WSC Assignment 1
Pritish Naik, 192422IT010
Feb 18th 2020

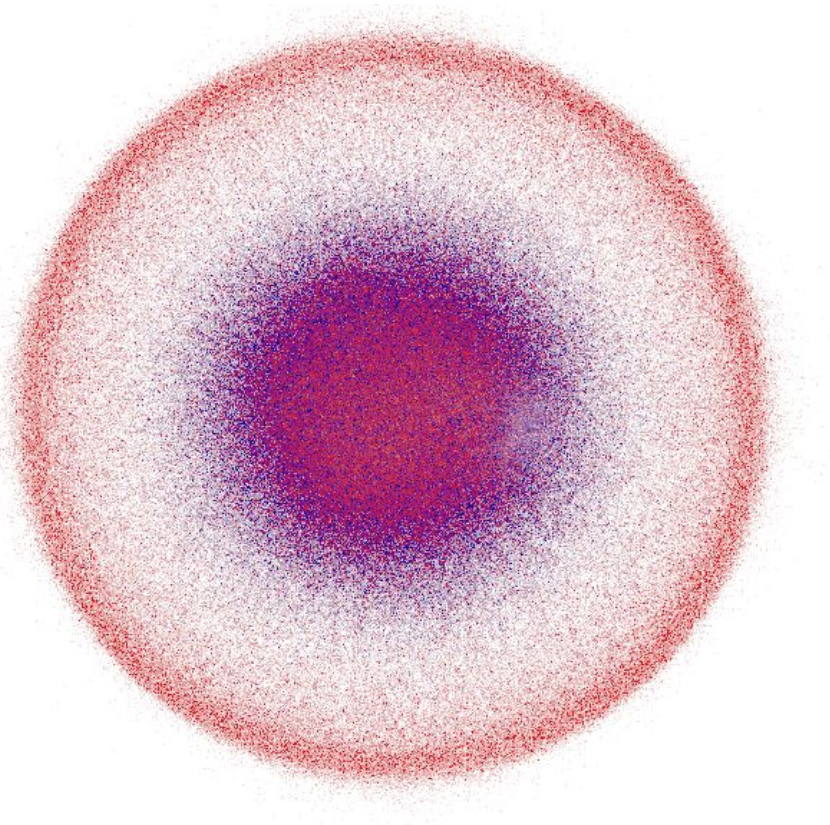
Large graph

Dataset:
DBLP collaboration network and ground-truth communities

Analysis:
Nodes: 425957
Edges: 333608
Type : Undirected Graph

The visualisation of DBLP graph was done using Forced Atlas 2 layout in Gephi. The Nodes are coloured based on degree and edges are coloured dark blue

Red - low
Yellow - medium
Blue - High degree



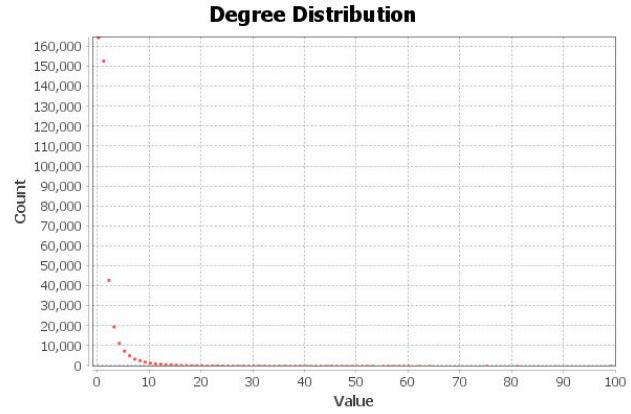
Results:

Average degree : 0.783

Diameter: 45

Results:

Average Degree: 0.783

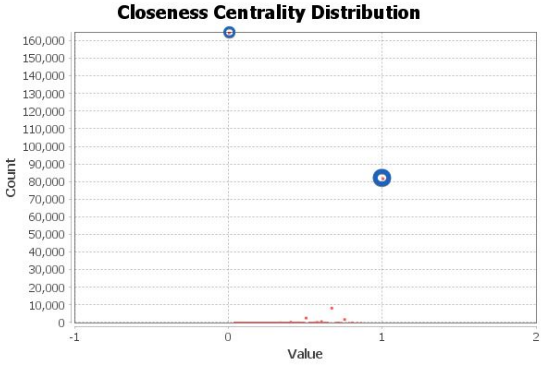


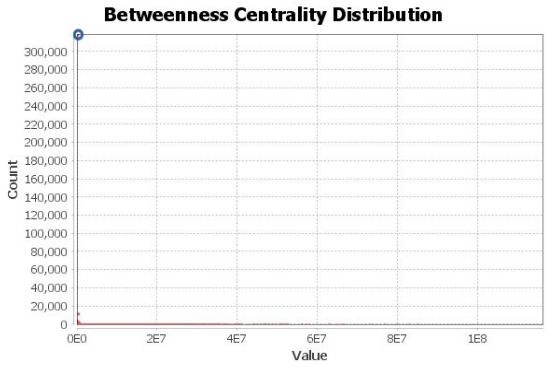
Betweenness centrality: number of shortest paths passing through a node. There exists one node which has high betweenness centrality.

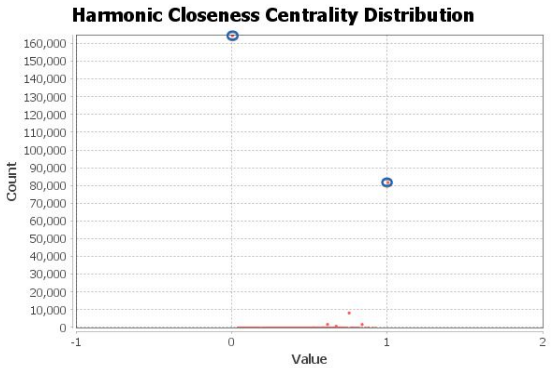
Closeness centrality: the avg. length of shortest paths from a node to all other webpages in the network.

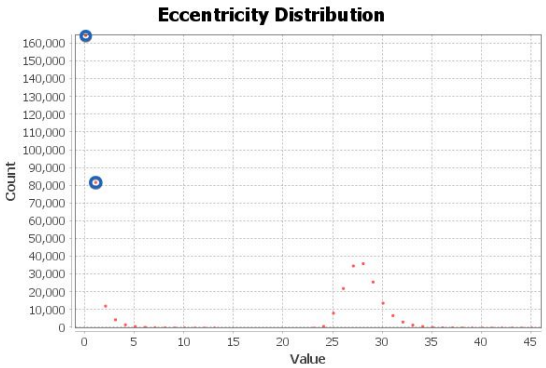
Harmonic centrality: reciprocal of closeness centrality of a node

Eigenvector centrality: measure of the influence of u in a network (like PageRank).





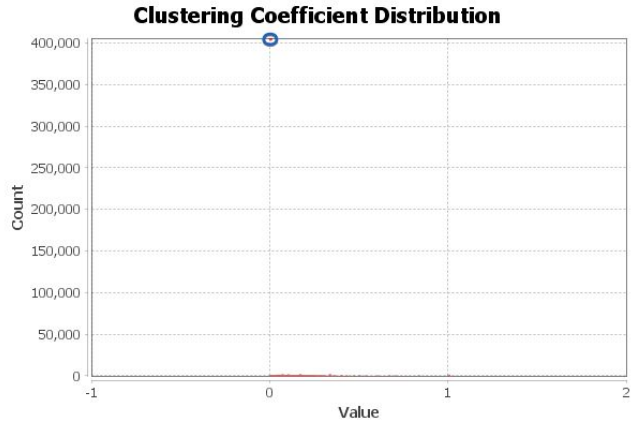




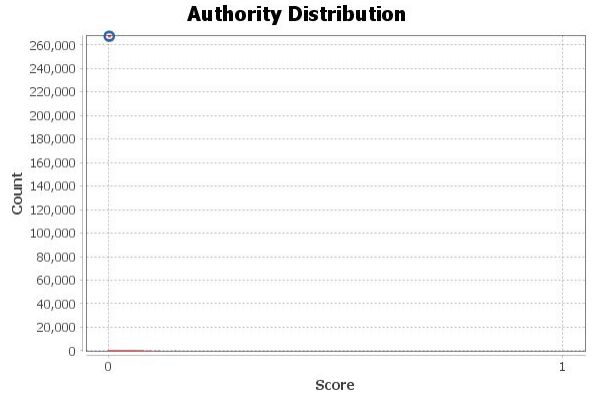
3. **Geodesic path length**
Average Path length: 110.036112770455

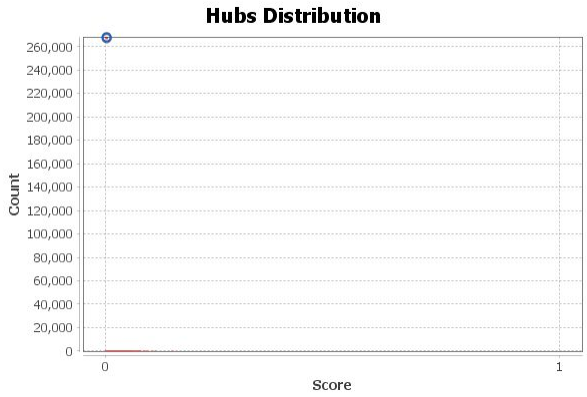
4. **Clustering coefficient & average clustering coefficient**

Average Clustering Coefficient: 0.026
Total triangles: 26707



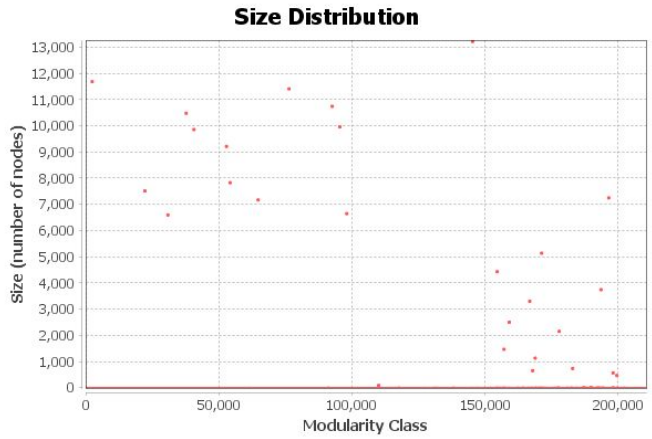
HITS Metric Report
Computes two separate values for each node.
Authority measures how valuable information stored at that node is.
Hub measures the quality of the nodes links.





