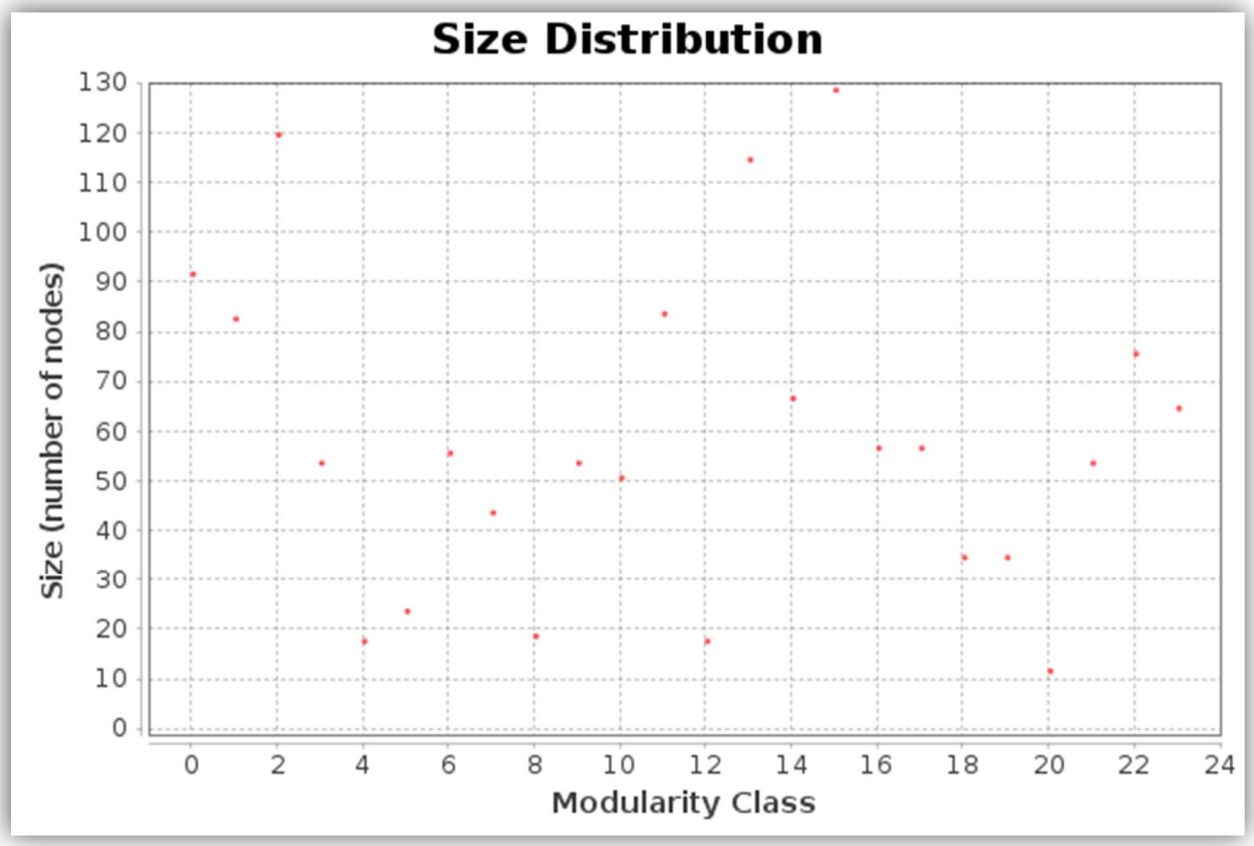
ii) ForceAtlas layout Visualization of community(Genes & Disease)



B) Modularity Value for Communities.

