GRAPH ANALYTICS

PROJECT 2 – ANALYZING LARGE NETWORKS

Guided by,

Dr: Zhong-Hui Duan

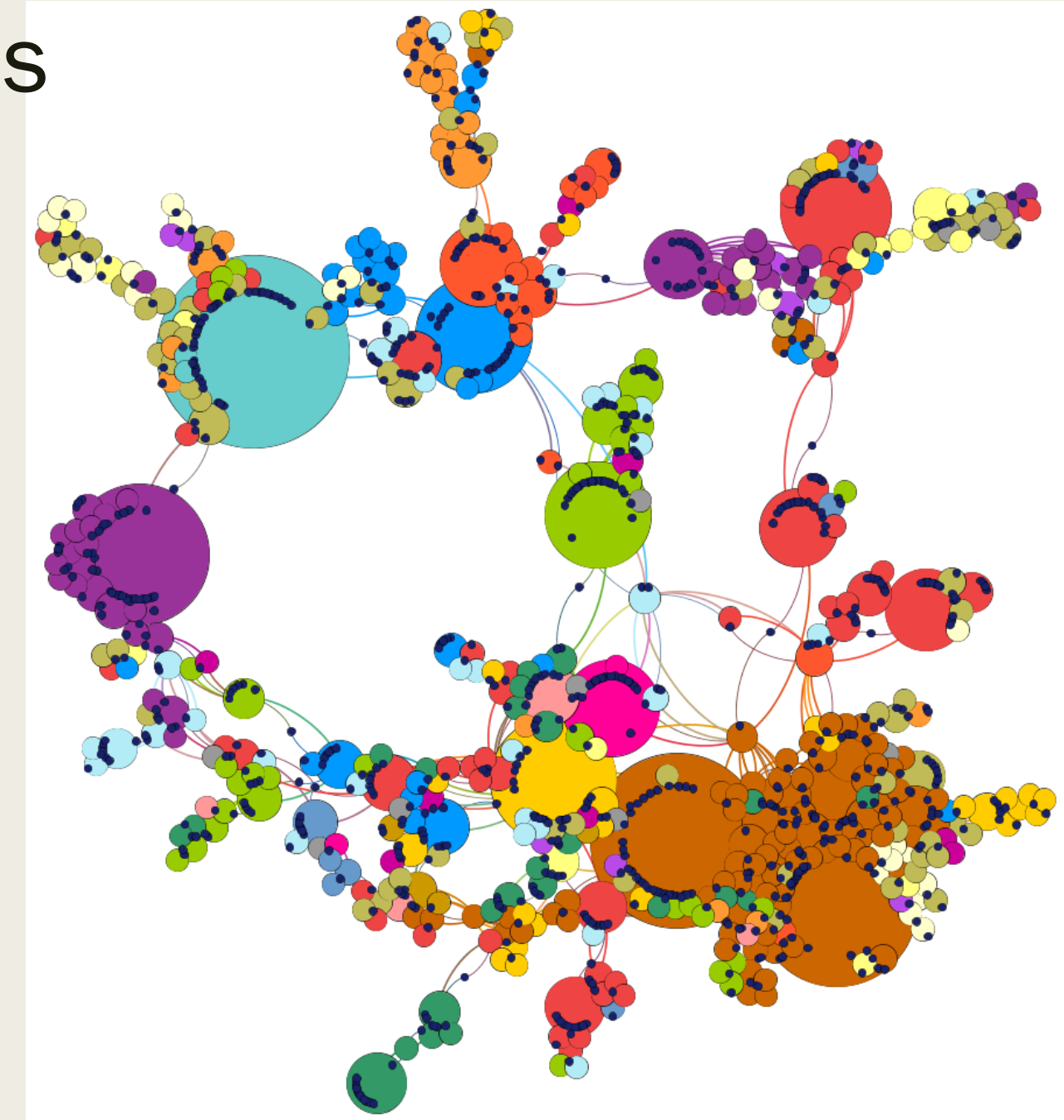
Team Members,

Bini Elsa Paul

Olaa Kasem

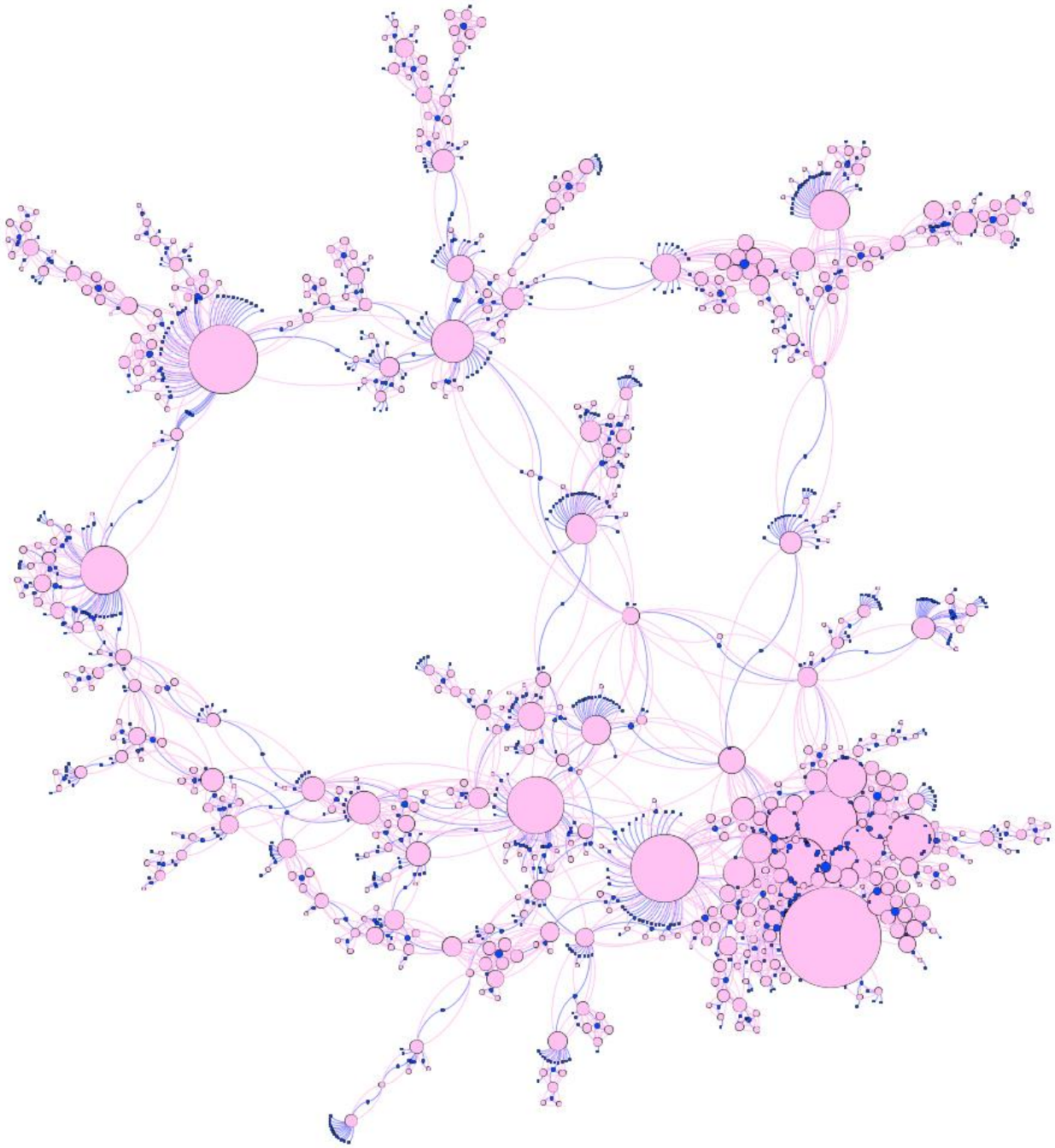
Materials and Methods

- Biological networks : - Diseasome
  - *A network of disorders and disease genes linked by known disorder–gene associations, indicating the common genetic origin of many diseases.*
  - *Nodes – 1419*
  - *Edges – 3926*
  - *Directed graph*
- Gephi



Over View

- Bini
  - *Diseasome*
  - *Nodes – 1419*
  - *Edges – 3926*
  - *Directed graph*
- Olaa
  - *Diseasome with only diseases*
  - *Nodes – 516*
  - *Edges – 2376*
  - *Directed graph*
- Comparison



Properties

- Network Diameter :- 15 – shortest distance between two most distant nodes
- Average Degree = 4.605
- Graph density :- 0.009 – sparse graph
- Connected components
  - *Weakly connected component :- 1*
  - *Strongly connected component :- 1(every vertex is reachable from every other vertex)*
- Average Path Length :- 6.509 :- follows small world effect
- Average Clustering Coefficient :- 0.636(Genes have 0 and 1)

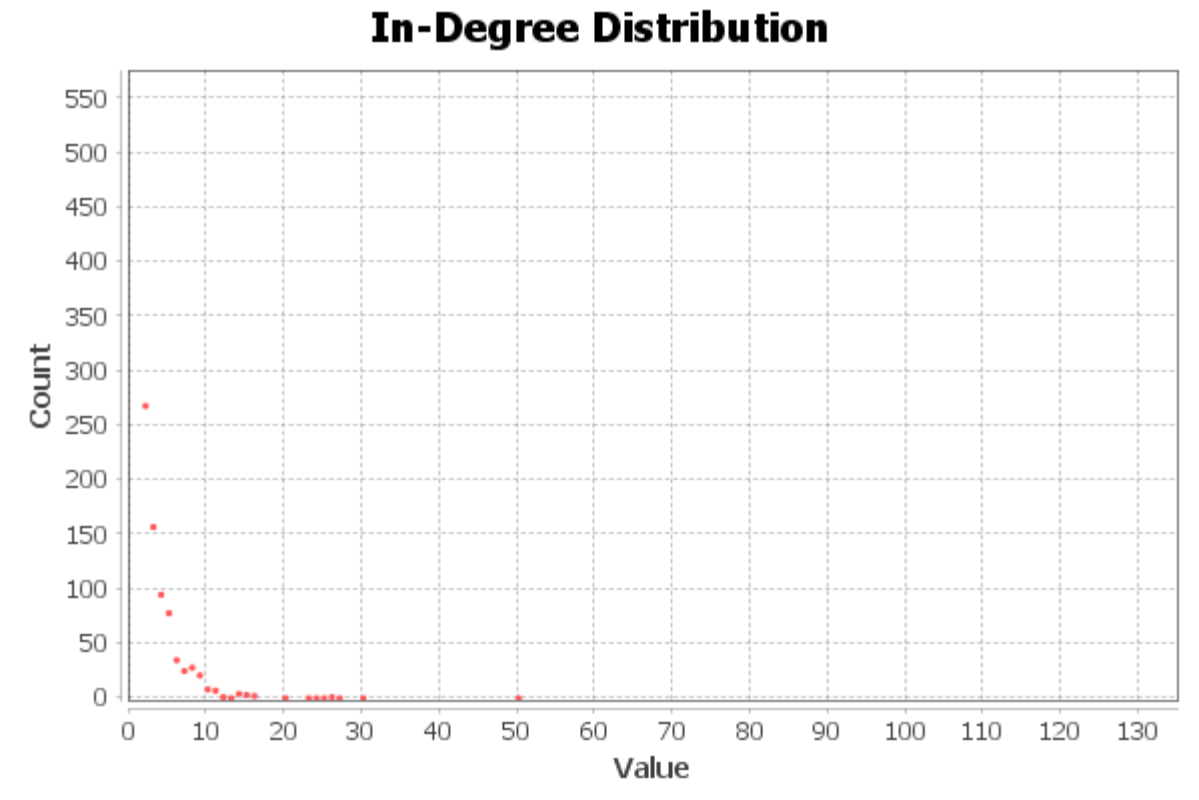
Nodes Edges | Configuration | Add node | Add edge | Search/Replace | Import Spreadsheet | Export table | More actions | Filter: | Id