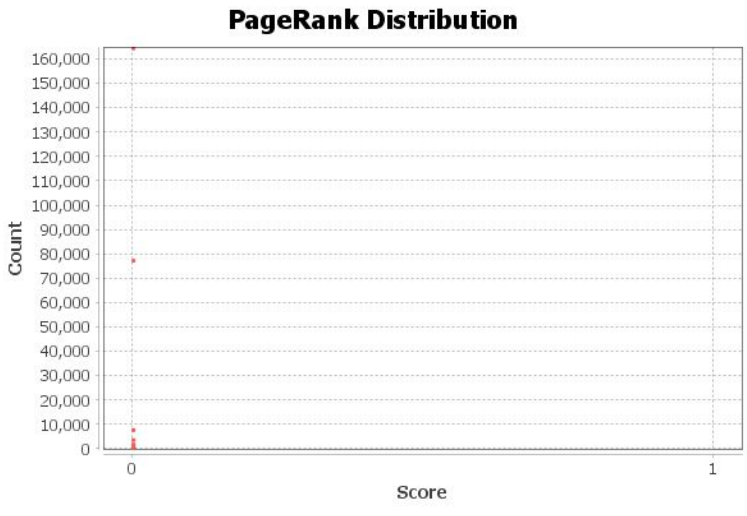
Modularity - Community detection algorithm.

Parameters:
Randomize: Off
Use edge weights: On
Resolution: 1.0
Results:
Modularity : 0.876
Modularity with resolution: 0.876
Number of communities: 210510



PageRank Report
Parameters:
Epsilon 0.001
probability = 0.85

Results:

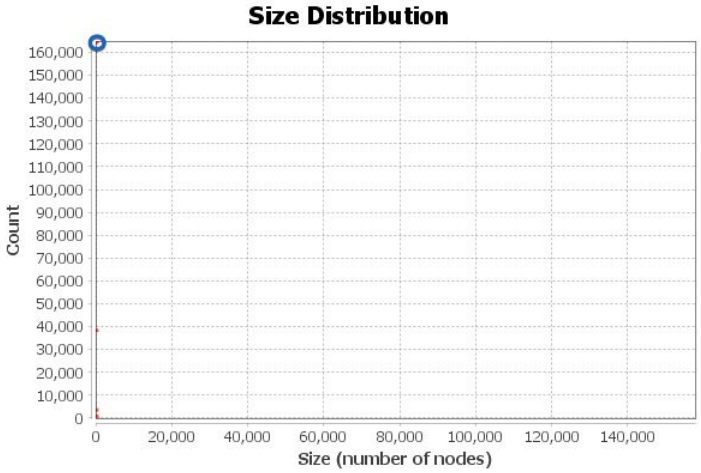


Connected components report

Parameters:

Network Interpretation: undirected

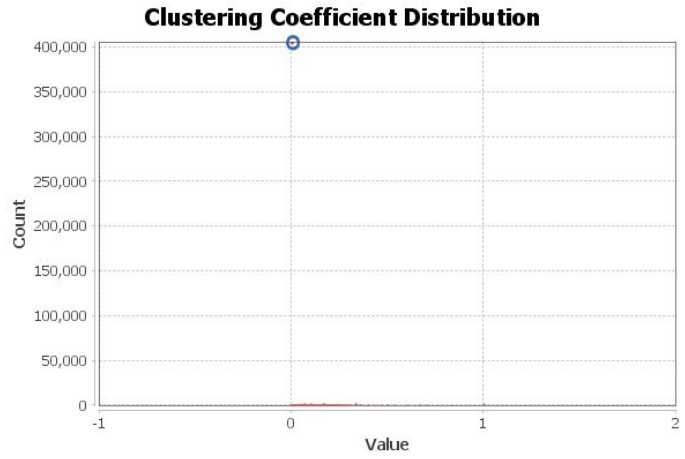
Number of weakly connected components : 210432



Clustering coefficient metrics report

Average clustering coefficient : 0.026

Total triangles: 26707

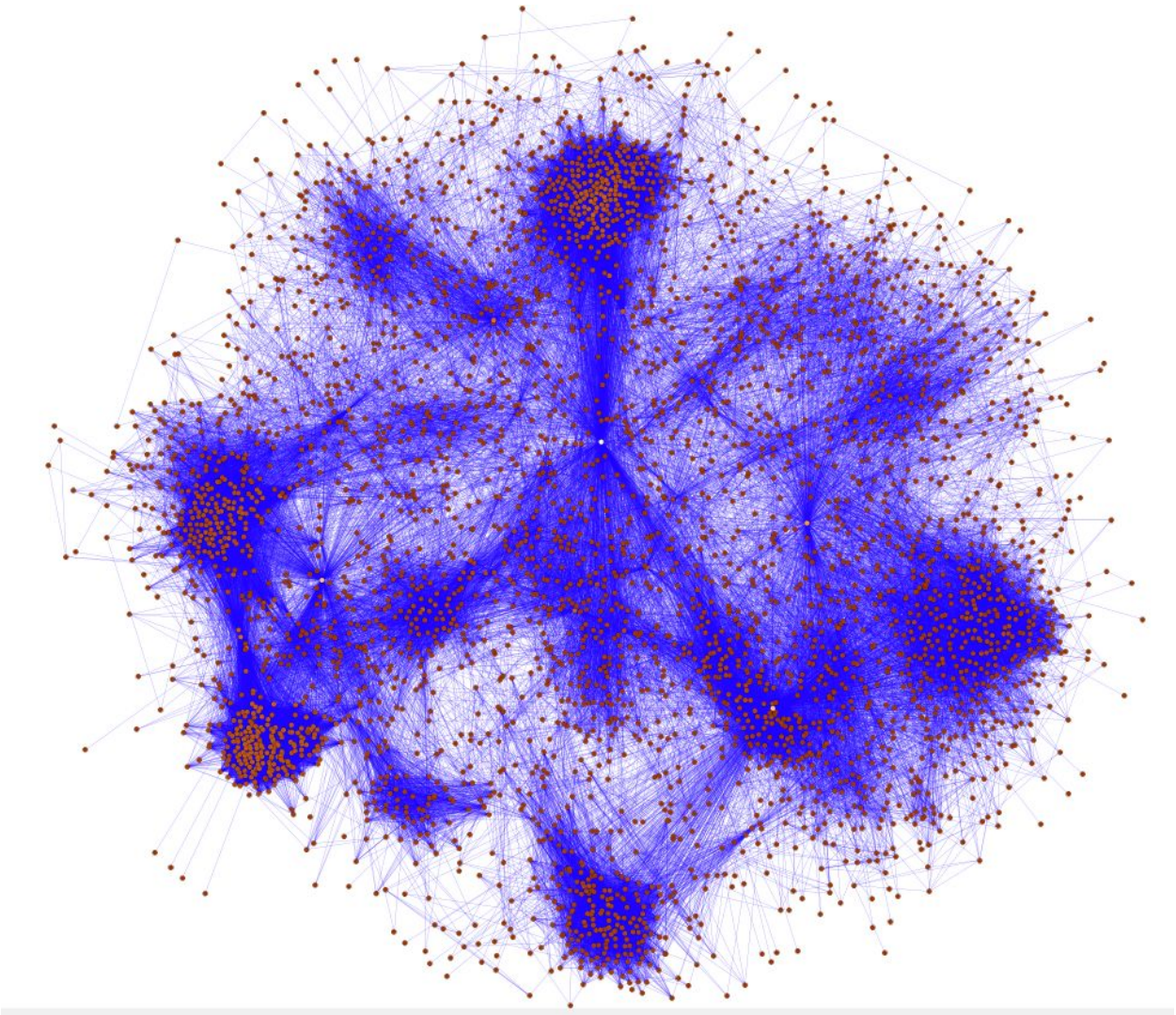


Medium Graph:

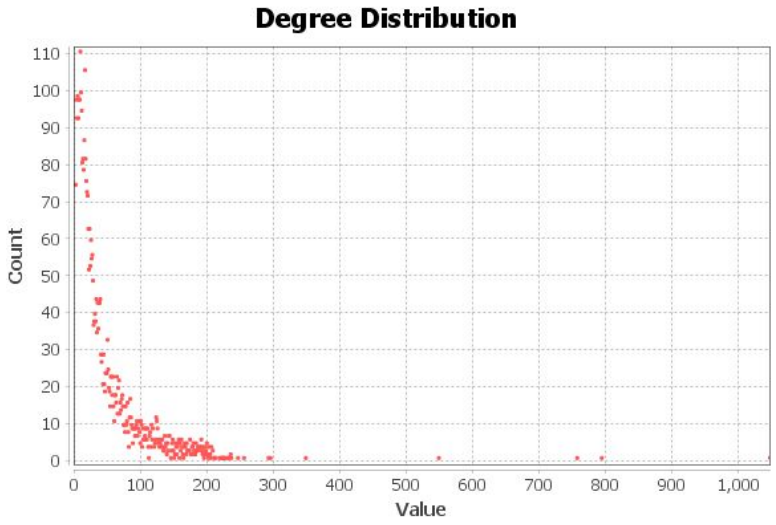
Facebook dataset
Number of Nodes: 4039
Number of Edges: 88234

The colour of the node is based on degree of the nodes
Darker the node, lesser is the degree

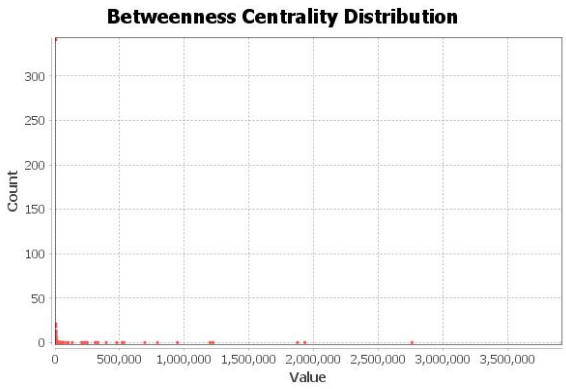
Force Atlas 2 visualisation of the graph