Parameters:
Randomize: On
Use edge weights: On
Resolution: 1.0

Results:
Modularity: 0.863
Modularity with resolution: 0.863
Number of Communities: 24



Algorithm:

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, *Fast unfolding of communities in large networks*, in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000

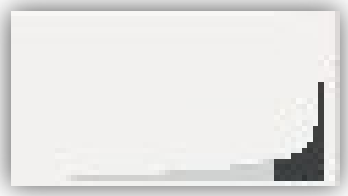
Resolution:

R. Lambiotte, J.-C. Delvenne, M. *Barahona Laplacian Dynamics and Multiscale Modular Structure in Networks* 2009

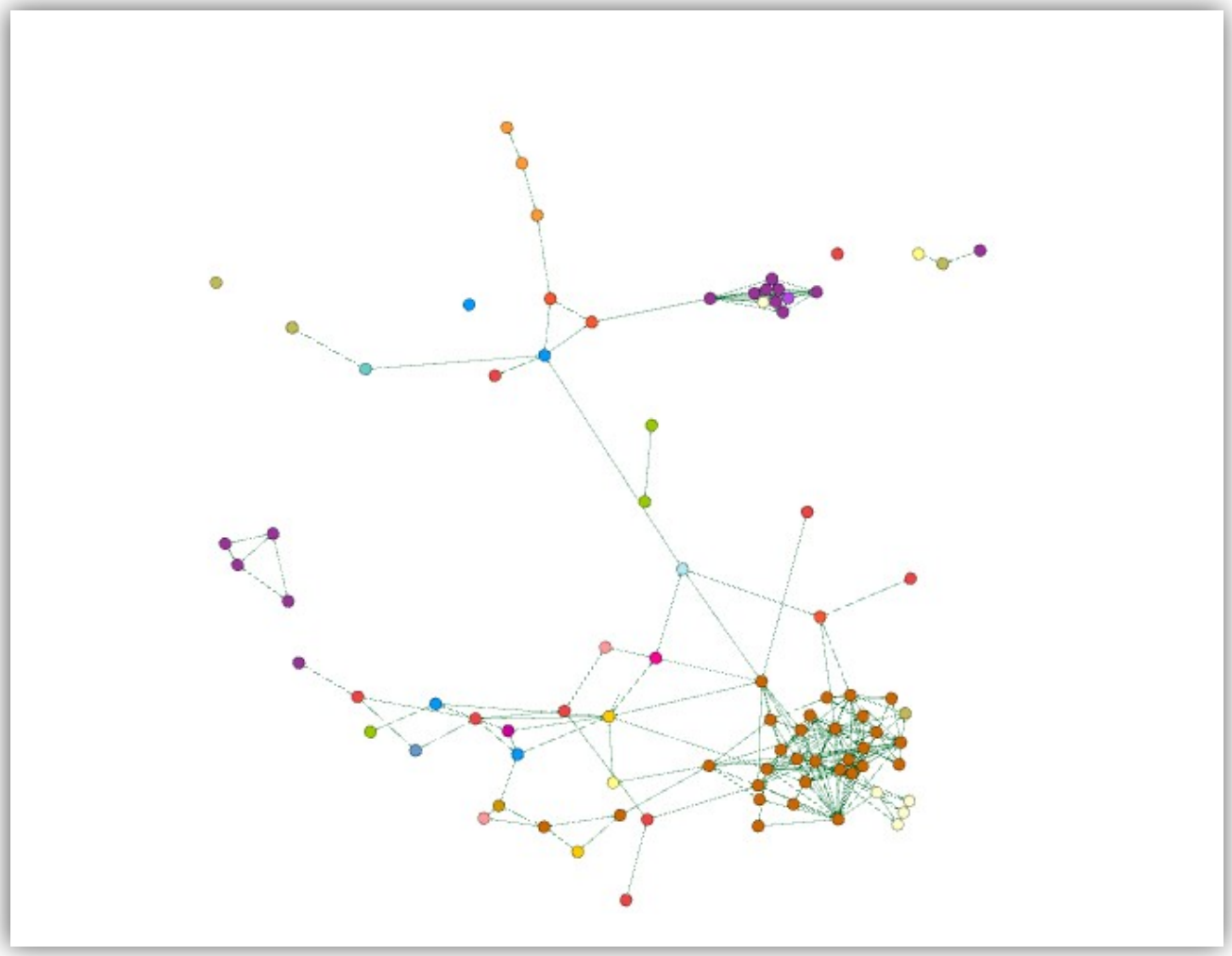
5. Filtering the Network (By Degree)