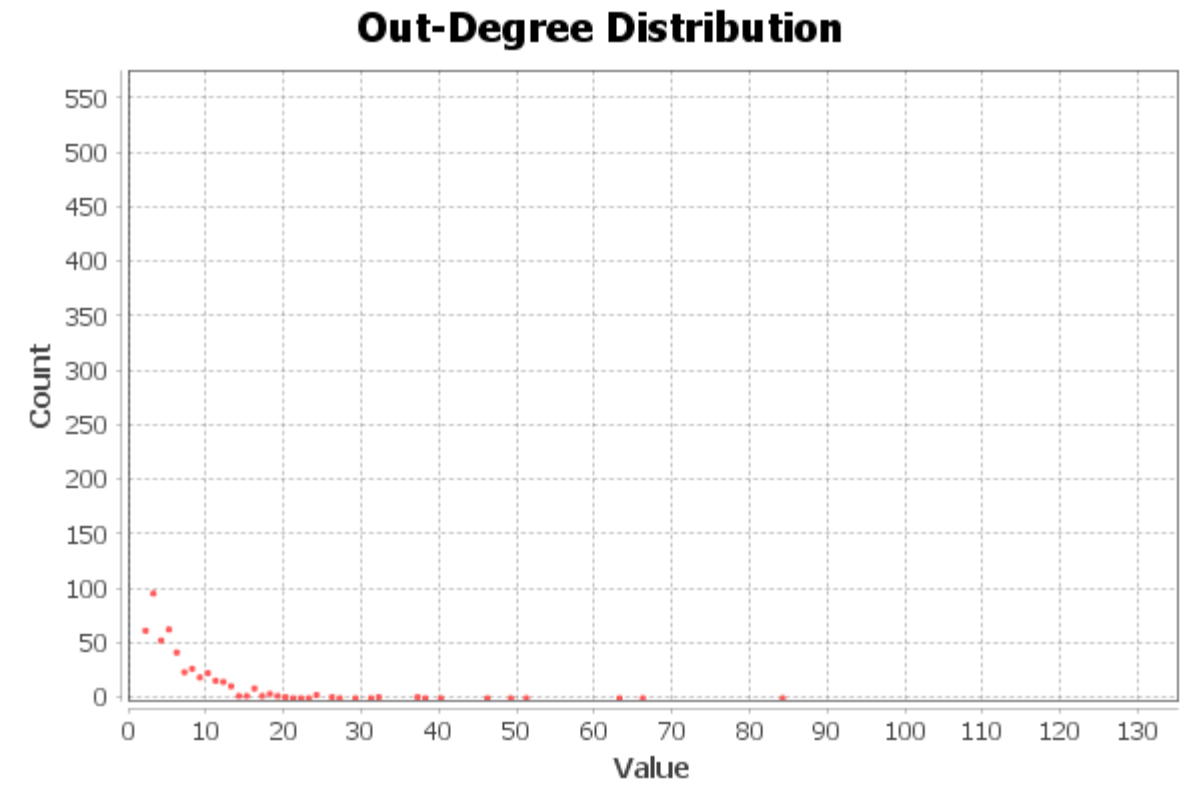
Id	Label	Interval	type	disclass	In-Degree	Out-Deg...	Degree	Weighted In-D...	Weighted Out-...	Weighted D...	Eccentri...	Closeness Cen...	Harmonic Closeness C...	Betweenness Ce...	Authority	Hub	Modularity ...	PageRank	Compone...
55	Deafness		disease	Ear,Nose,...	25	66	91	25.0	66.0	91.0	9.0	0.197081	0.249394	142497.426917	0.000134	0.00016	4	0.004068	0
47	Leukemia		disease	Cancer	26	63	89	26.0	63.0	89.0	11.0	0.204854	0.27537	92966.954187	0.120864	0.145771	2	0.003648	0
114	Colon can...		disease	Cancer	50	84	134	50.0	84.0	134.0	10.0	0.219777	0.296173	92802.124039	0.345284	0.457335	2	0.005352	0
45	Retinitis pi...		disease	Ophtham...	16	46	62	16.0	46.0	62.0	10.0	0.164711	0.208451	62231.515855	0.000006	0.000008	7	0.002105	0
87	Diabetes ...		disease	Endocrine	24	51	75	24.0	51.0	75.0	9.0	0.244272	0.296071	206496.991521	0.02994	0.036639	3	0.003211	0
54	Cardiomy...		disease	Cardiovas...	15	40	55	15.0	40.0	55.0	8.0	0.226771	0.270707	333925.143584	0.00134	0.001576	6	0.002053	0
81	Mental ret...		disease	Neurological	14	38	52	14.0	38.0	52.0	12.0	0.14434	0.174919	30921.0	0.000065	0.000081	8	0.002366	0
48	Blood group		disease	Hematolo...	8	31	39	8.0	31.0	39.0	12.0	0.16099	0.196024	50990.0	0.00009	0.000098	5	0.001268	0
70	Obesity		disease	Nutritional	8	29	37	8.0	29.0	37.0	9.0	0.228084	0.266276	57781.152201	0.01231	0.014349	3	0.001218	0
137	Breast ca...		disease	Cancer	30	49	79	30.0	49.0	79.0	10.0	0.2156	0.281731	62915.657697	0.271522	0.319507	2	0.002916	0
223	Muscular ...		disease	Muscular	8	26	34	8.0	26.0	34.0	9.0	0.191311	0.224788	70420.833333	0.000148	0.000183	6	0.001259	0
390	Charcot-...		disease	Neurological	5	23	28	5.0	23.0	28.0	10.0	0.179448	0.208352	18921.333333	0.003102	0.003658	6	0.000896	0
220	Epilepsy		disease	Neurological	5	22	27	5.0	22.0	27.0	10.0	0.197879	0.231512	64188.0	0.007342	0.008624	8	0.000945	0
325	Cataract		disease	Ophtham...	11	26	37	11.0	26.0	37.0	10.0	0.1655	0.196329	110322.0	0.000015	0.000016	1	0.001621	0
68	Asthma		disease	Respiratory	11	24	35	11.0	24.0	35.0	10.0	0.20524	0.240576	53815.756949	0.002192	0.002392	5	0.001709	0
208	Compleme...		disease	Immunolo...	1	14	15	1.0	14.0	15.0	14.0	0.116968	0.131164	6180.0	0.0	0.0	3	0.000527	0
213	Spinocere...		disease	Neurological	5	18	23	5.0	18.0	23.0	12.0	0.159973	0.186378	16740.0	0.000511	0.000517	5	0.000999	0
30	Alzheimer ...		disease	Neurological	15	27	42	15.0	27.0	42.0	10.0	0.217318	0.25893	87962.103221	0.004233	0.004354	3	0.002037	0
65	Hyperten...		disease	Cardiovas...	9	21	30	9.0	21.0	30.0	10.0	0.207705	0.243412	41100.928644	0.003365	0.003331	3	0.001353	0
139	Prostate c...		disease	Cancer	20	32	52	20.0	32.0	52.0	10.0	0.198655	0.249799	23646.599046	0.158507	0.175458	2	0.002257	0
185	Leigh syn...		disease	Neurological	6	18	24	6.0	18.0	24.0	9.0	0.186114	0.213701	19802.5	0.000115	0.000113	6	0.000959	0
202	Fanconi a...		disease	Multiple	4	15	19	4.0	15.0	19.0	11.0	0.178837	0.217945	4643.625	0.055342	0.05857	2	0.000698	0
302	Epidermol...		disease	Dermatolo...	9	20	29	9.0	20.0	29.0	10.0	0.16254	0.189131	57443.0	0.000014	0.000013	0	0.001694	0
364	Parkinson ...		disease	Neurological	6	17	23	6.0	17.0	23.0	11.0	0.189168	0.223553	40683.732257	0.005689	0.006598	5	0.001071	0
634	Thyroid c...		disease	Cancer	26	37	63	26.0	37.0	63.0	11.0	0.193557	0.250949	40047.983786	0.225486	0.239616	2	0.003055	0
53	Myopathy		disease	Muscular	9	19	28	9.0	19.0	28.0	9.0	0.192873	0.225134	116958.166667	0.000165	0.000186	6	0.001294	0
86	Hemolytic ...		disease	Hematolo...	8	18	26	8.0	18.0	26.0	13.0	0.139361	0.165595	12305.0	0.000009	0.00001	5	0.001452	0
99	Myocardia...		disease	Cardiovas...	10	20	30	10.0	20.0	30.0	10.0	0.210355	0.246026	62534.722326	0.003824	0.003847	3	0.001357	0
117	Gastric ca...		disease	Cancer	27	37	64	27.0	37.0	64.0	10.0	0.209794	0.263603	39061.357414	0.16367	0.173737	2	0.003046	0

Centrality

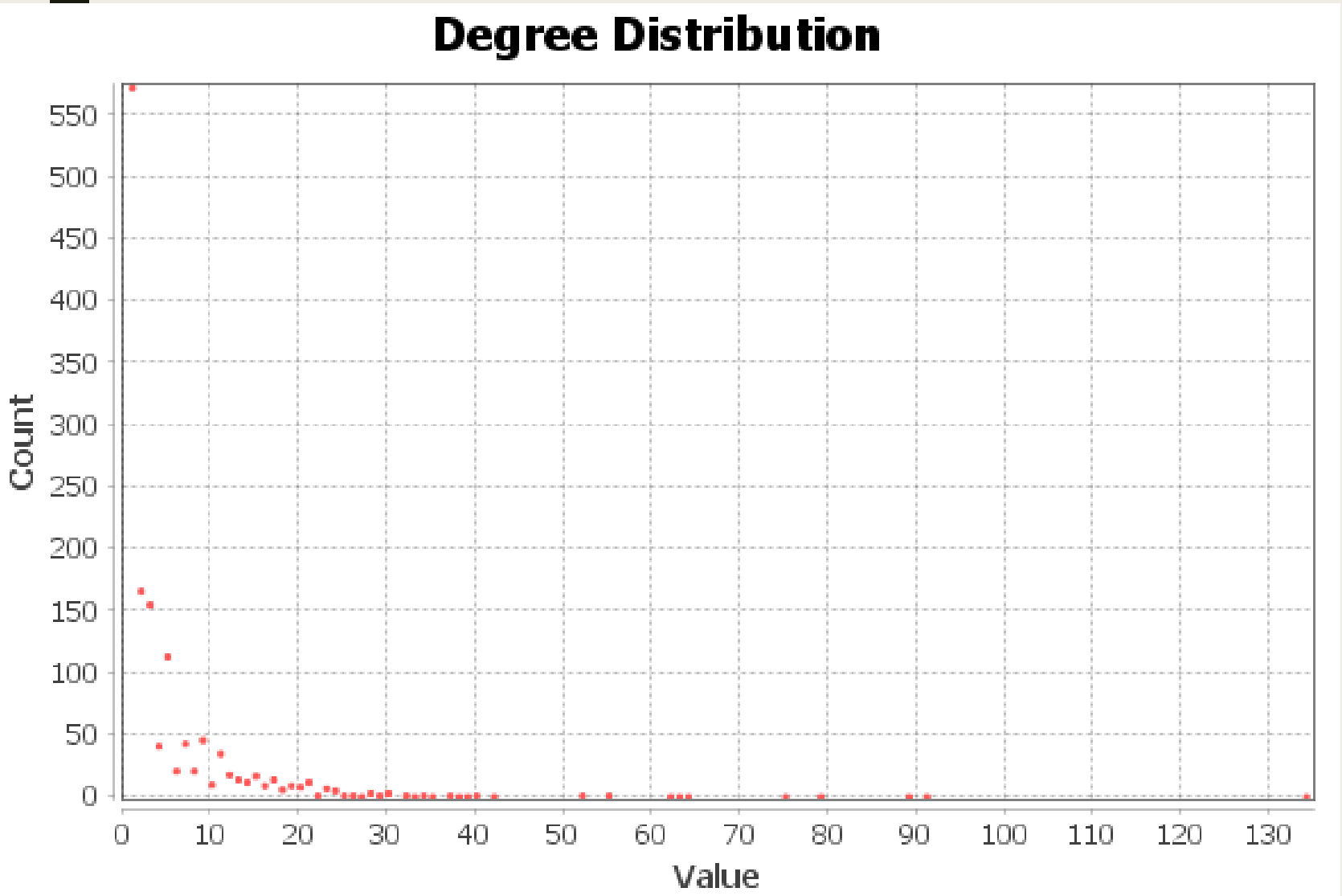
1. Degree Distribution
2. Eigenvector Centrality
3. Page Rank
4. Closeness Centrality
5. Betweenness Centrality
6. Harmonic Closeness Centrality
7. Eccentricity
8. Hubs Distribution