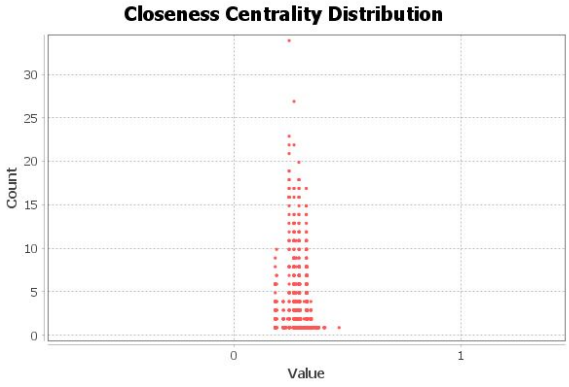


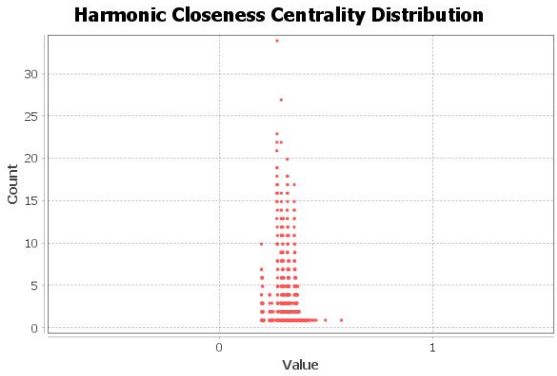
Degree Distribution
Average Degree: 43.691

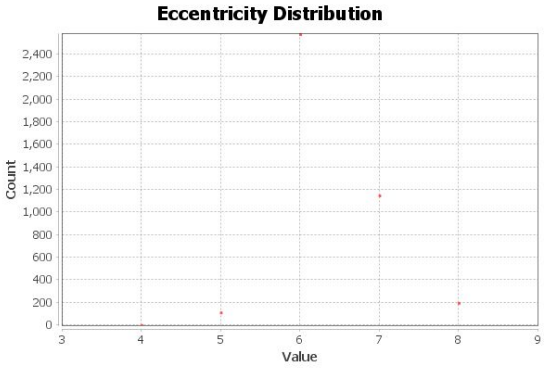


Graph Distance Report
Parameters:
Network Interpretation: undirected
Results:
Diameter: 8
Radius: 4
Average Path length: 3.6925068496963913

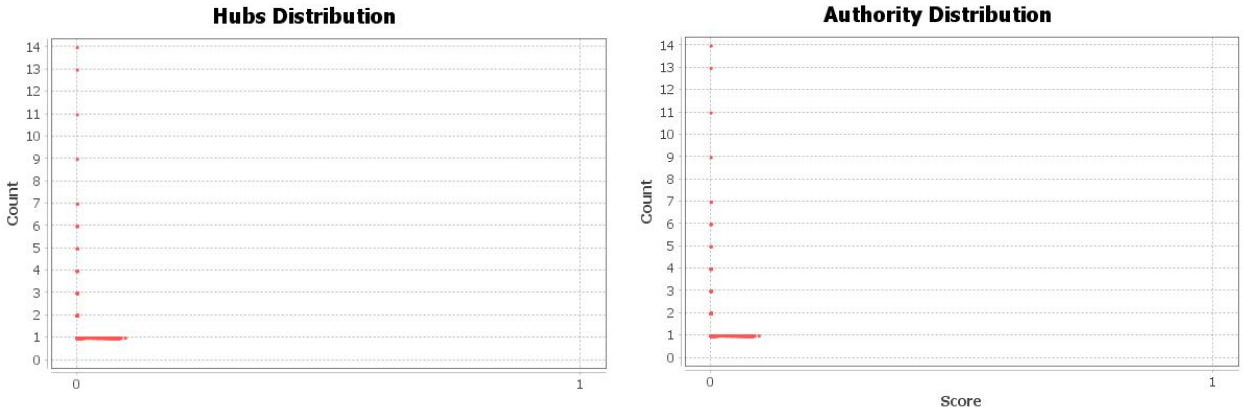




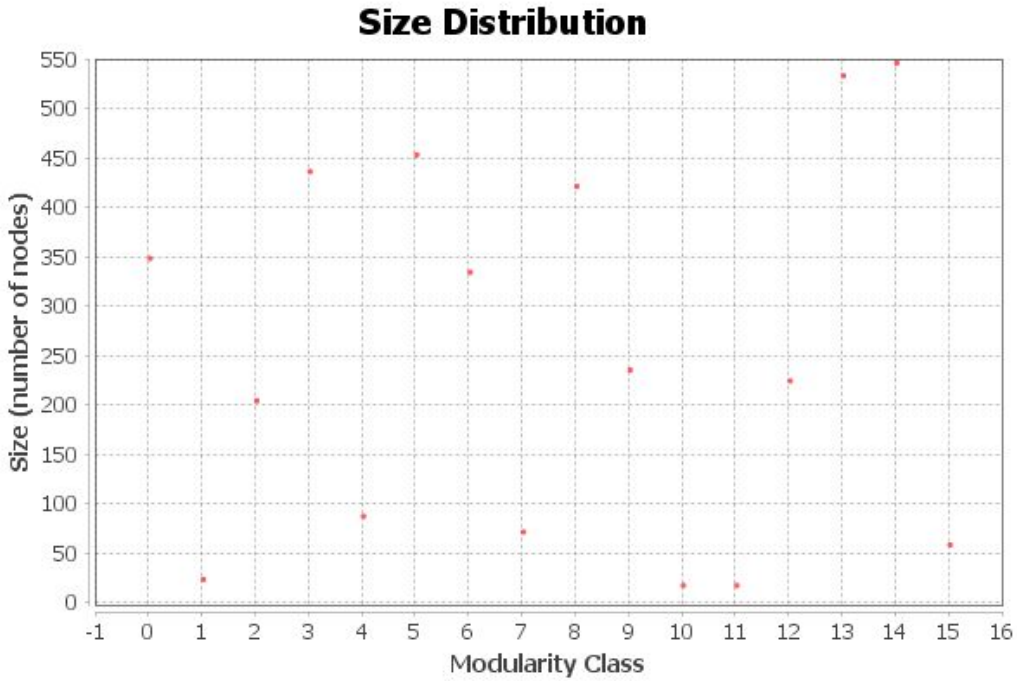




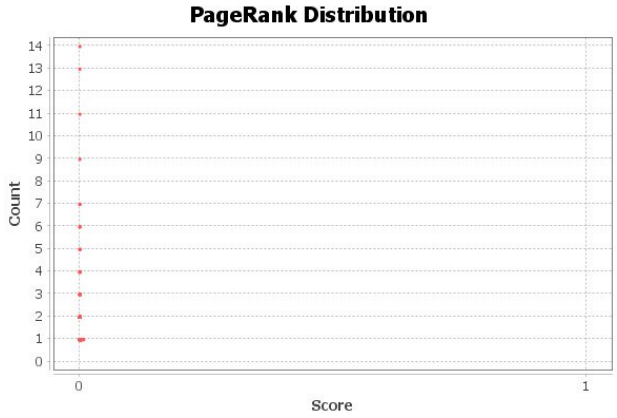
Graph density report
Density: 0.011
Parameters:
E = 1.0E-4
HITS Metric Report



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

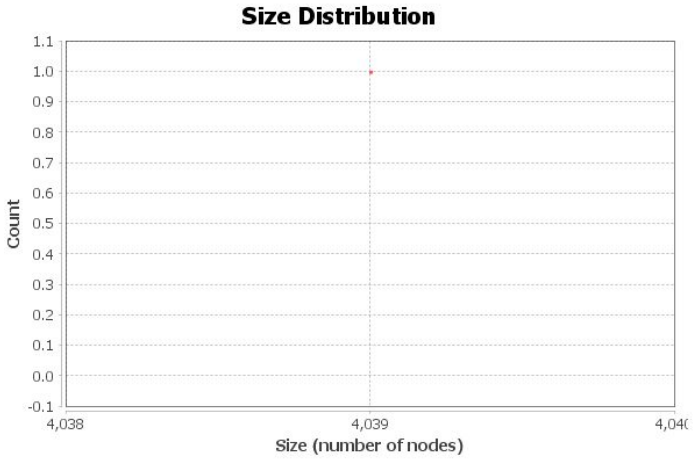


PageRank Report
Parameters:
Epsilon = 0.001
Probability = 0.85

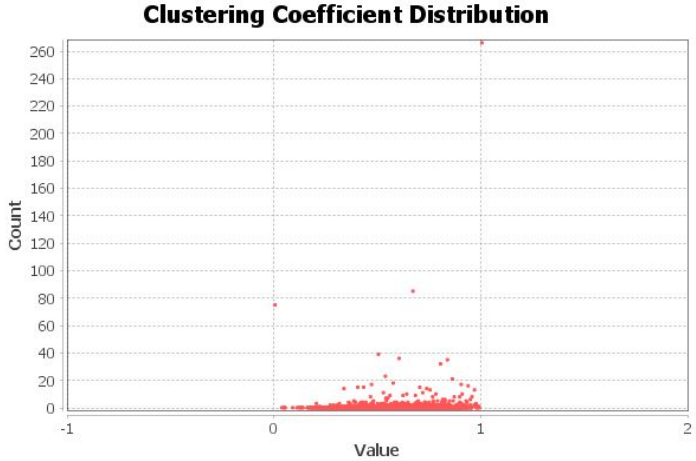


Connected components

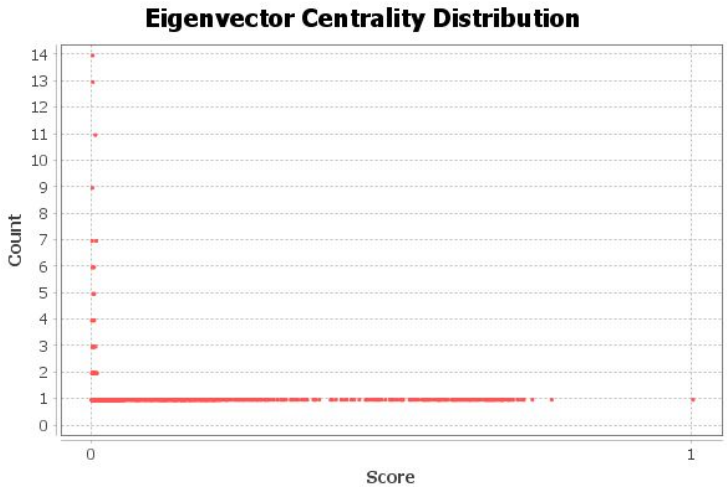
Number of Weakly Connected Components: 1