A) Top 5%

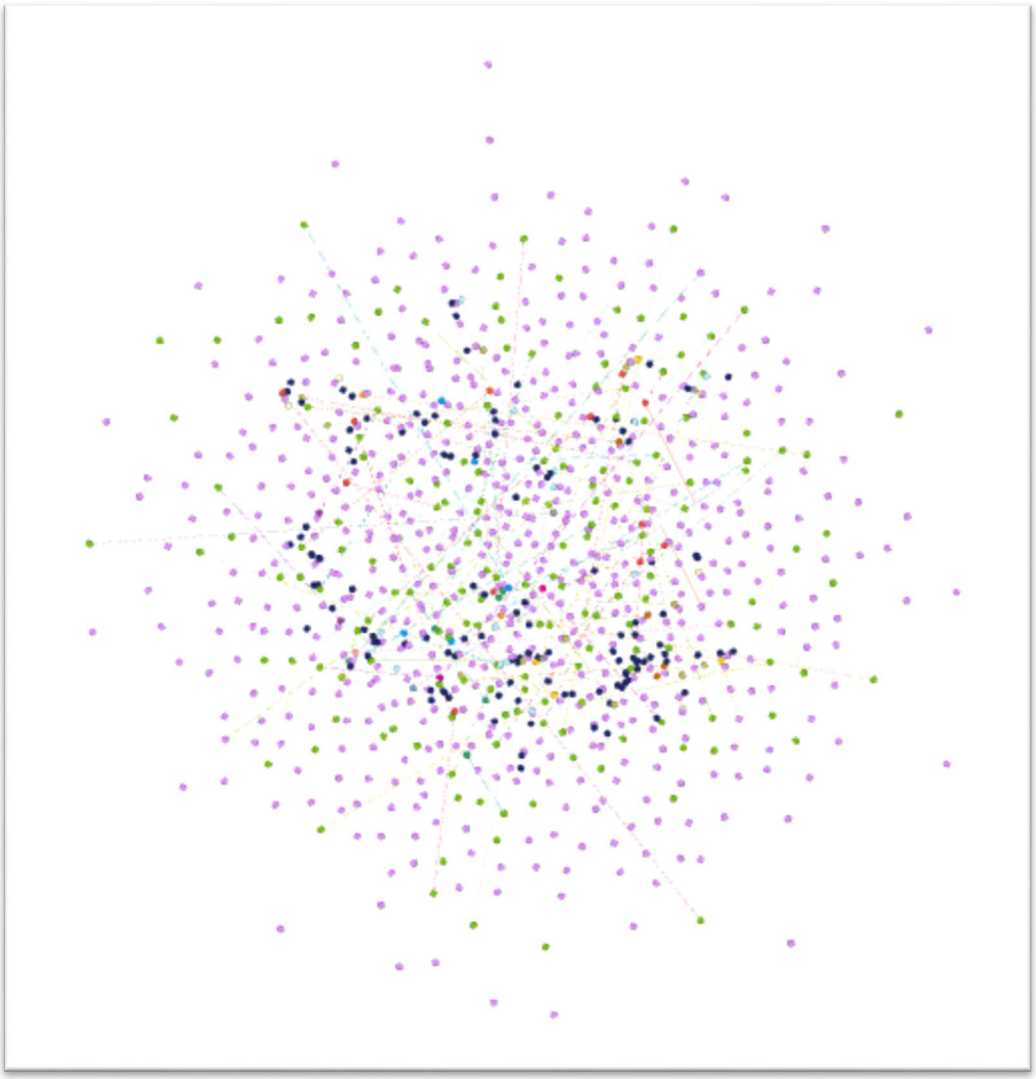
i) Degree Distribution



ii) Visualization.



