

## Coronary Artery Disease Network Analysis

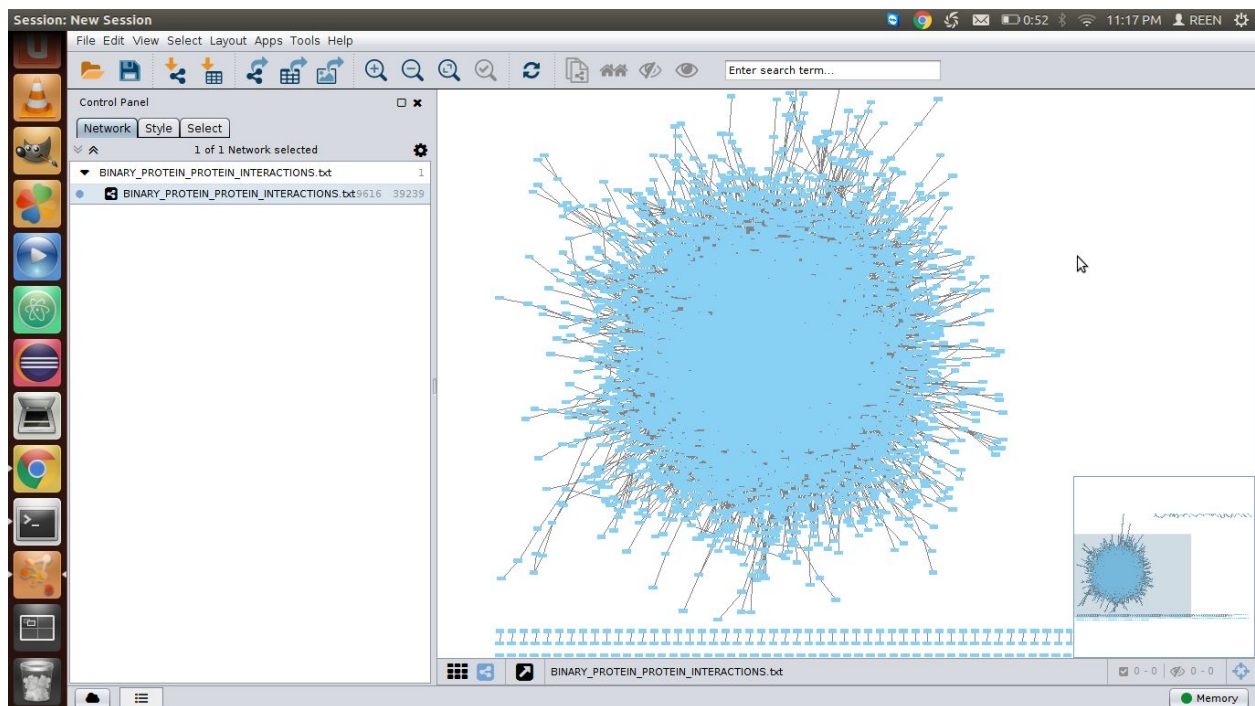
-Avneet Kaur(201402)

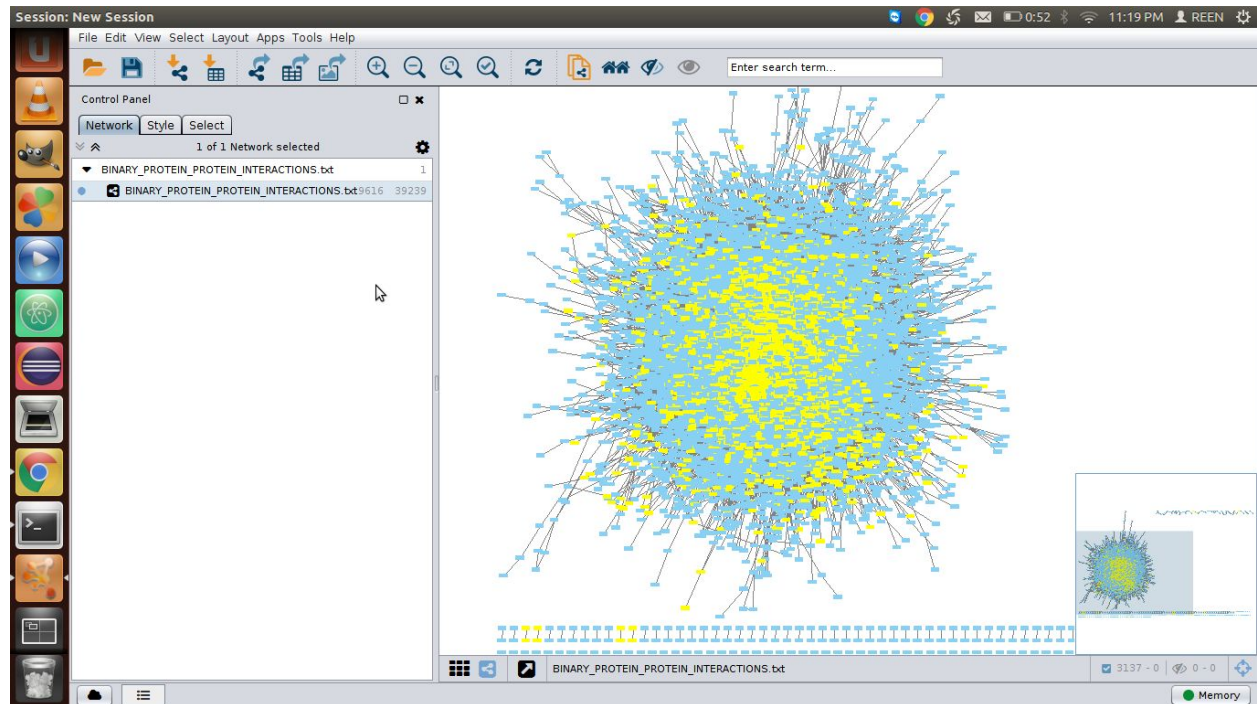
In this project, network analysis was done on a network formed using the genes of coronary artery disease.