Average Clustering Coefficient: 0.617
Total triangles: 1612010

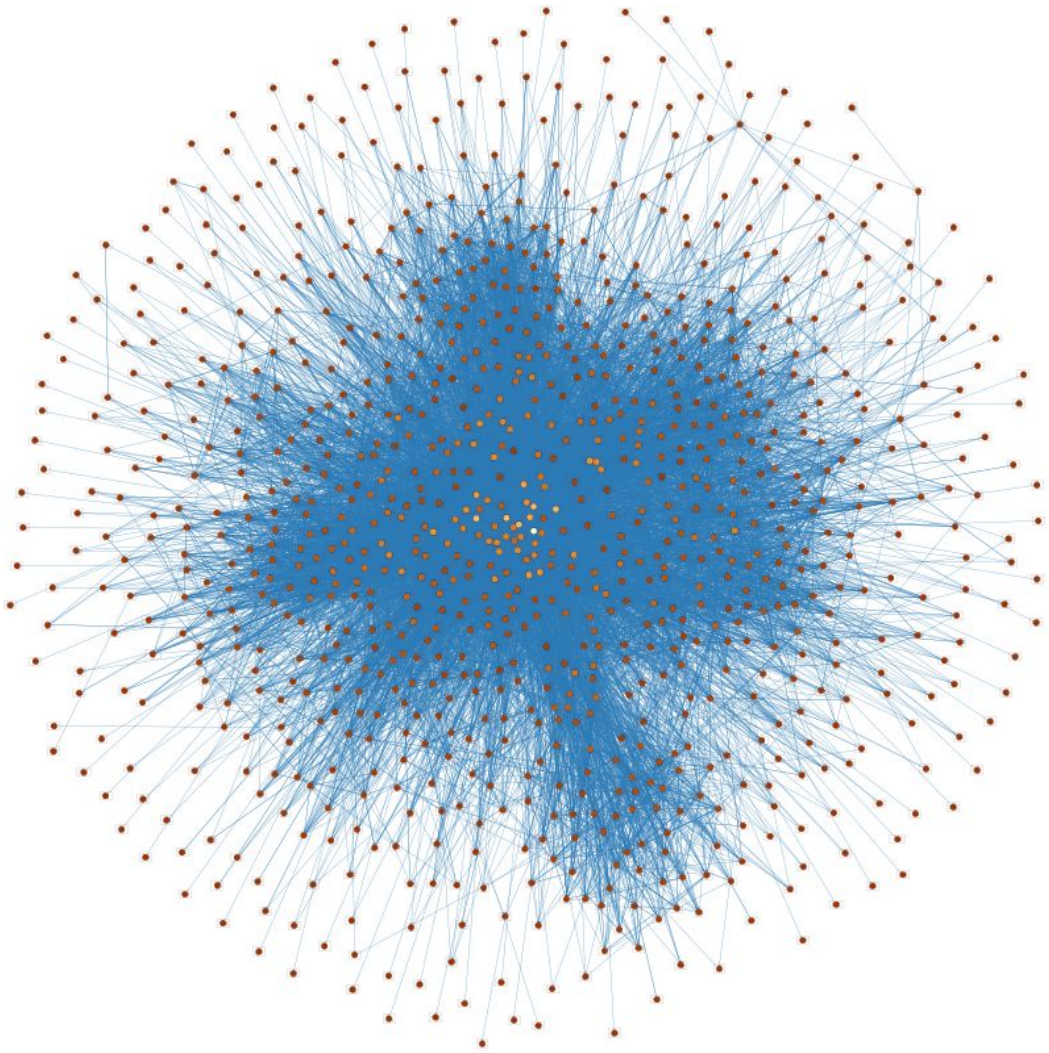


Eigenvector Centrality Report
Parameters:
Network Interpretation: undirected
Number of iterations: 100
Sum change: 0.4279073669804361



3. Small Dataset

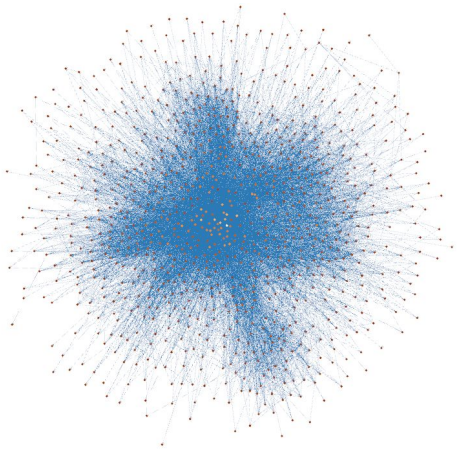
Email - CU dataset



Analysis:
Nodes: 1005
Edges: 25571
Directed graph

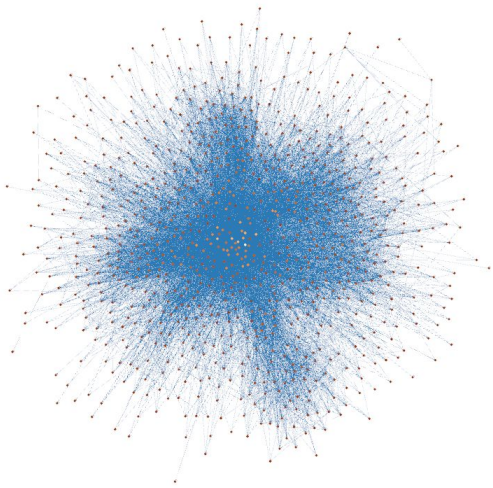
The colour of the node is based on the degree of the nodes. Darker the node, lesser is the degree

K = 1



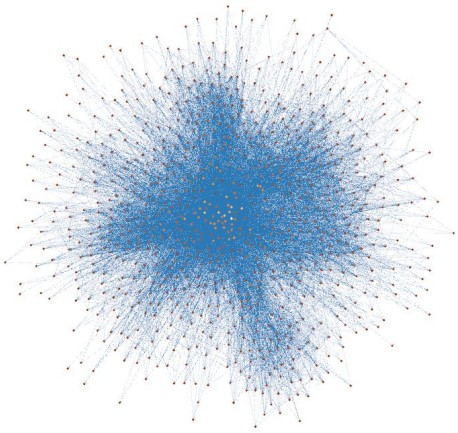
Nodes: 934 (92.94% visible)
Edges: 25500 (99.72% visible)
Directed Graph

K= 2



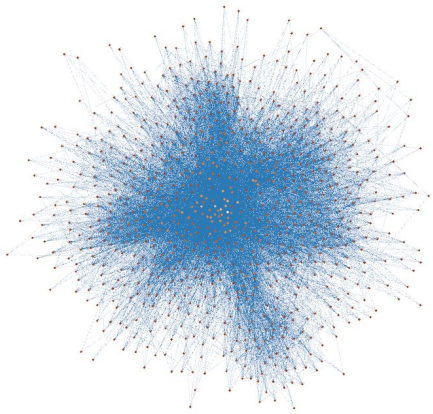
Nodes: 886 (88.16% visible)
Edges: 25423 (99.42% visible)
Directed Graph

K = 3



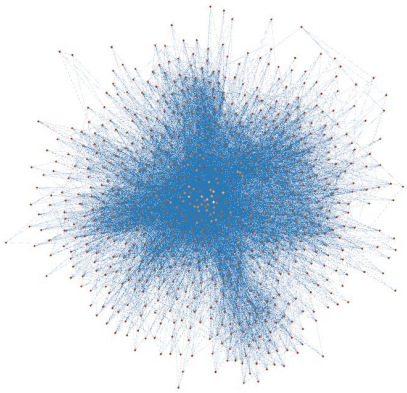
Nodes: 853 (84.88% visible)
Edges: 25335 (99.08% visible)

K = 4



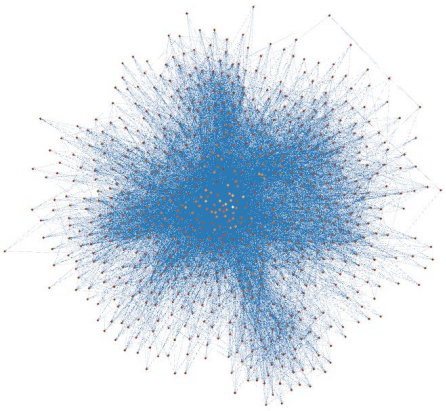
Nodes: 821 (81.69% visible)
Edges: 25218 (98.62% visible)

K = 5



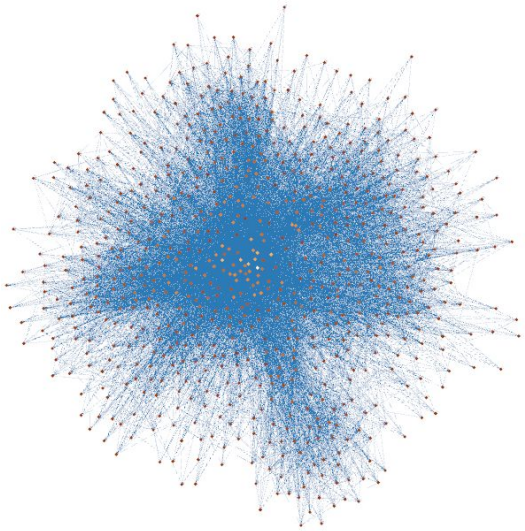
Nodes: 802 (79.8% visible)
Edges: 25131 (98.28% visible)

K = 6



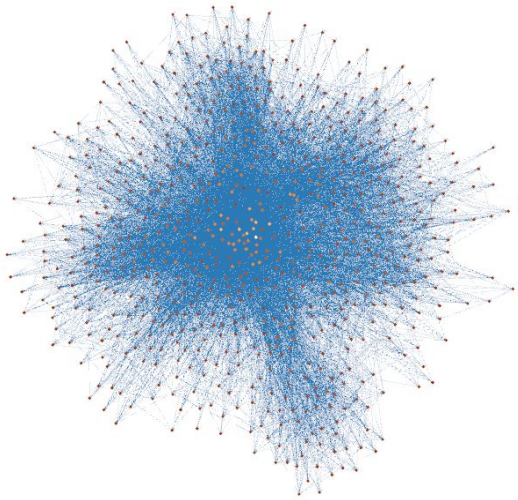
Nodes: 779 (77.51% visible)
Edges: 25004 (97.78% visible)