B) Bottom 10%



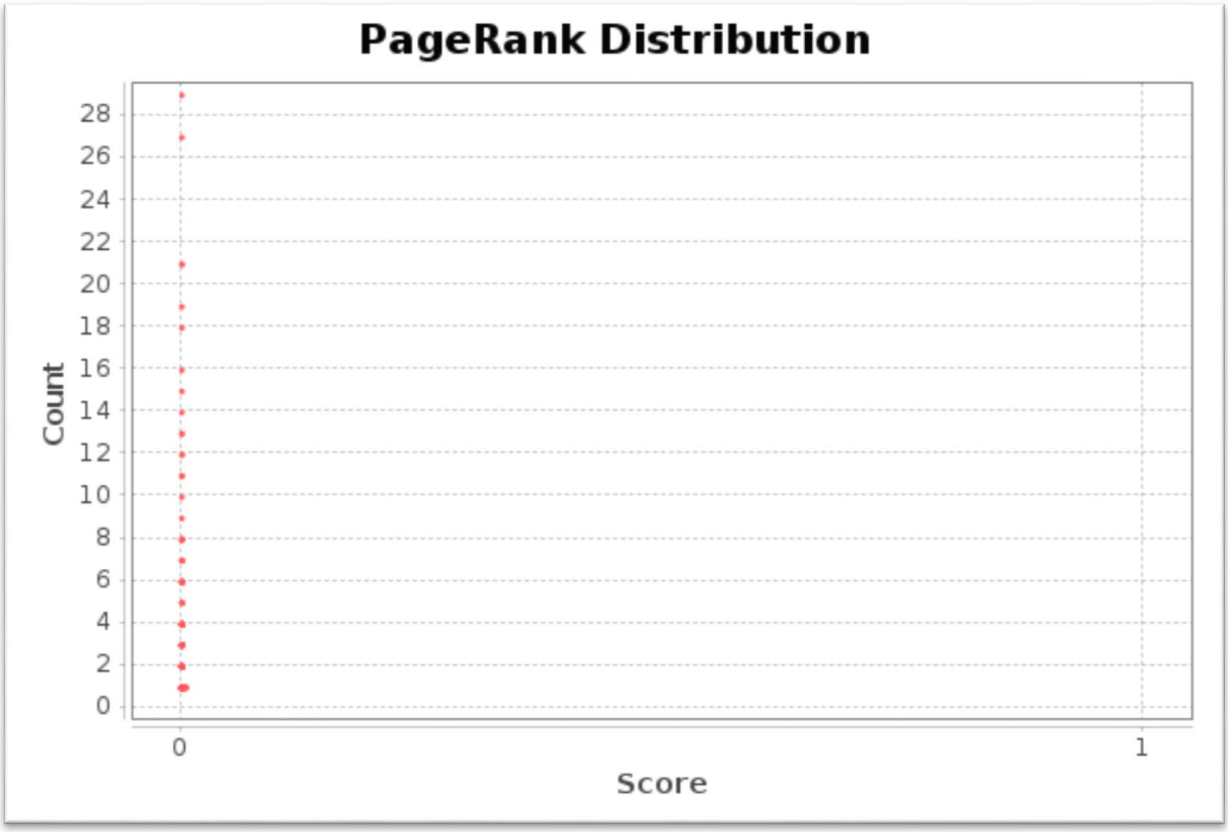
6. Pagerank & Hits Distribution

A) Pagerank

Parameters:

Epsilon = 0.001
Probability = 0.85

Results:

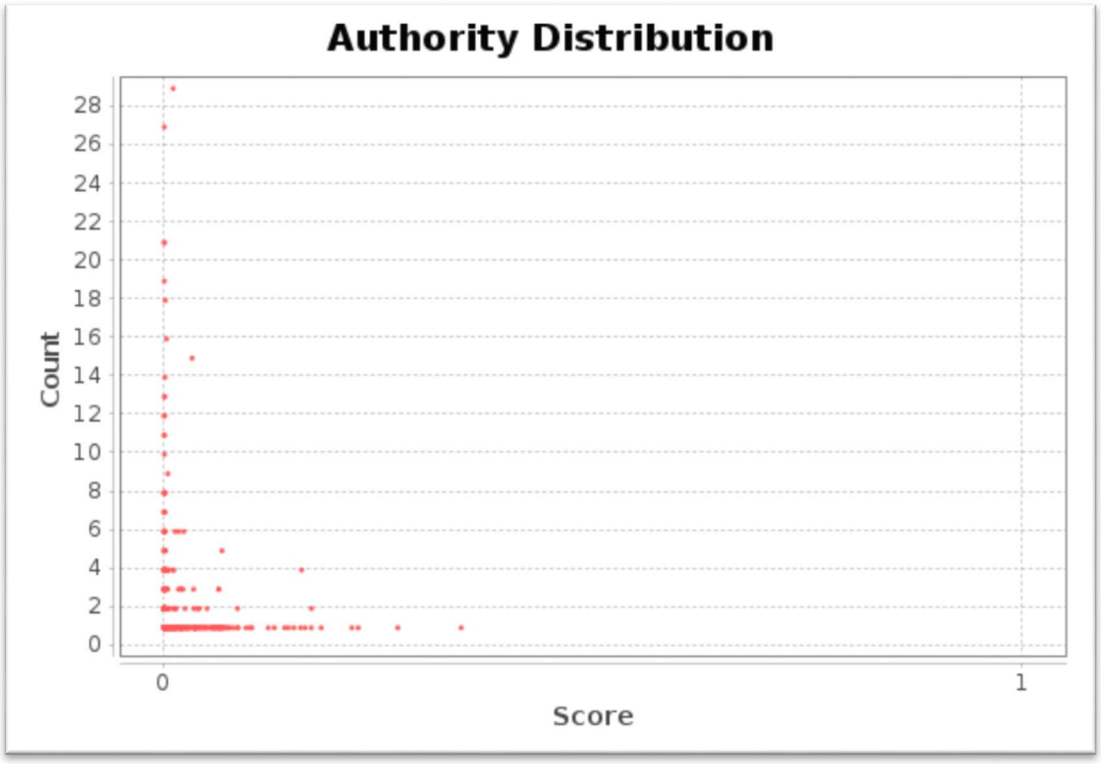


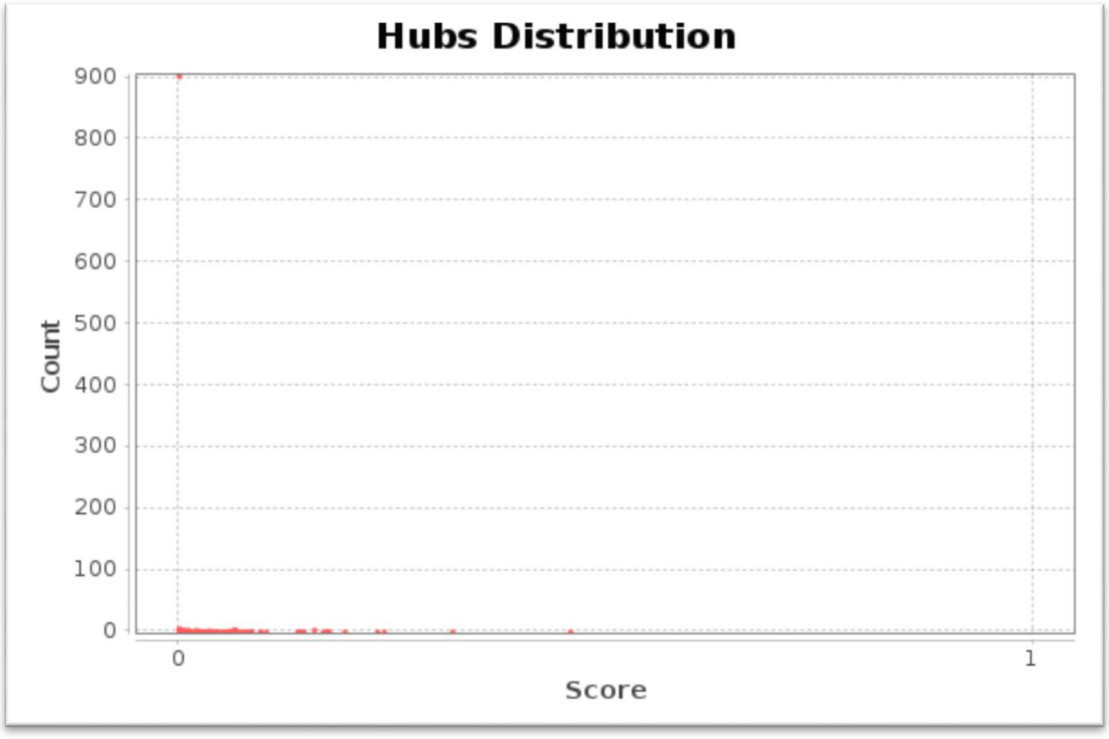
Algorithm:

Sergey Brin, Lawrence Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, in Proceedings of the seventh International Conference on the World Wide Web (WWW1998):107-117

B) Hits Metric Report

Results: **E = 1.0E-4**





Algorithm: Jon M. Kleinberg, *Authoritative Sources in a Hyperlinked Environment*, in Journal of the ACM 46 (5): 604–632 (1999)

7. Components of Network

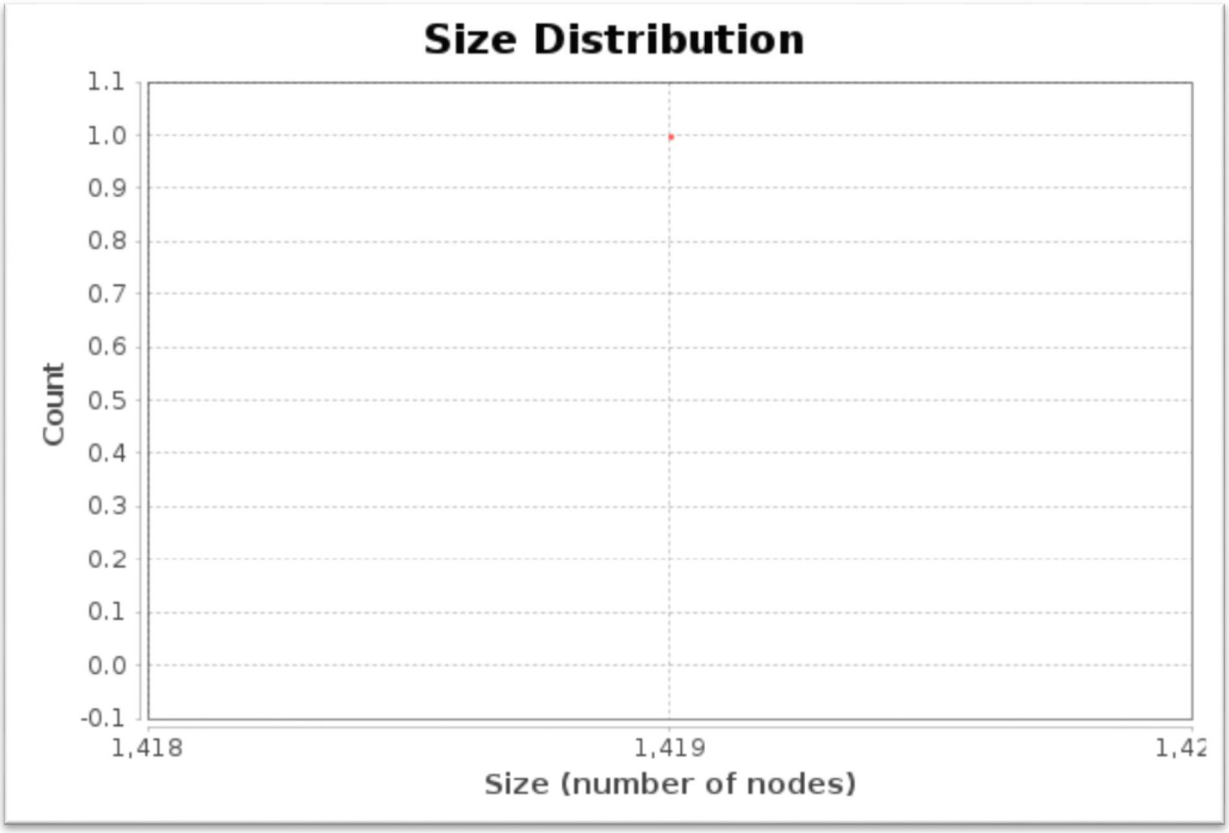
A) **Connected Component.**

Parameters:

Network Interpretation: directed

Results:

Number of Weakly Connected Components: 1
Number of Strongly Connected Components: 904



Algorithm:

Robert Tarjan, *Depth-First Search and Linear Graph Algorithms*, in SIAM Journal on Computing 1 (2): 146–160 (1972)

B) Giant Component.