Degree Distribution





- In – Degree
  - Lowest = 1 (260 nodes)
  - Highest = 50 (1 node)
- Out – Degree
  - Lowest = 0 (55 nodes)
  - Highest = 84 (1 node)

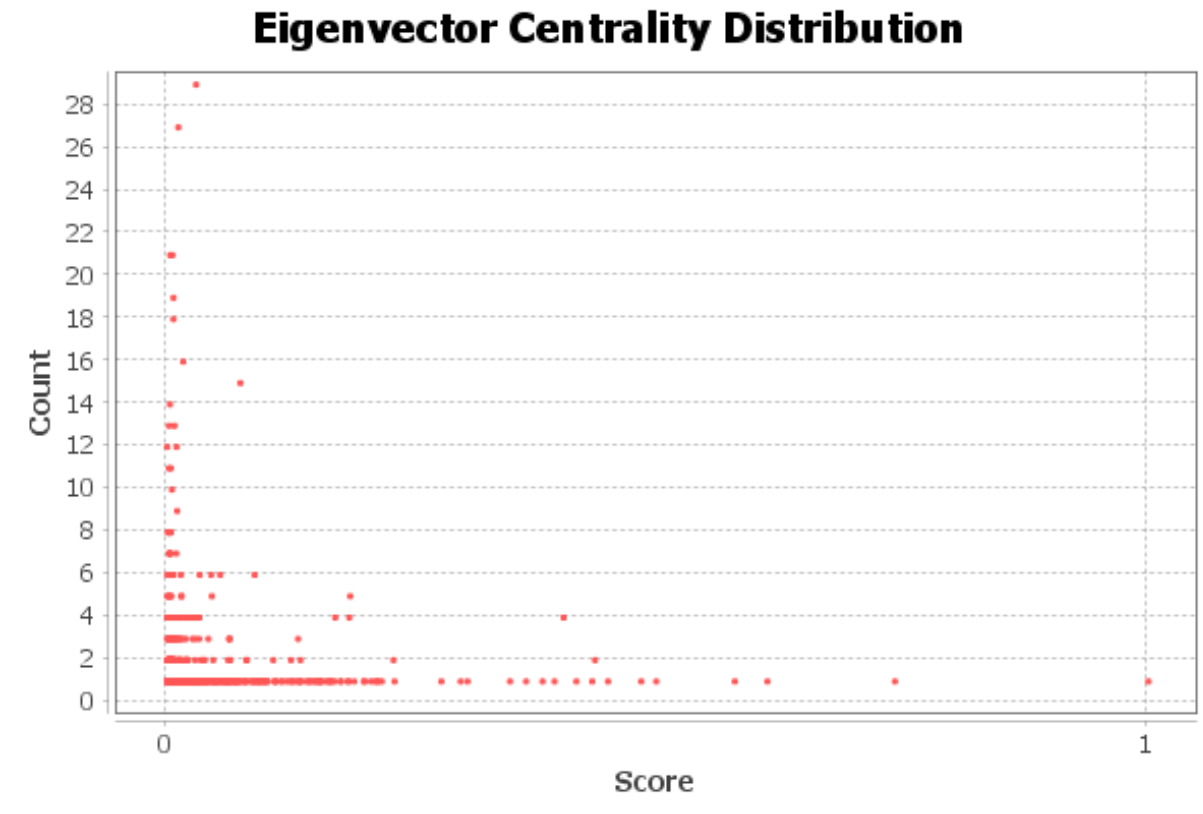


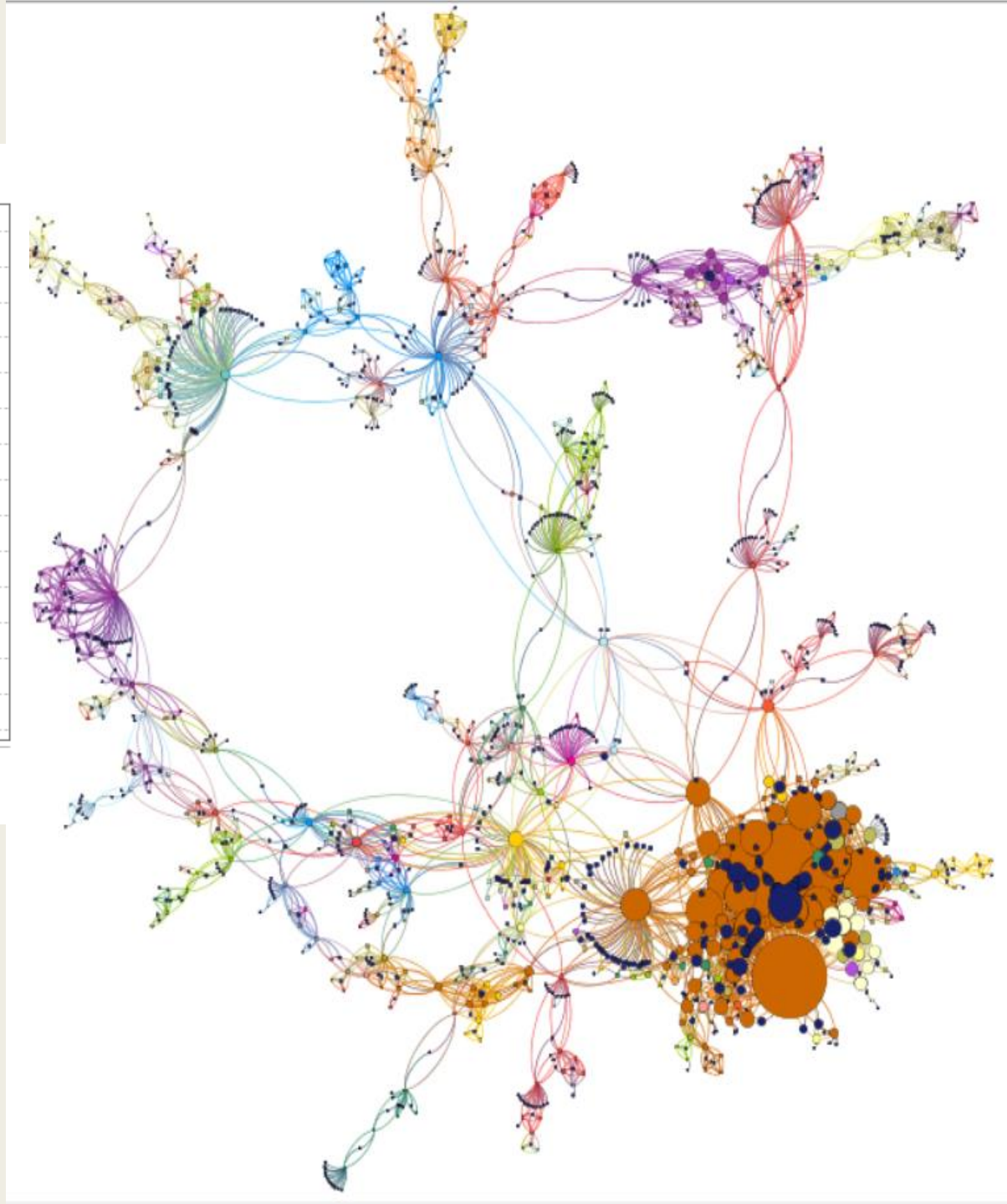
- Lowest degree = 1 (572 nodes)
- Highest degree = 134 (1 node)

- Since there are 1419nodes in total the node with largest degree is connected to 9% of the whole network in one hop
- Such nodes are the hubs
- Similarly 40 % of the nodes have degree 1 (572 nodes)

- The highest out–degree, in-degree and degree is contributed by the same node :- Colon Cancer
- It is a scale-free network since it has long tails
- Right – Skewed
  - Many nodes with small degrees, few with extremely high