K = 7



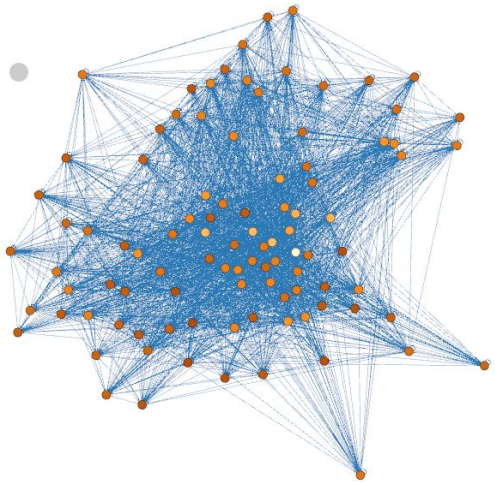
Nodes: 751 (74.73% visible)
Edges: 24819 (97.06% visible)

K = 8



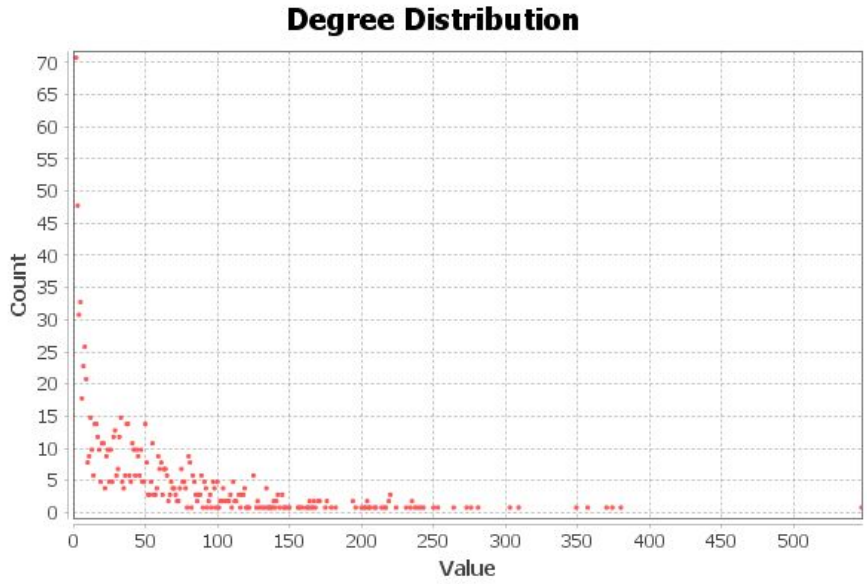
Nodes: 732 (72.84% visible)
Edges: 24675 (96.5% visible)

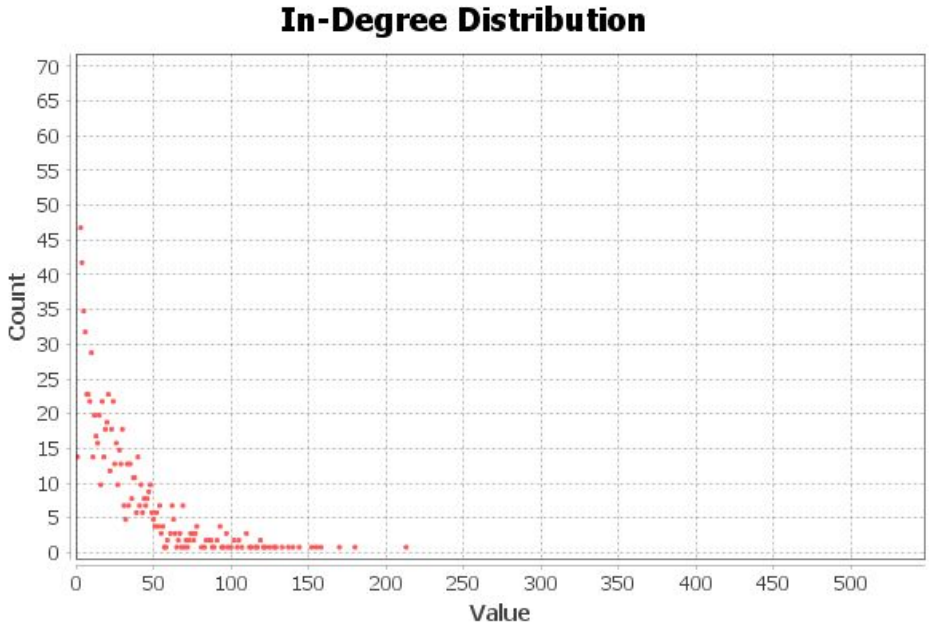
K = 49

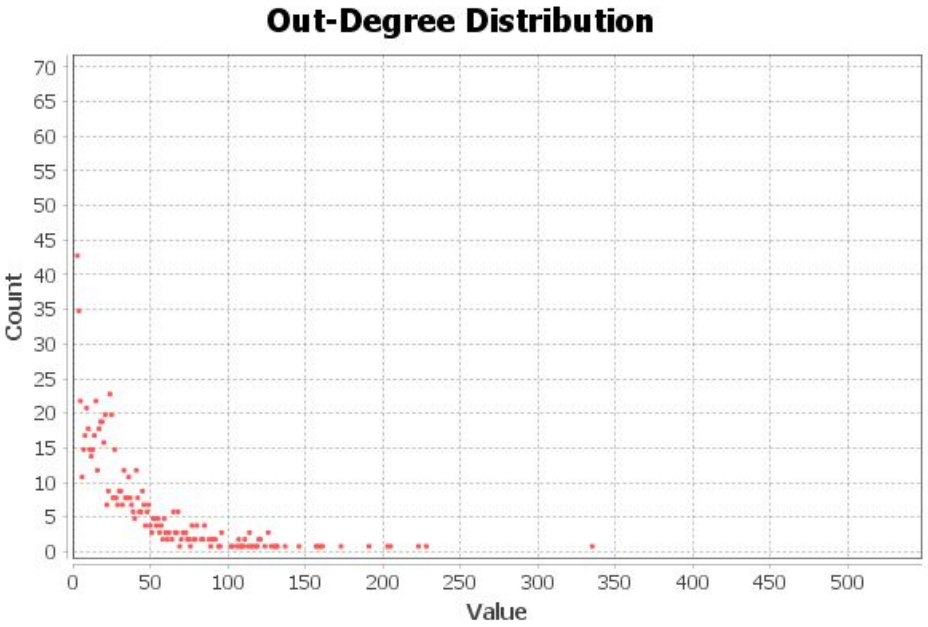


Here White nodes are the one with highest degree and red nodes with lowest degree

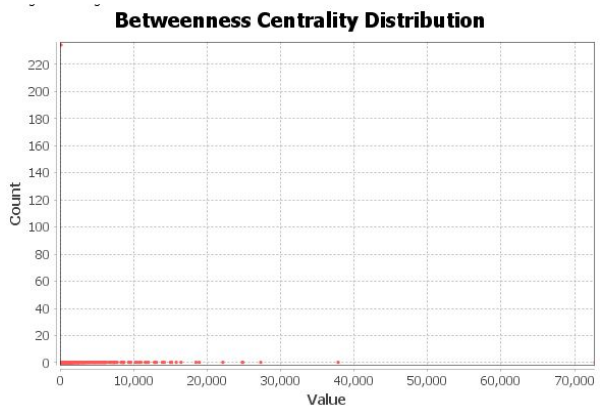
1. Degree distribution
Average Degree: 25.444

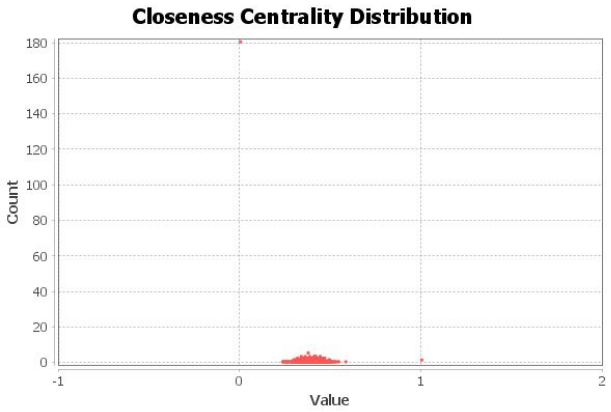


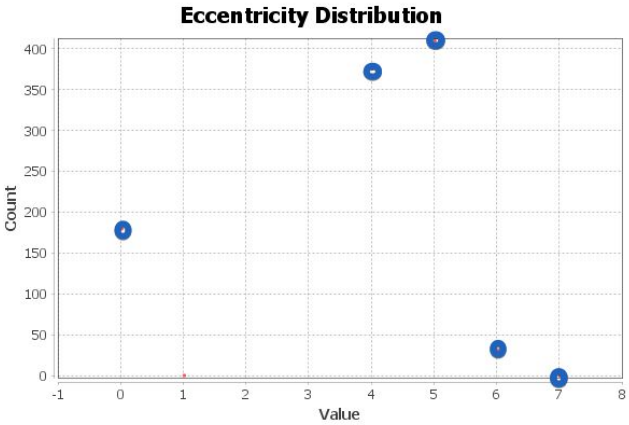


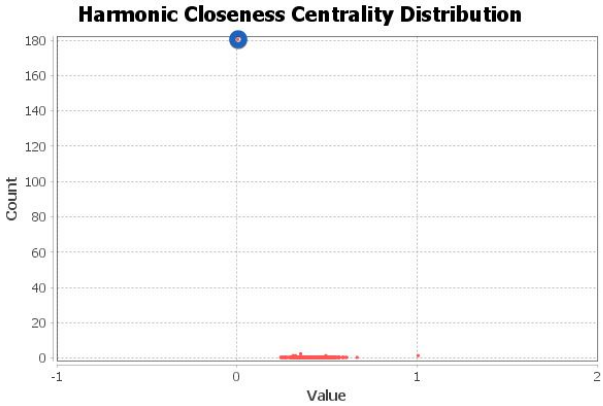


Diameter: 7
Radius: 0
Average Path length: 2.6528193693062723


Betweenness Centrality Distribution
Count
Value
0
20
40
60
80
100
120
140
160
180
200
220
0
10,000
20,000
30,000
40,000
50,000
60,000
70,000