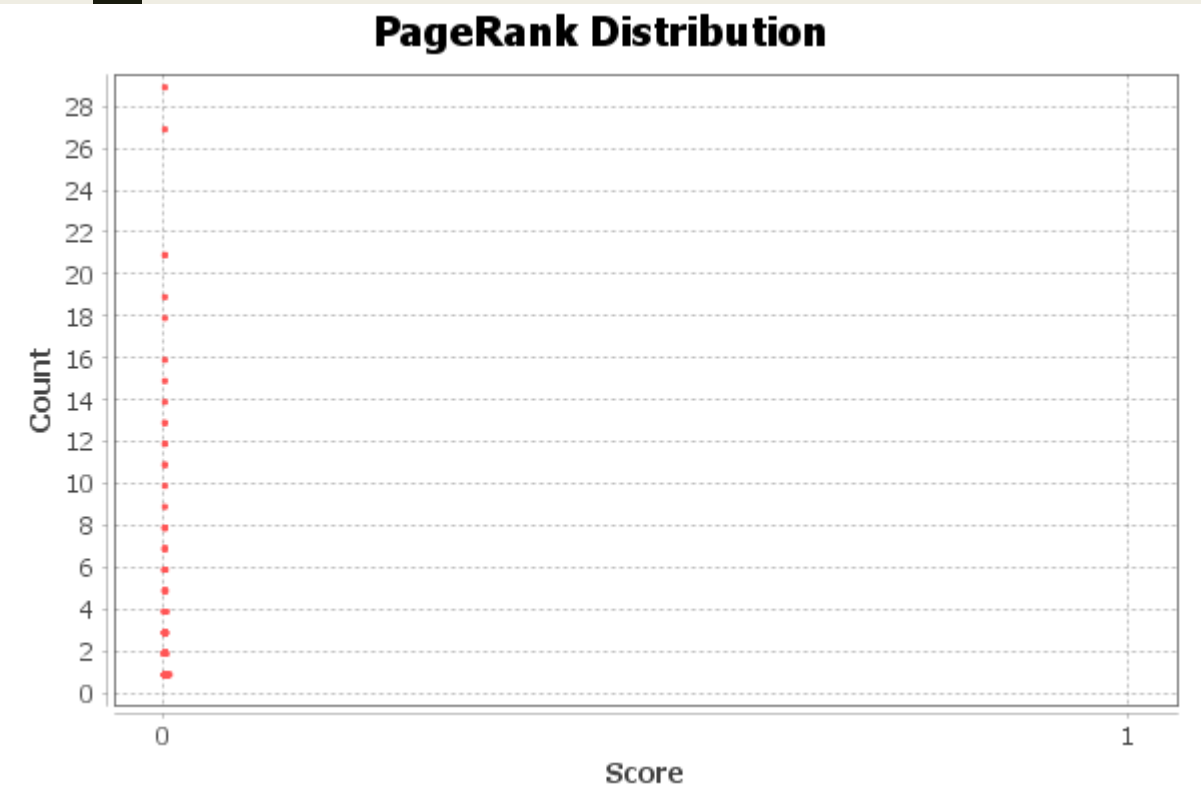
Eigenvector Centrality





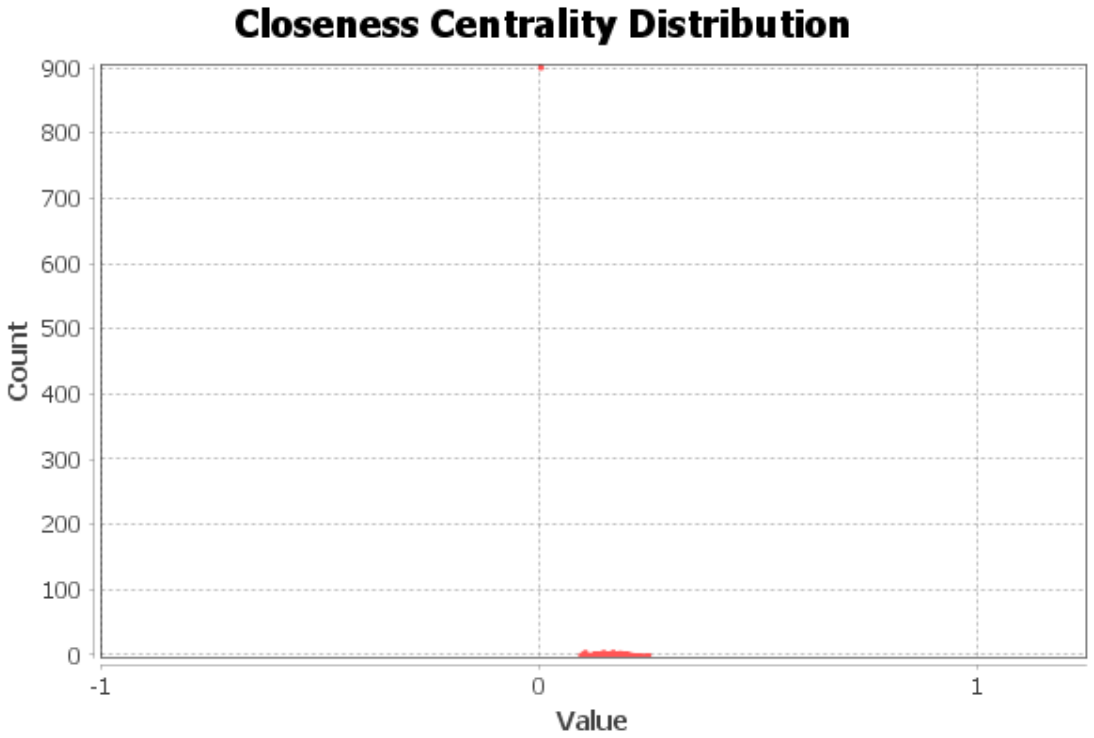
- Highest Eigenvector centrality :- Colon Cancer
  - Meaning it is highly linked (from degree centrality) and/or this node might have important linkers.
- High Eigenvector centralities are from cancer group

Page Rank

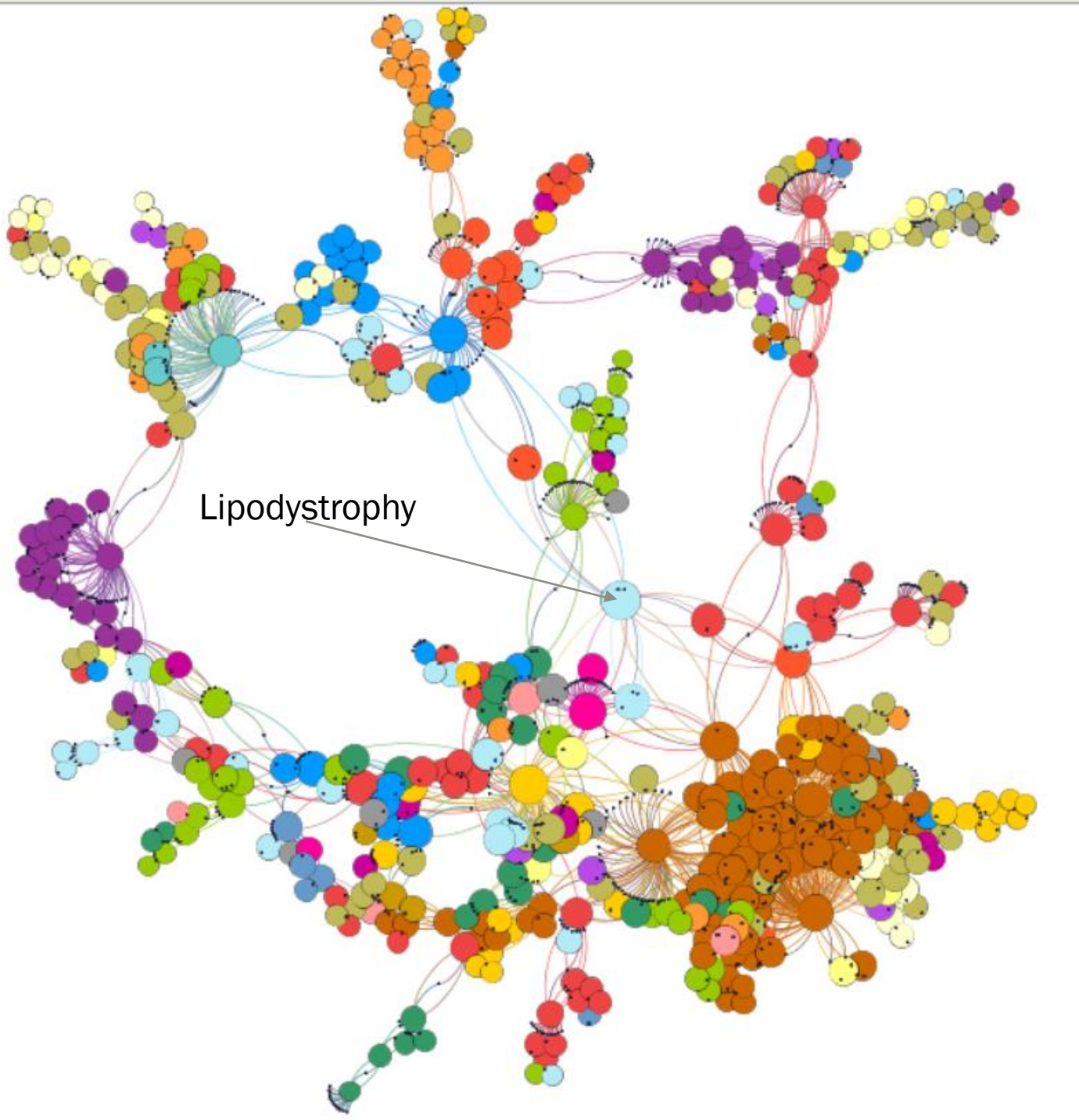


- Lies between 0.000449 and 0.005352
- Highest Page Rank: Colon Cancer
- Cancer group has high page ranks
  - They are connected to each other

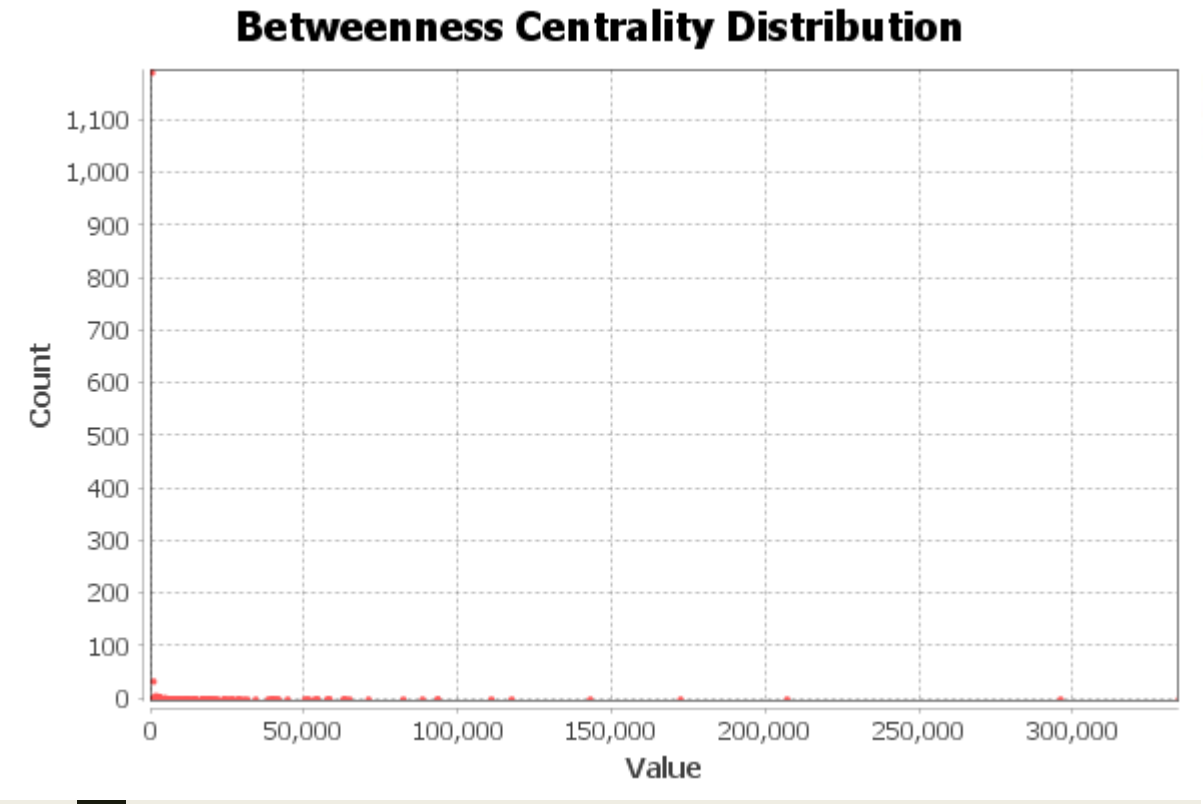
Closeness Centrality



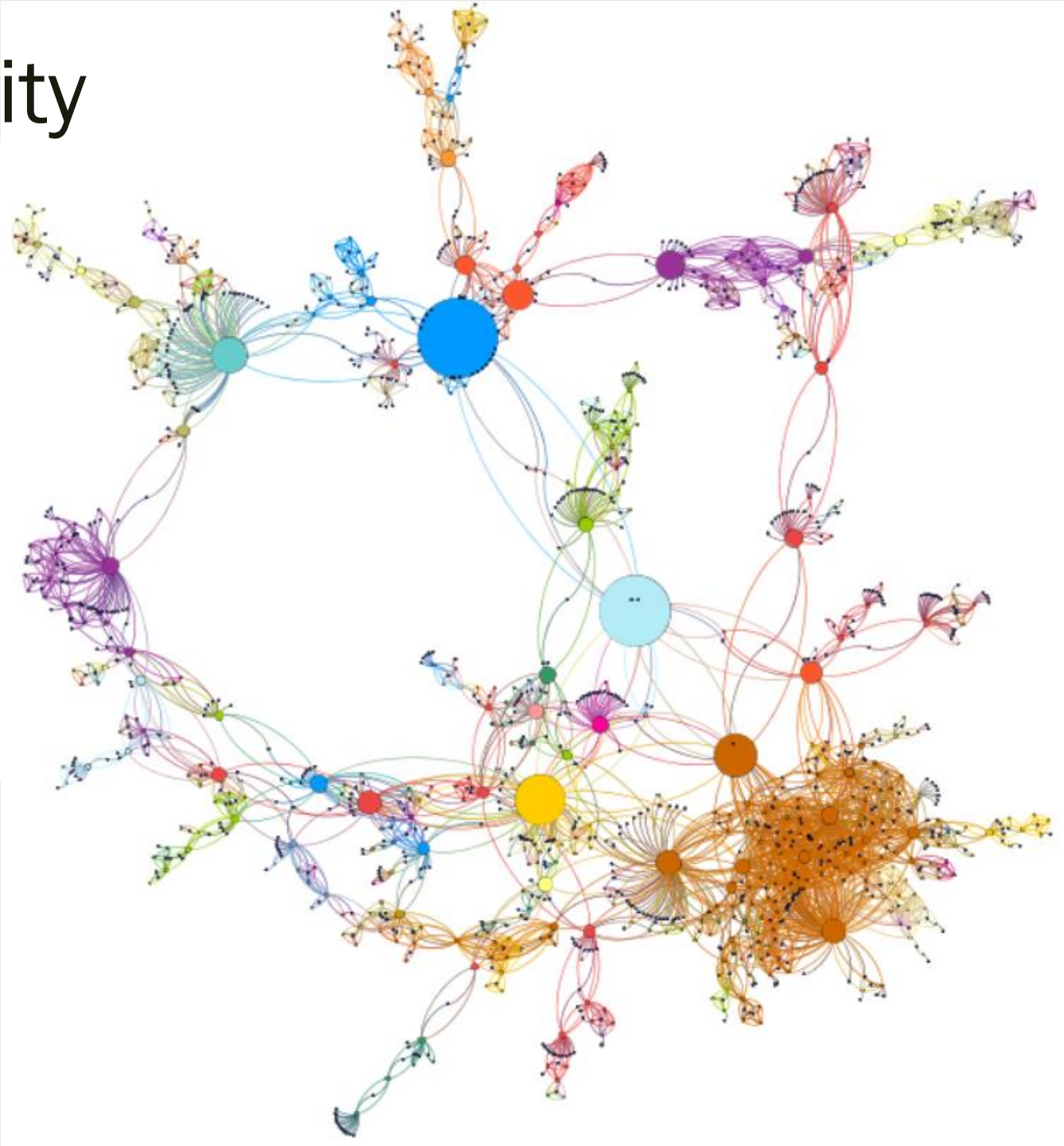
- Varies between 0.0 and 0.2454
- Highest :- Lipodystrophy
- Colon Cancer :- 0.2198 (7th highest)
- Cancer group dominates
- Genes have values 0 – no outgoing edges
- Nodes have a similar score (except for genes) since it is a highly connected network



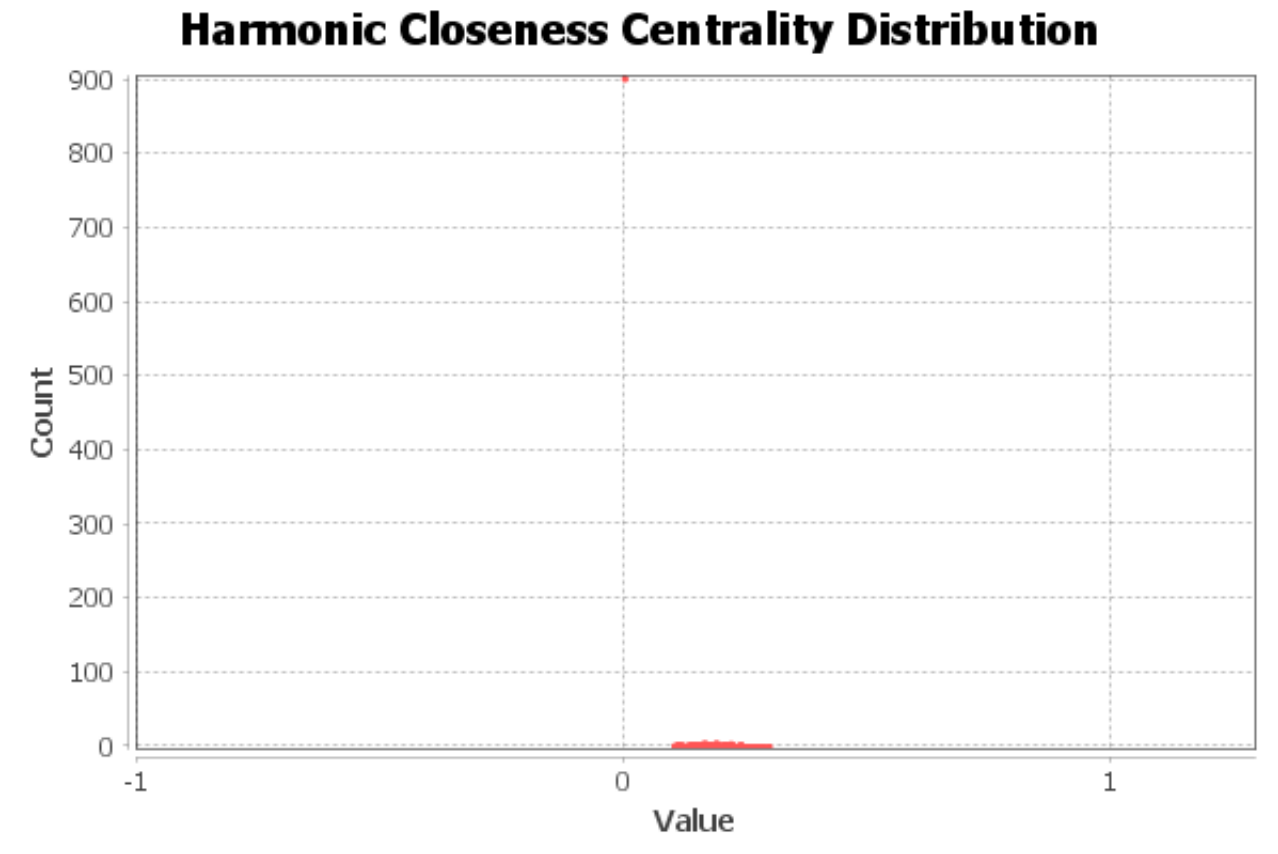
Betweenness Centrality

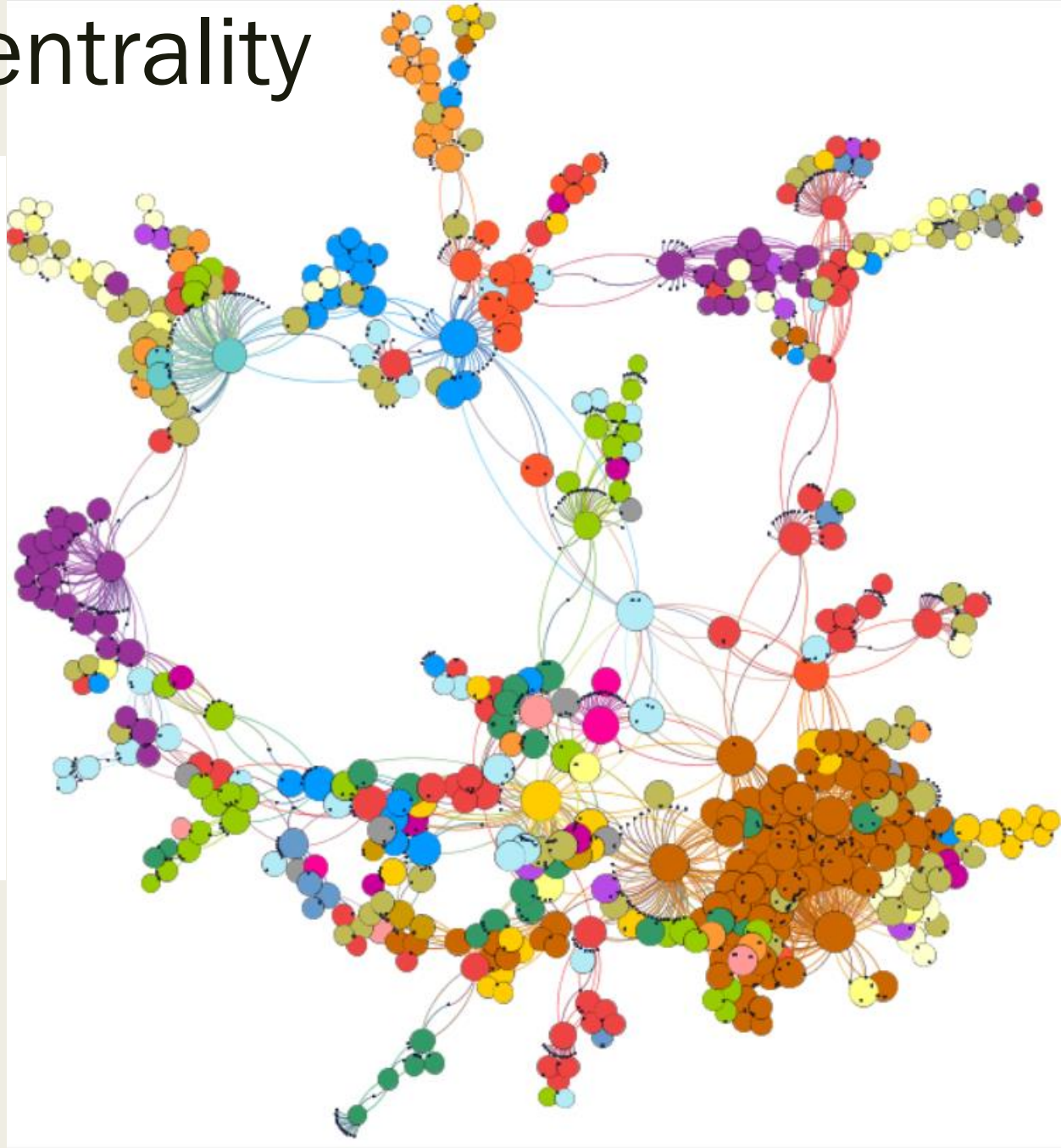


- Lies between 0.0 and 333925.1436
- Highest:- cardiomyopathy
- Genes – 0 :- no outgoing edges
- Some diseases also have 0 values