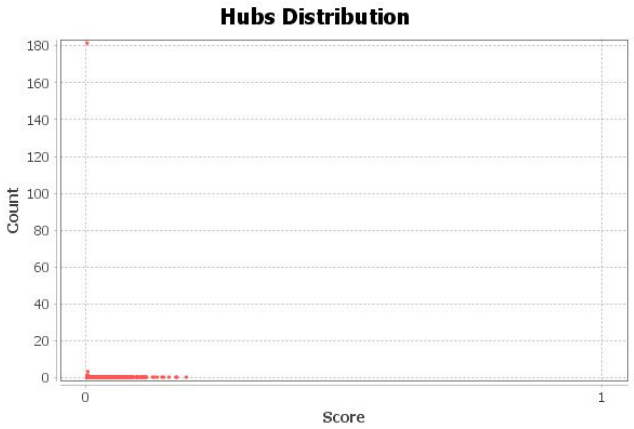

Closeness Centrality Distribution
Count
Value
0
20
40
60
80
100
120
140
160
180
-1
0
1
2

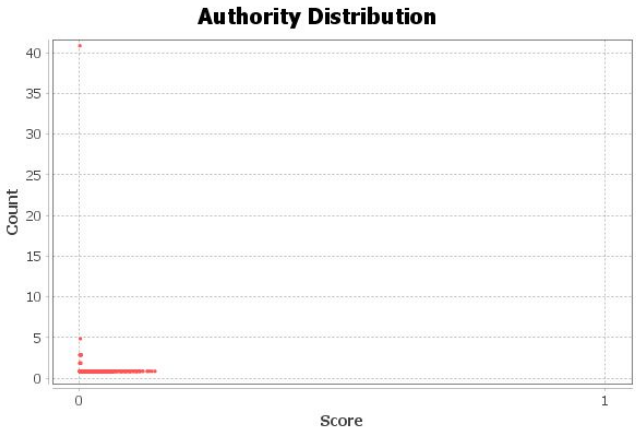

Eccentricity Distribution
Count
Value
0
50
100
150
200
250
300
350
400
-1
0
1
2
3
4
5
6
7
8


Harmonic Closeness Centrality Distribution
Count
Value
0
20
40
60
80
100
120
140
160
180
-1
0
1
2

Graph Density Report
Density: 0.025

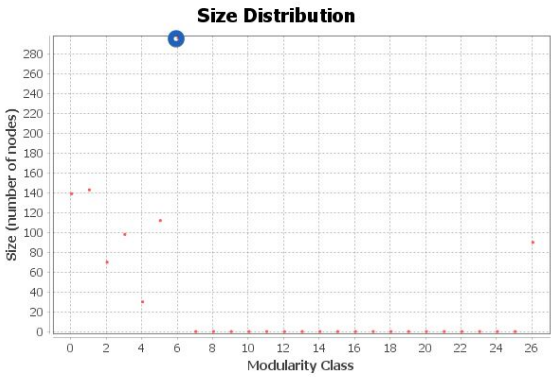
Parameters:
E = 1.0E-4



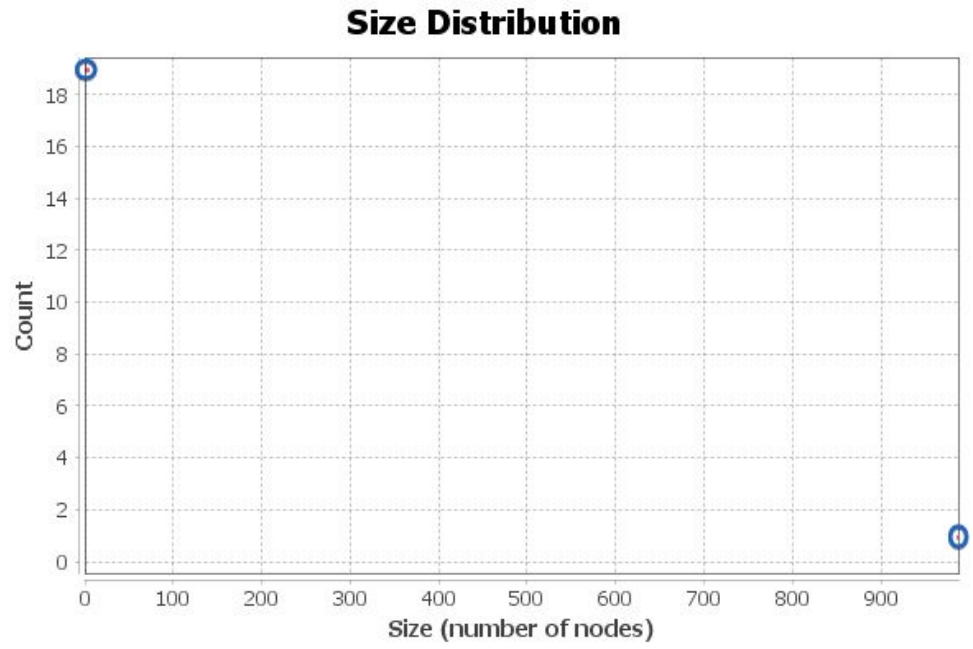


Parameters:
Randomize: On
Use edge weights: On
Resolution: 1.0

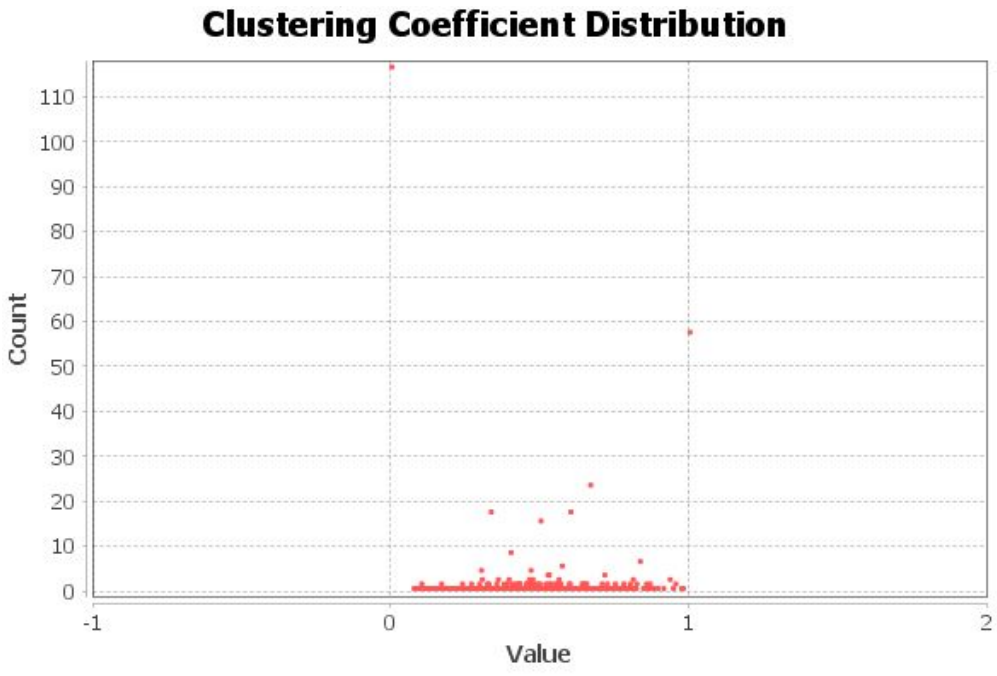
Results:
Modularity: 0.411
Modularity with resolution: 0.411
Number of Communities: 27



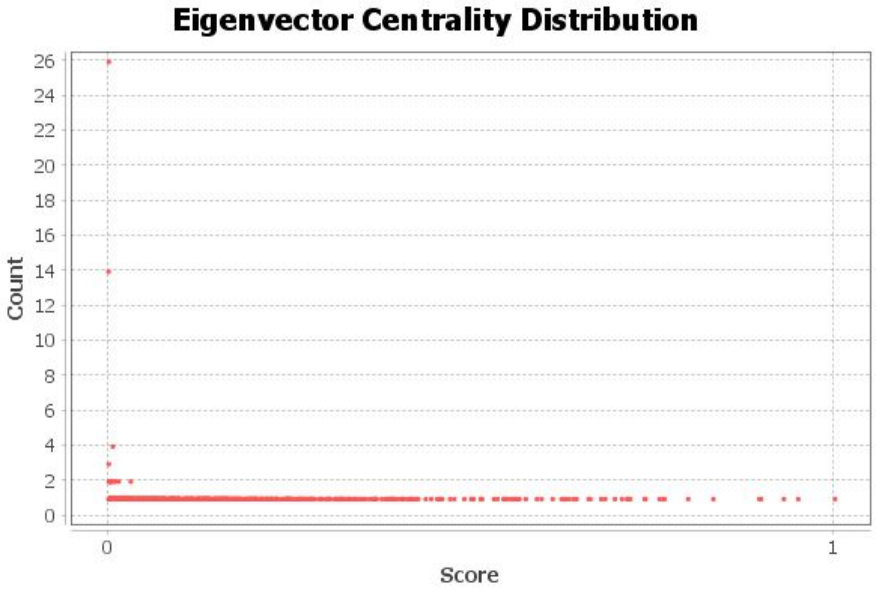
Number of Weakly Connected Components: 20
Number of Strongly Connected Components: 203



Clustering coefficient:
Network Interpretation: directed
Average Clustering Coefficient: 0.372
Network Interpretation: undirected
Average Clustering Coefficient: 0.473
Total triangles: 111711