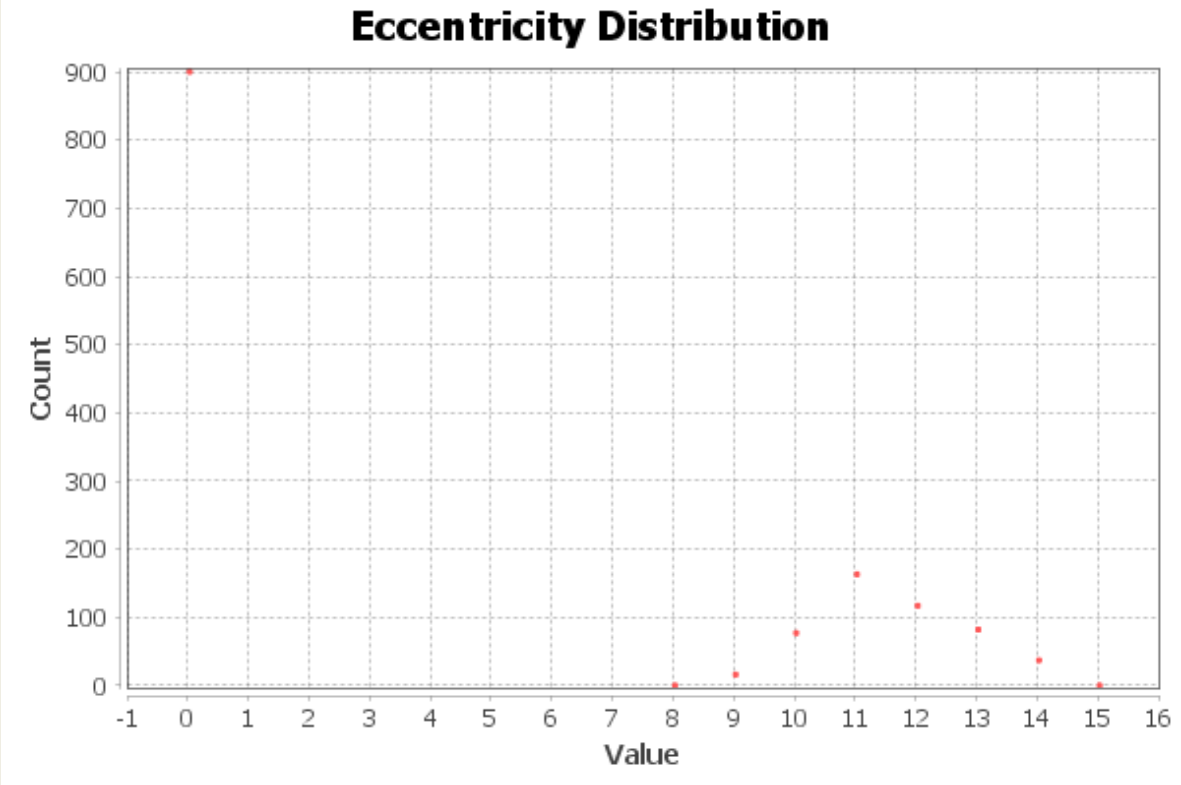
Harmonic Closeness Centrality



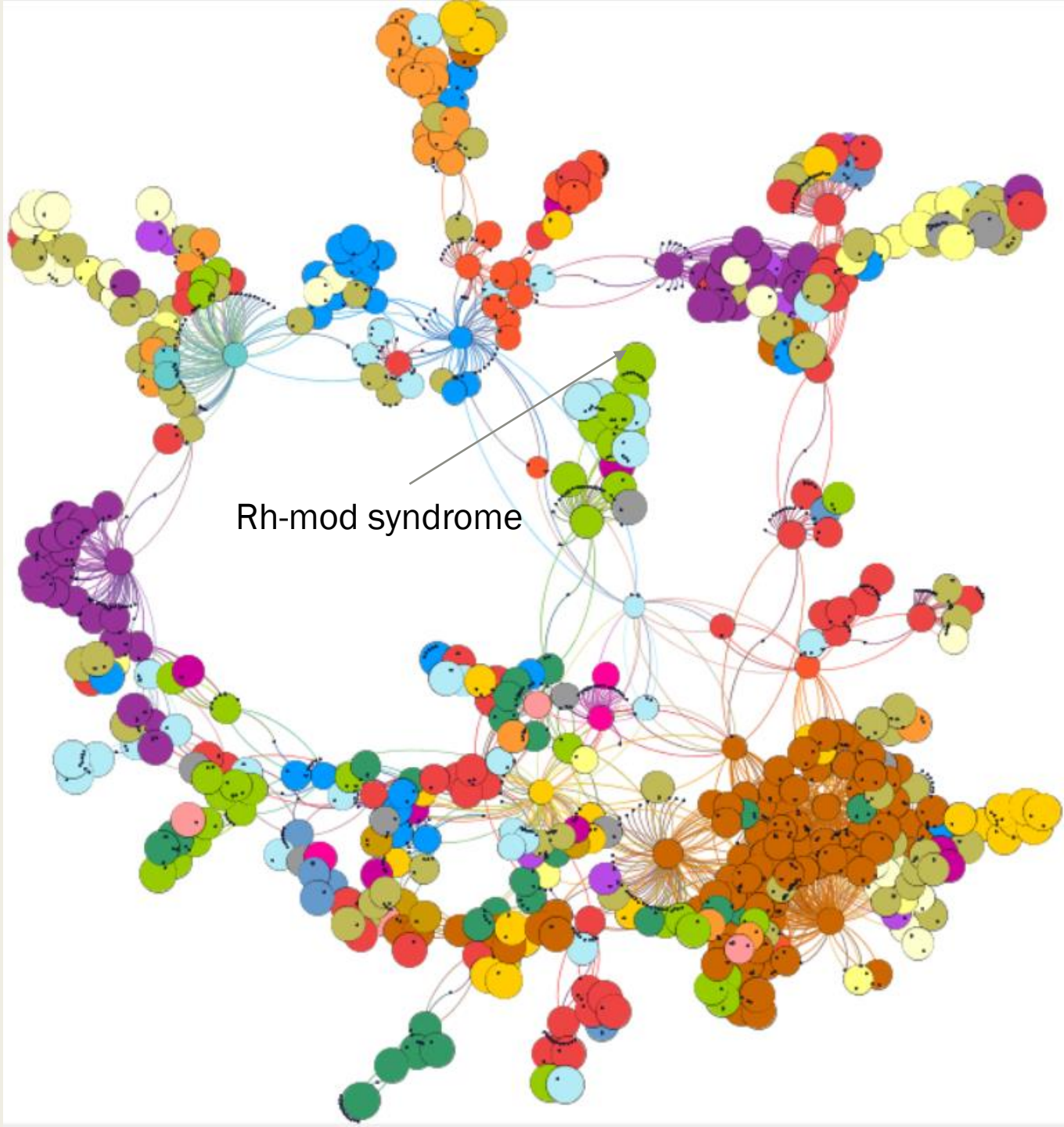


- Between 0.0 and 0.2962
- Highest :- Colon Cancer

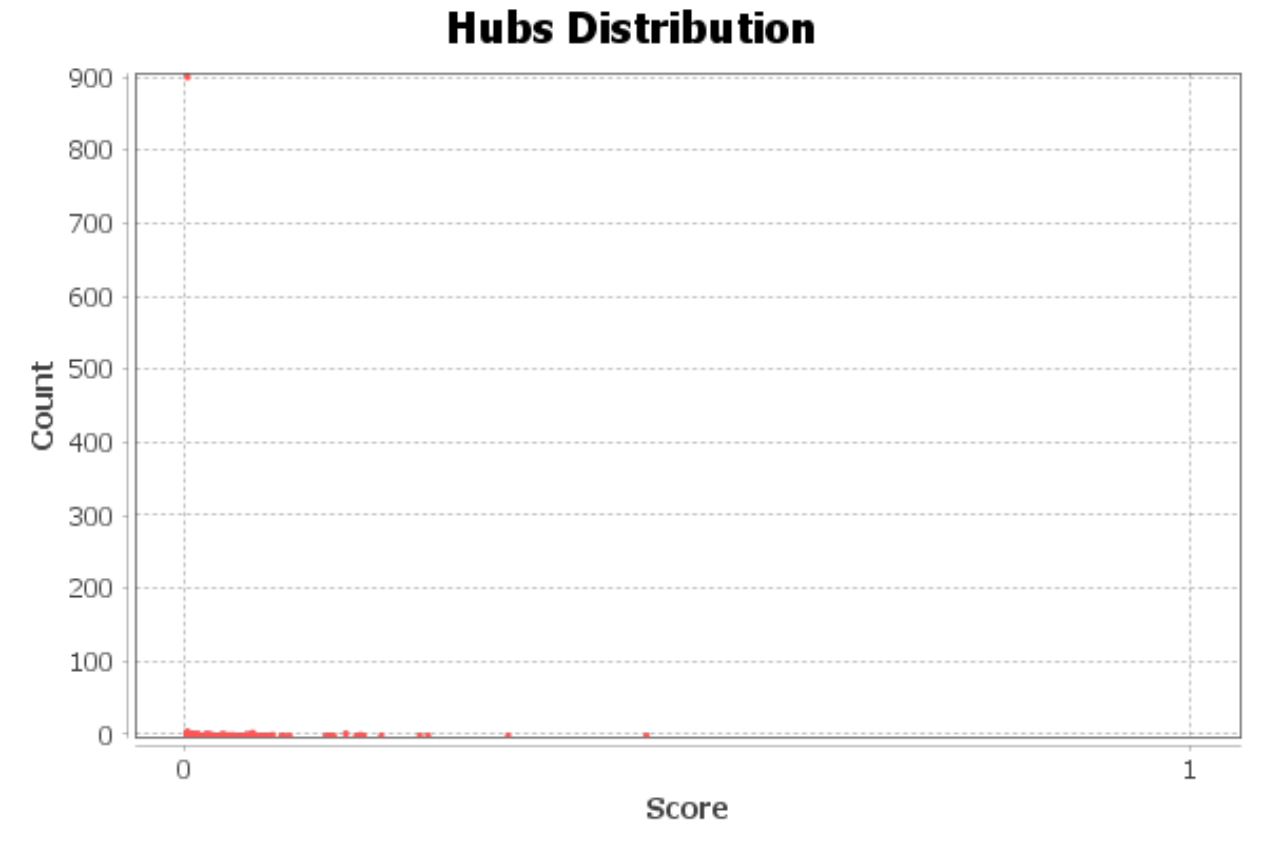
Eccentricity



- Between 0.0 and 15.0
- Highest Rh-mod syndrome
  - Max distance to any other node is 15
  - Network diameter also 15