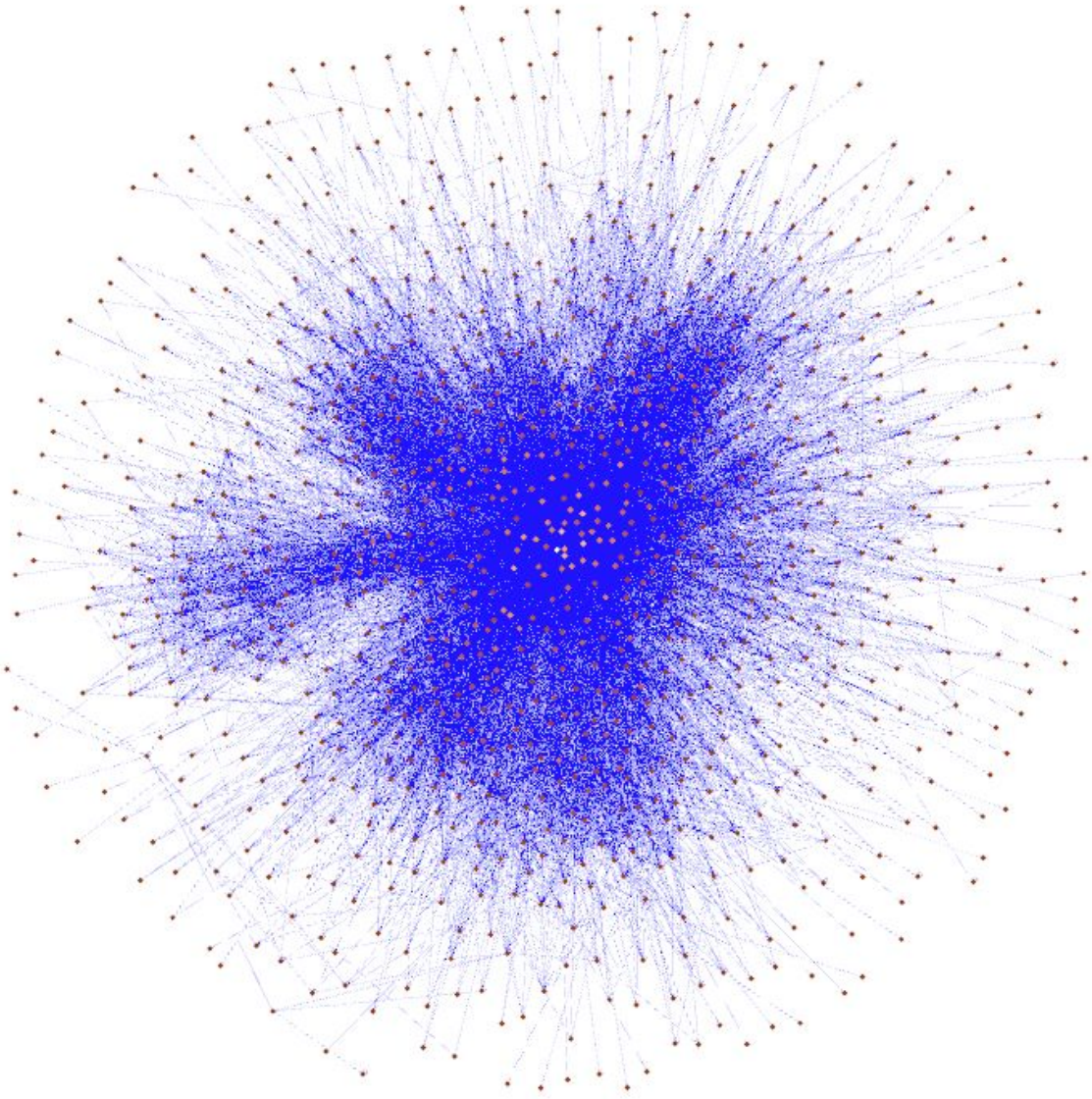


Eigenvector Centrality Report
Network Interpretation: directed
Number of iterations: 100
Sum change: 0.019065886249298433



Giant component: The 98% of the network is a giant component

Nodes: 986 (98.11% visible)
Edges: 25552 (99.93% visible)

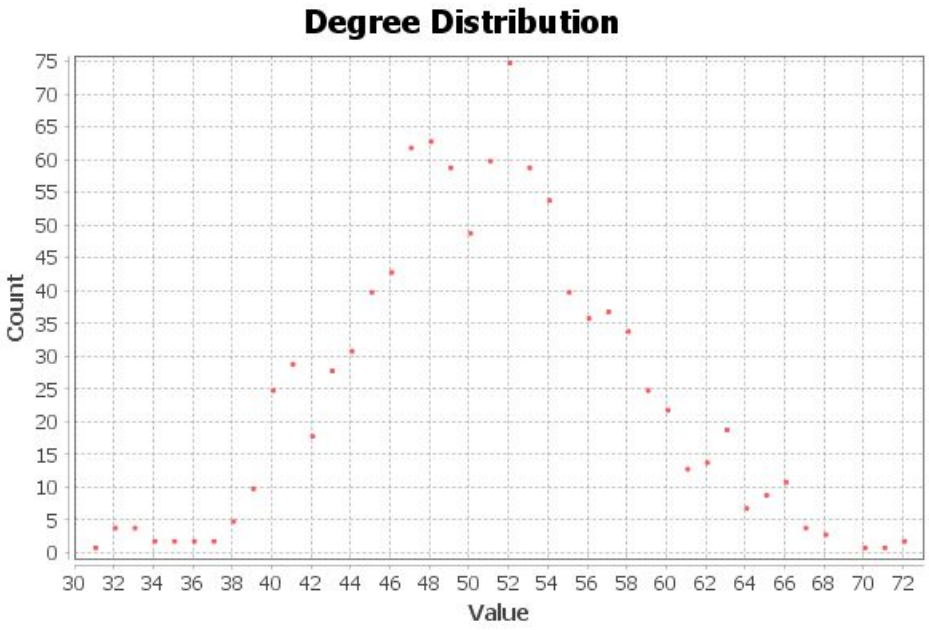


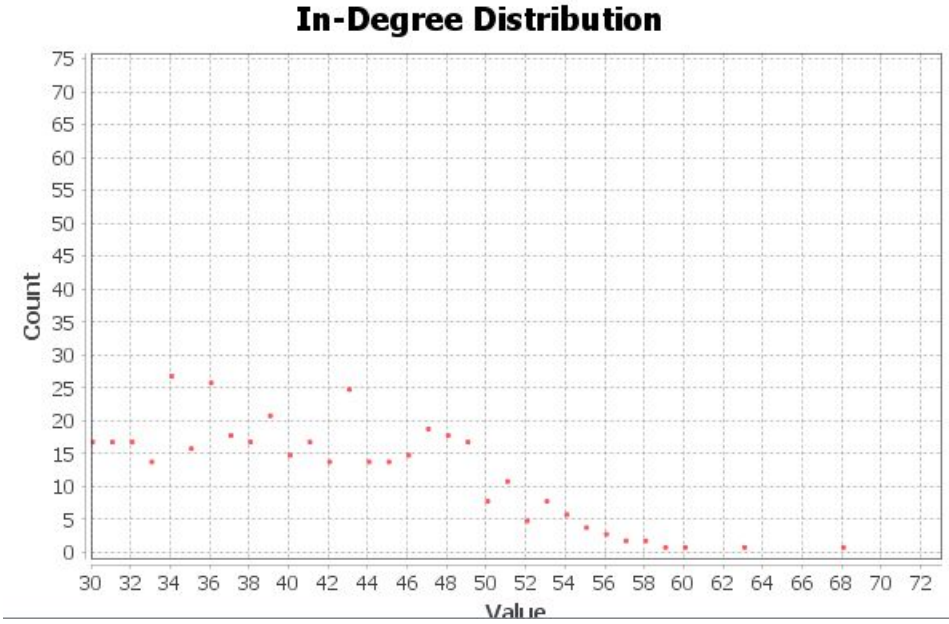
Part 2:

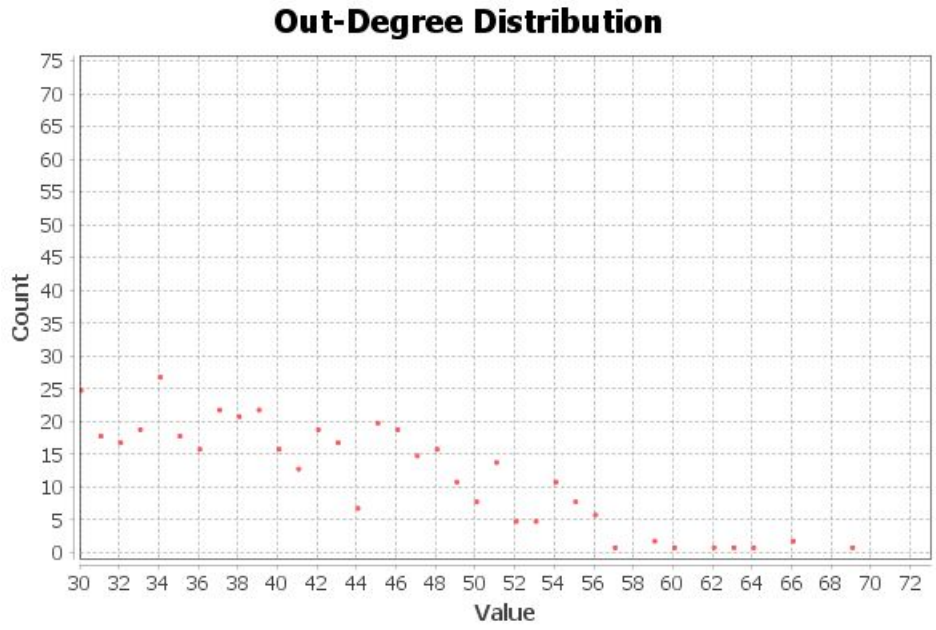
ER model for Email-CU dataset

Nodes: 1005
Edges: 25597
Directed Graph
value of p : 0.0509

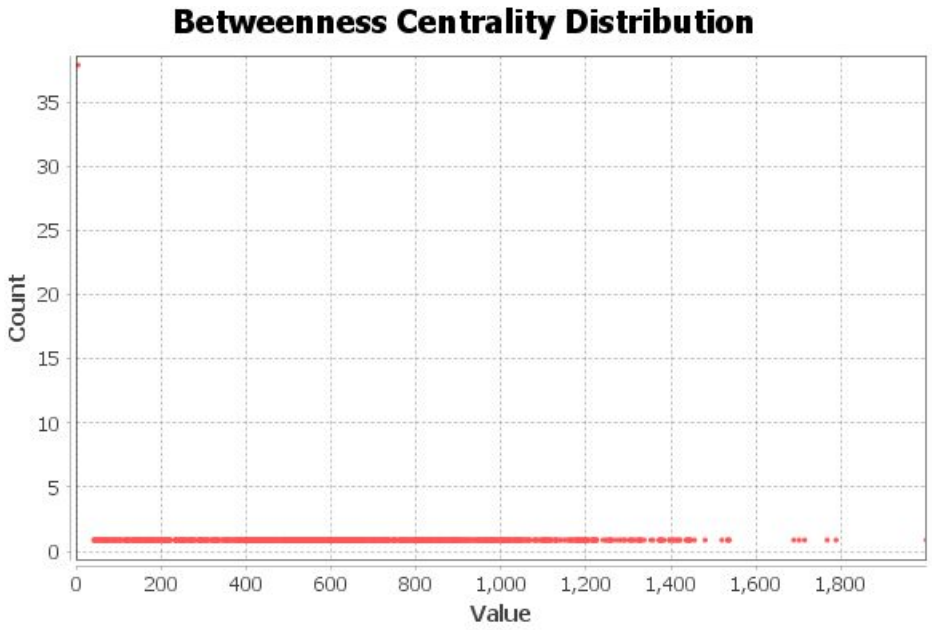
Average Degree: 25.470



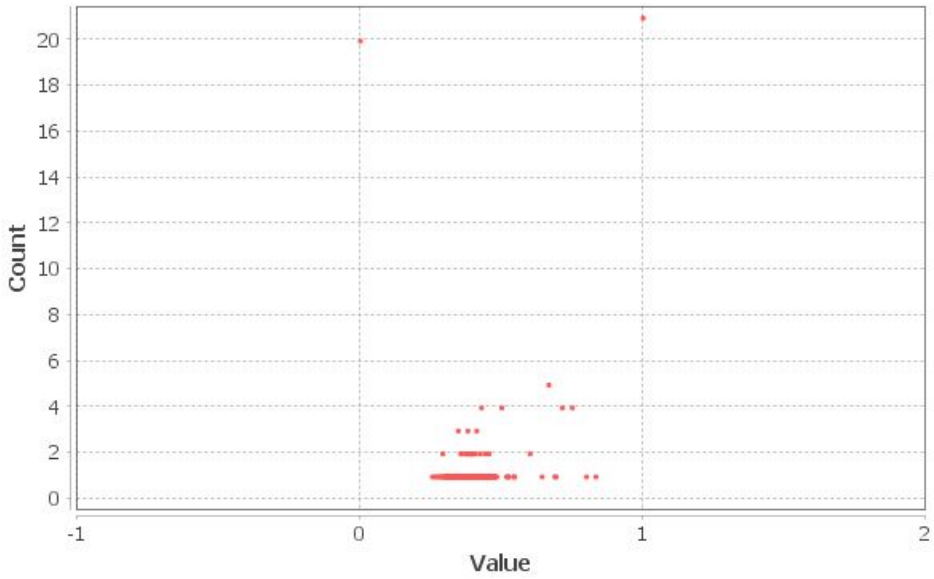




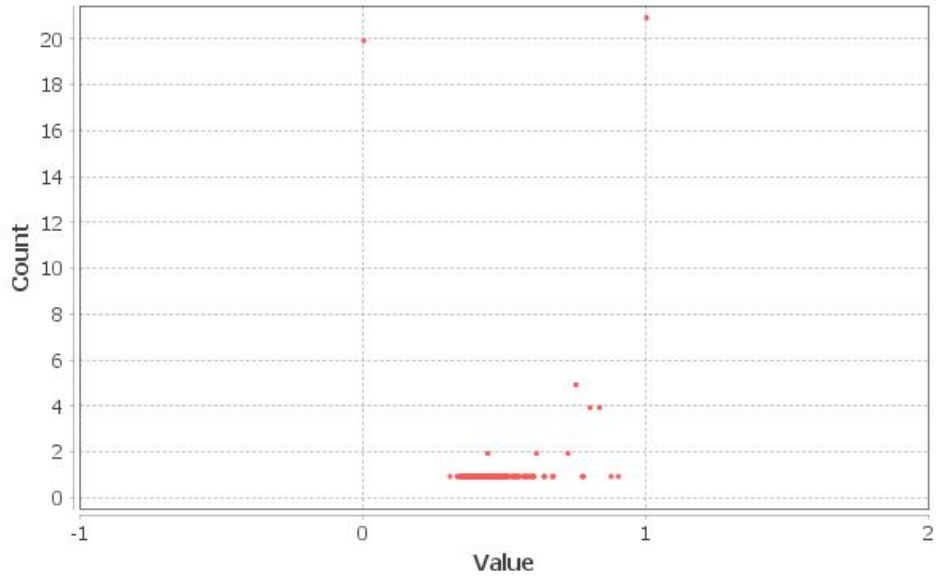
Diameter: 12
Radius: 0
Average Path length: 2.433669099217802

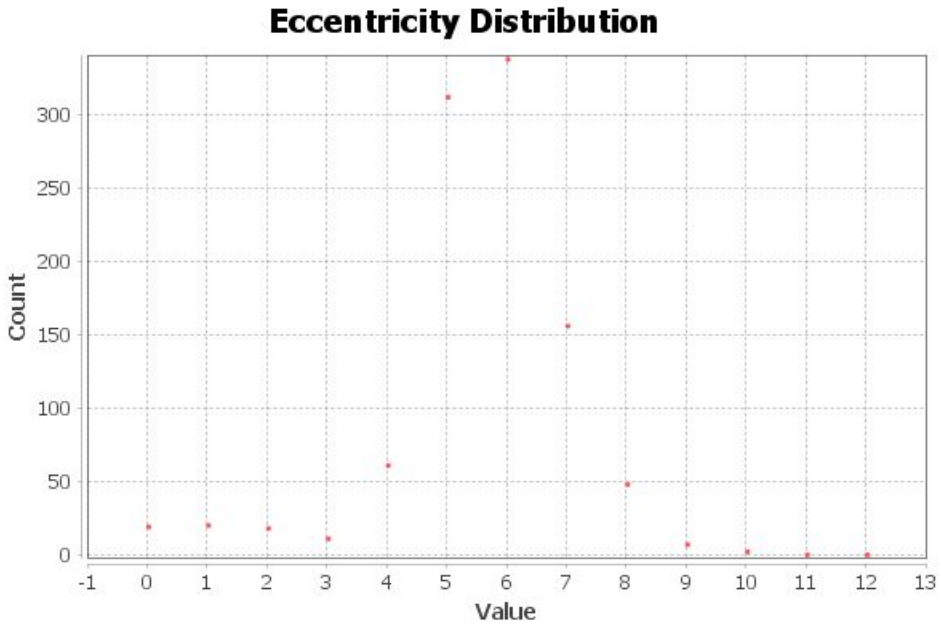


Closeness Centrality Distribution

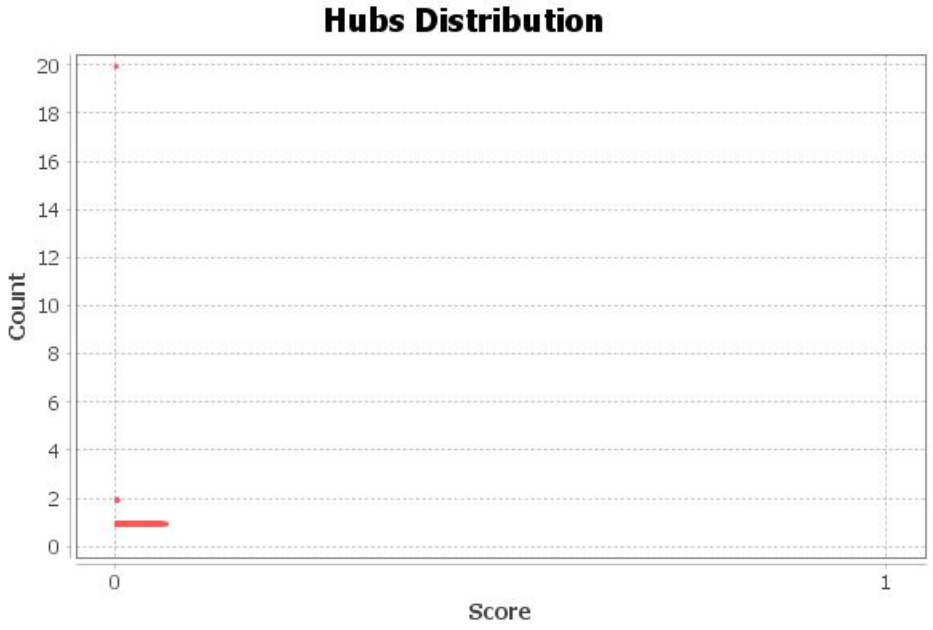


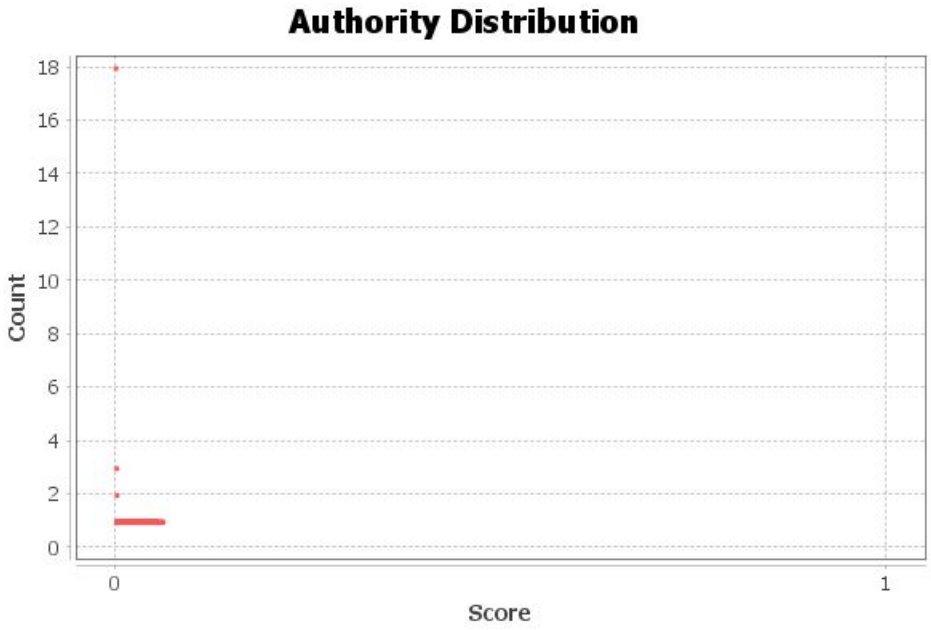
Harmonic Closeness Centrality Distribution





HITS Metric Report




Authority Distribution
18
16
14
12
10
8
6
4
2
0
Count
0
1
Score