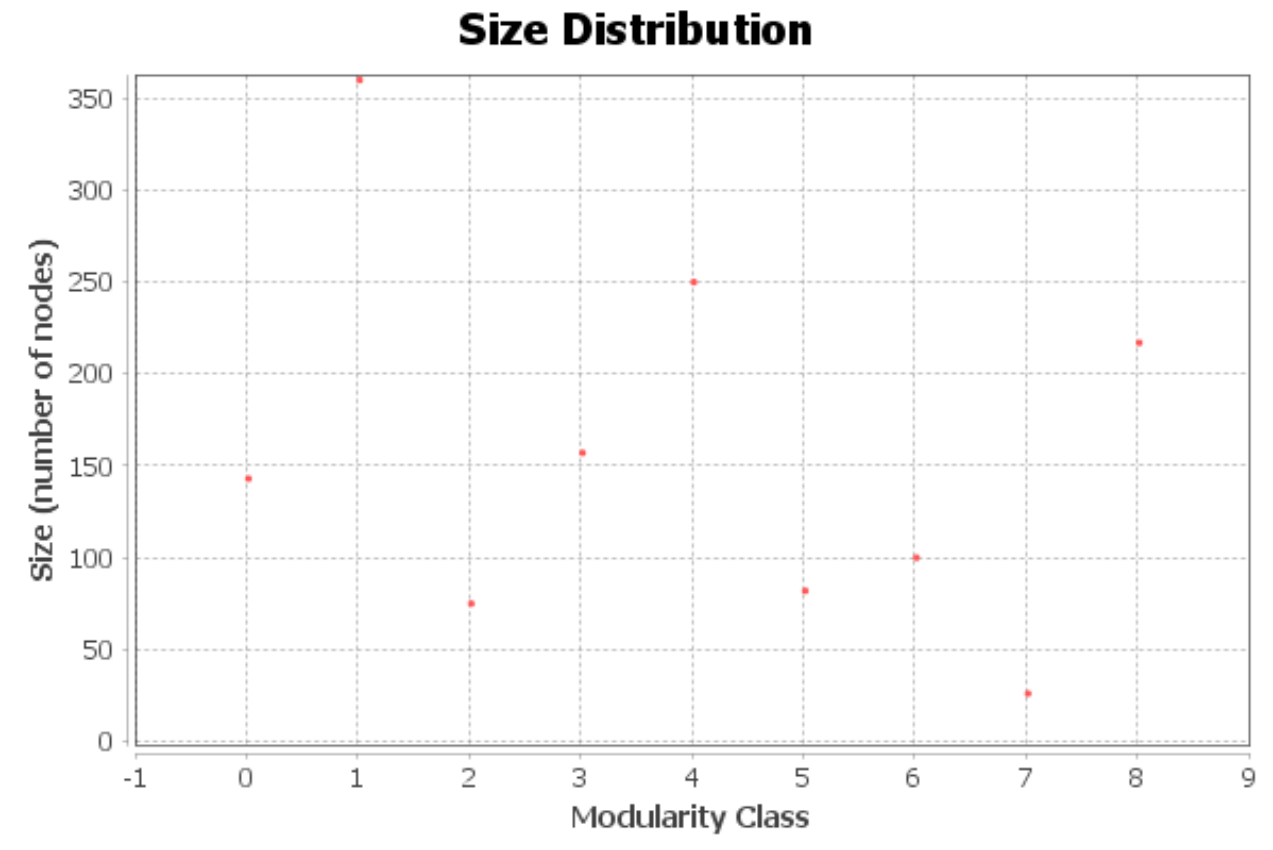


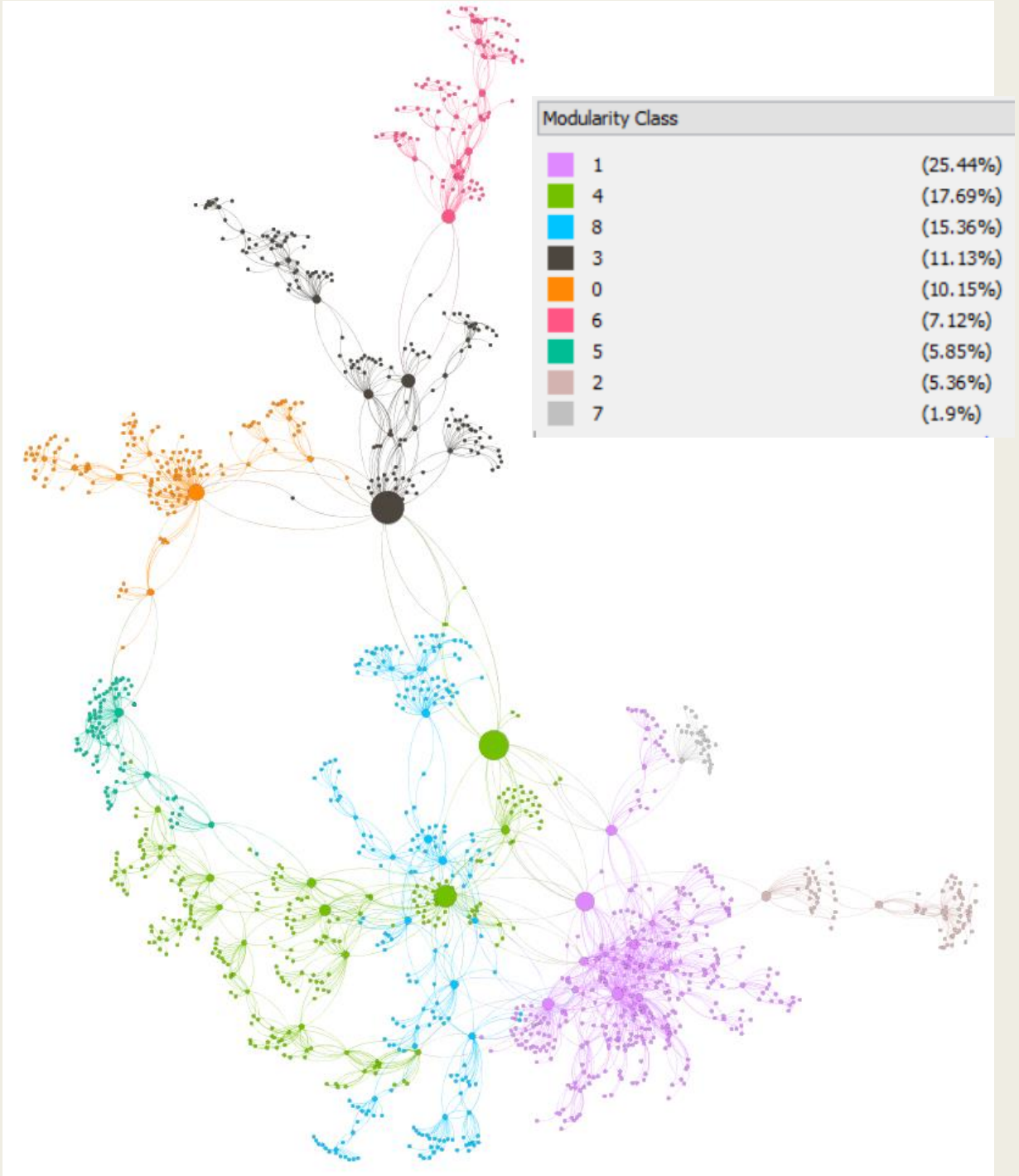
Hubs Distribution



- a hub is a node with a number of links that greatly exceeds the average
- Nodes in the Cancer group are the hubs in this network
- Between 0.0 and 0.457335
- Colon Cancer has the highest
- This is a scale free network

Modularity

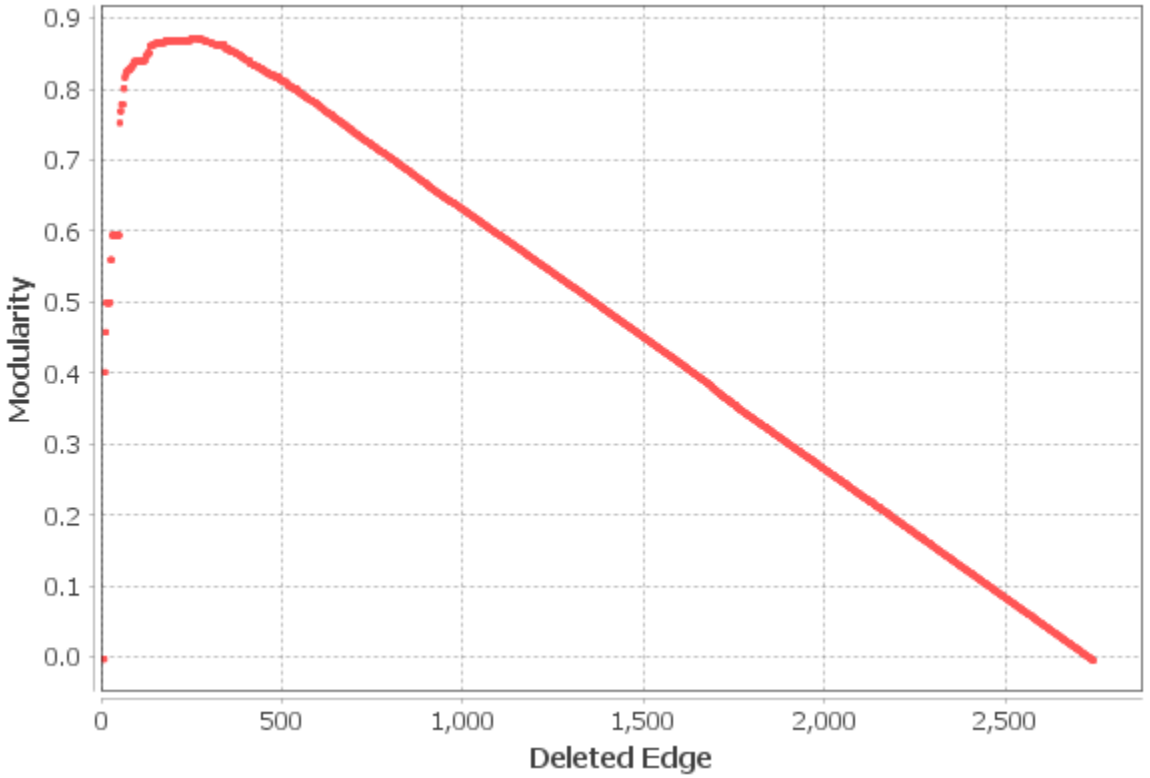


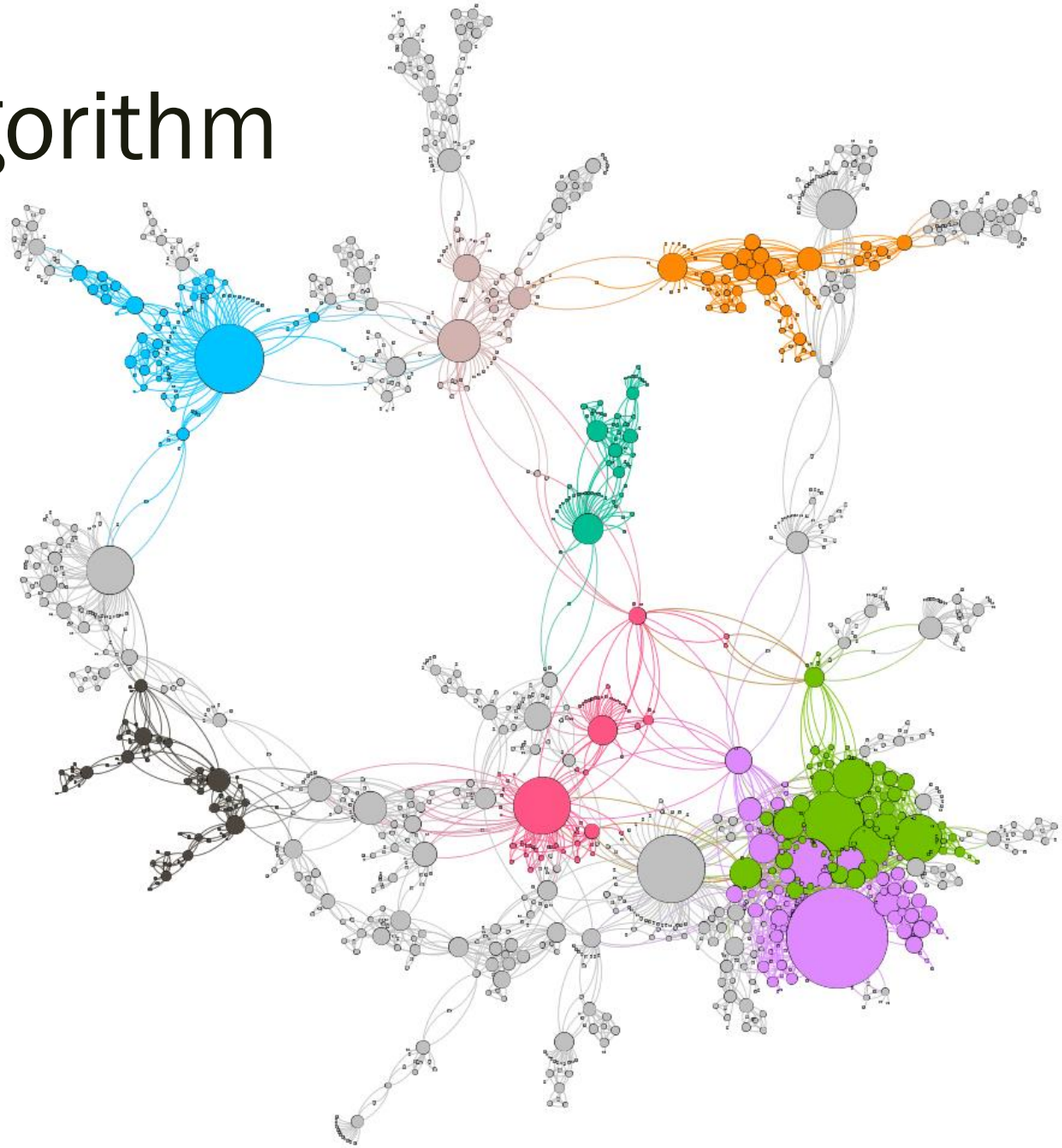


- No of communities 9
- Modularity : 0.806
- Highest – Group 1 – 360 nodes (25.44 %)
- Lowest – Group 7 – 26 (1.9 %)

- All the group contains diseases as well as genes

Clustering :
Girvan – Newman Algorithm

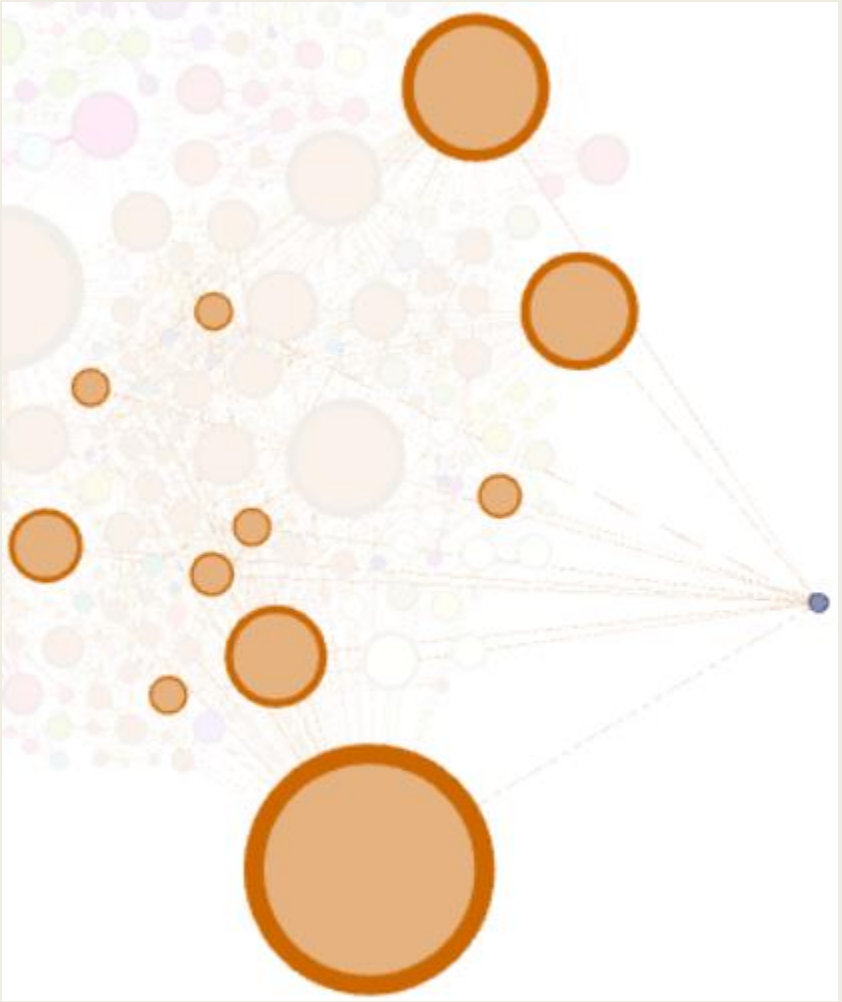




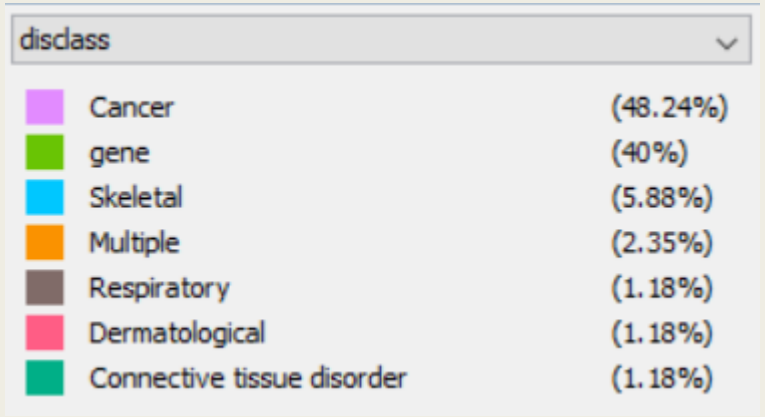
- No of communities 34
- Modularity : 0.87496436
- All the group contains diseases as well as genes

Observations

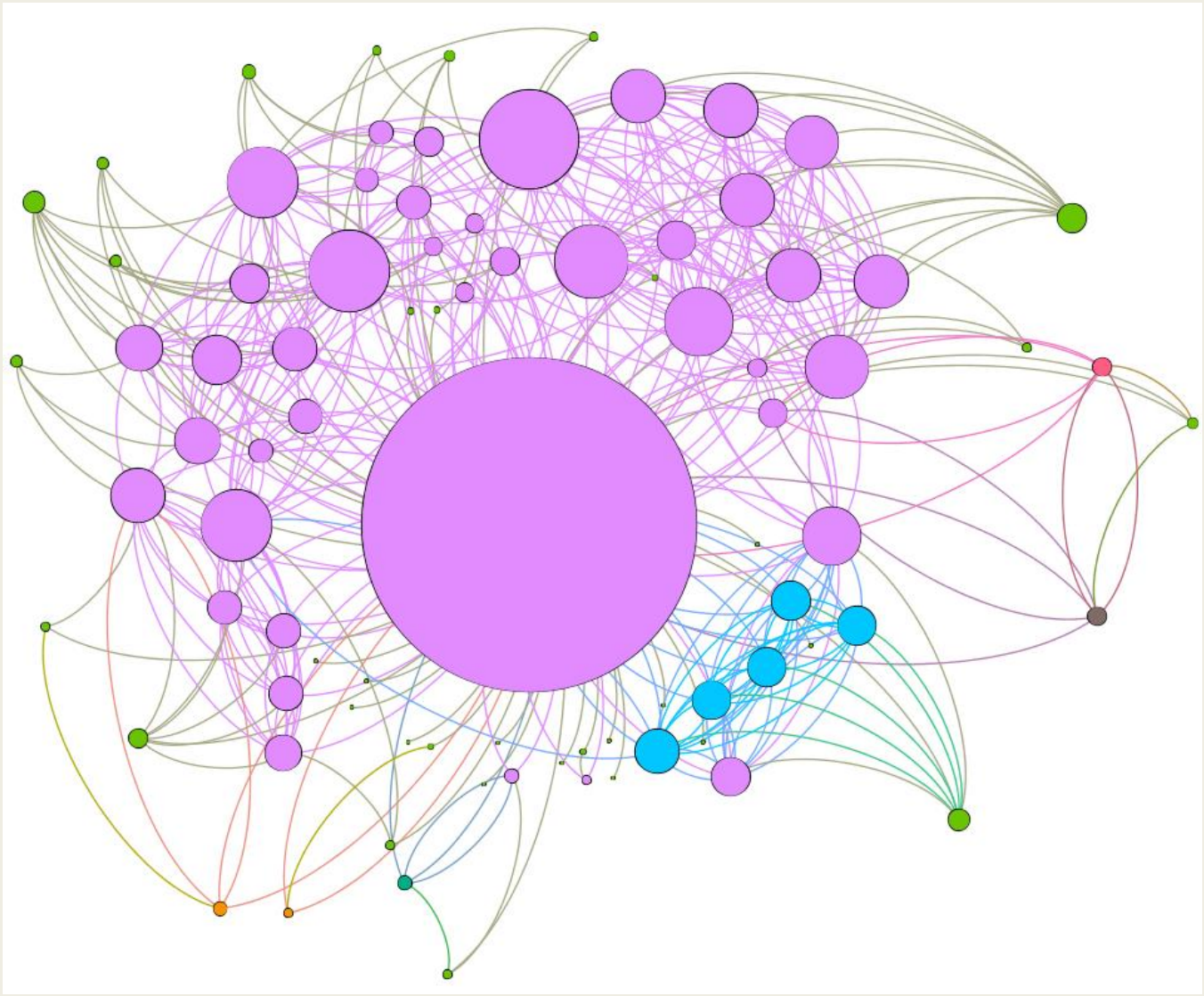
- Degree
  - *Genes out degrees are 0*
  - *Gene with highest in-degree(11) is tp53 which is pointed from the cancer group*
  - *Any changes/ issues to this gene can trigger many types of cancers which again connected to many diseases.*
- Clustering Coefficient
  - *0 and 1 for Genes*
- Centrality
  - *Cancer group has high centrality*
  - *Colon cancer has the highest values*
  - *Page rank and hub shows cancer groups connected together*



Colon Cancer and its neighbors

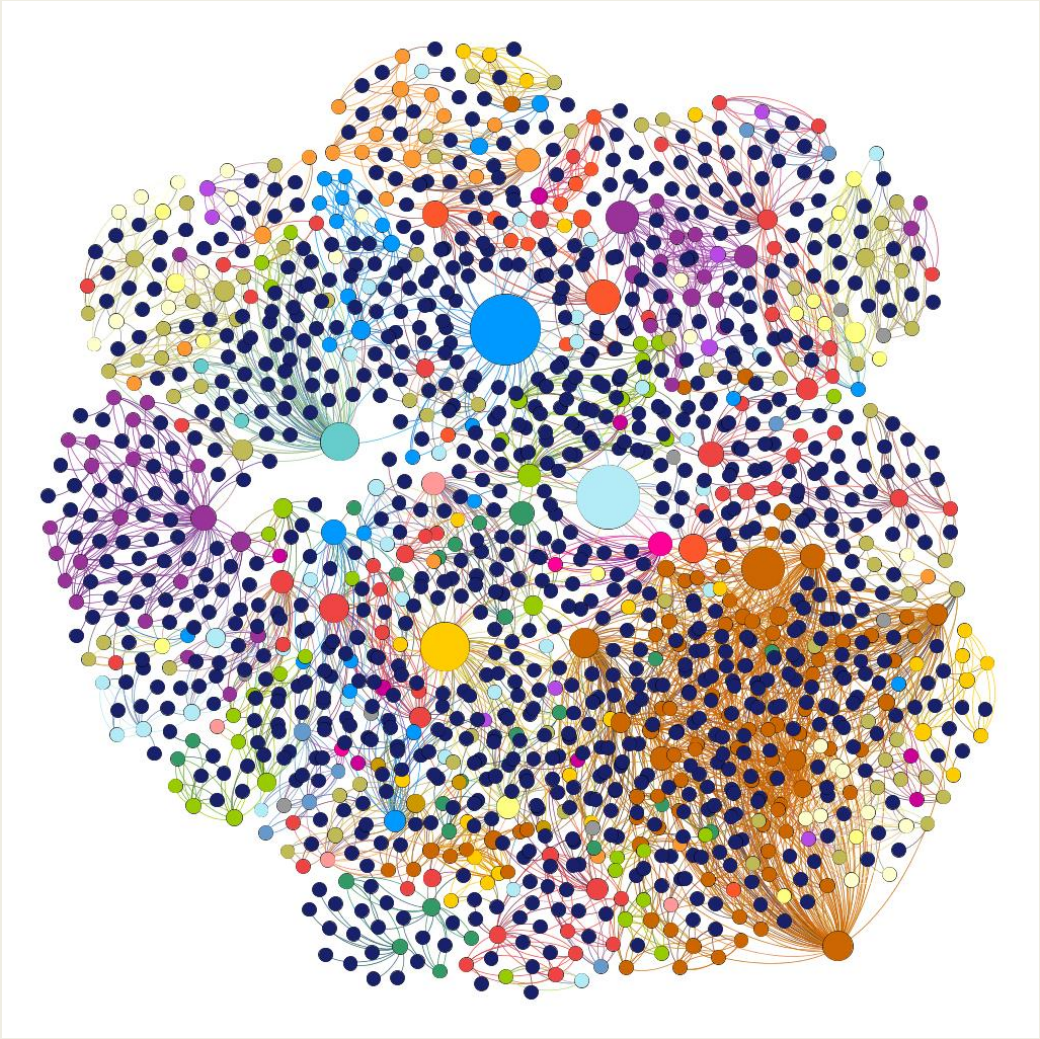


Degree : 134 (in : 50 – out: 84)
Nodes : 85
Edges : 504
Genes Connected : 34



Betweenness Centrality

- Eventhough degree centrality for colon cancer is higher, cardiomyopathy has the highest betweenness centrality
- Colon Cancer mainly related to its groups
- Cardiomyopathy connected to other groups as well



Comparison

- Removal of genes increased the values of statistics lightly
- No other effects
  - *Can be because genes doesn't have any out-going edges*
- Scale-free network : removal or adding new edges doesn't change the network

Bibliography

- Gephi - The Open Graph Viz Platform.” [Online]. Available: https://gephi.org/. [Accessed: 24-Sep-2018].
- [2] *Gephi: The Open Graph Viz Platform. Contribute to gephi/gephi development by creating an account on GitHub*. Gephi, 2018.
- [3] “Social network analysis: Centrality measures,” *Cambridge Intelligence*, 03-Dec-2014. [Online]. Available: https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/. [Accessed: 02-Dec-2018].
- [4] “Closeness centrality,” *Wikipedia*. 04-Oct-2018.
- [5] “Eccentricity.” [Online]. Available: http://www.cbmc.it/fastcent/doc/Eccentricity.htm. [Accessed: 03-Dec-2018].
- [6] “Modularity (networks),” *Wikipedia*. 14-Nov-2018.

Stay Healthy