Database source : CADGene Disease database

The network was constructed by mapping the genes from the CAD database onto the human binary interaction network from HPRD. Cytoscape was used for the same. Yellow nodes indicate mapped network consisting of the list of genes(604) in CAD database as well as other nodes.

Original and mapped network shown below :

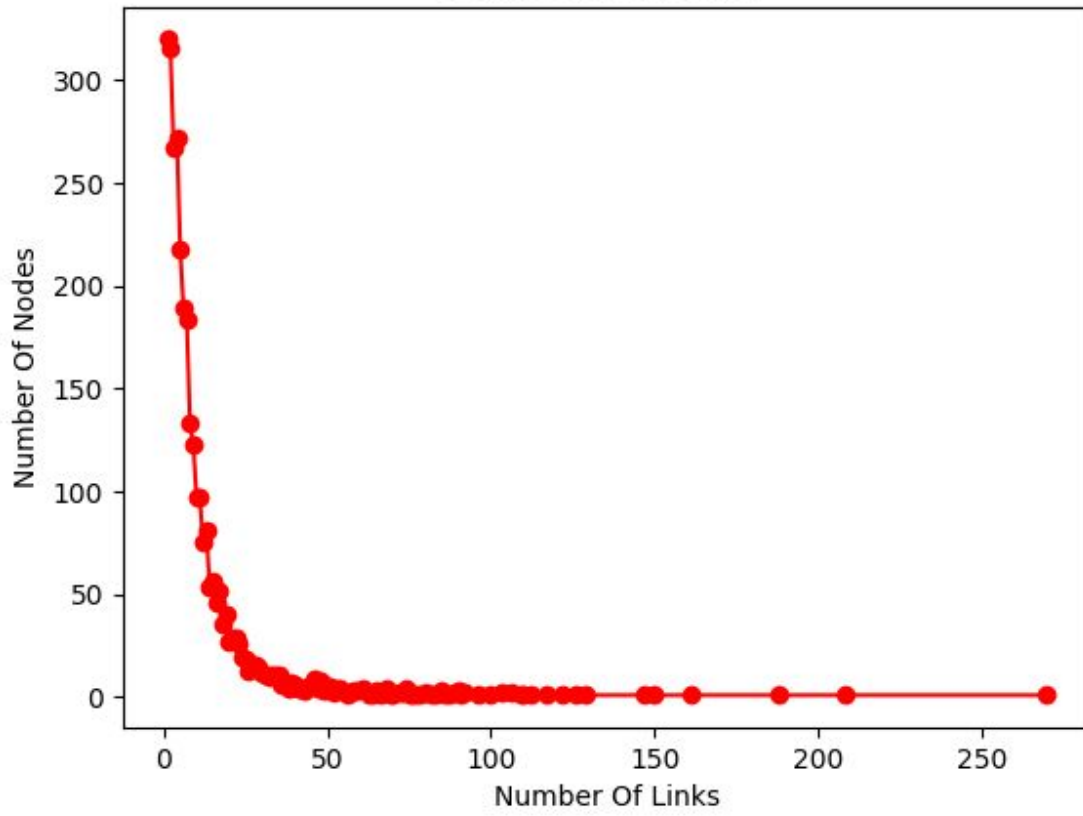


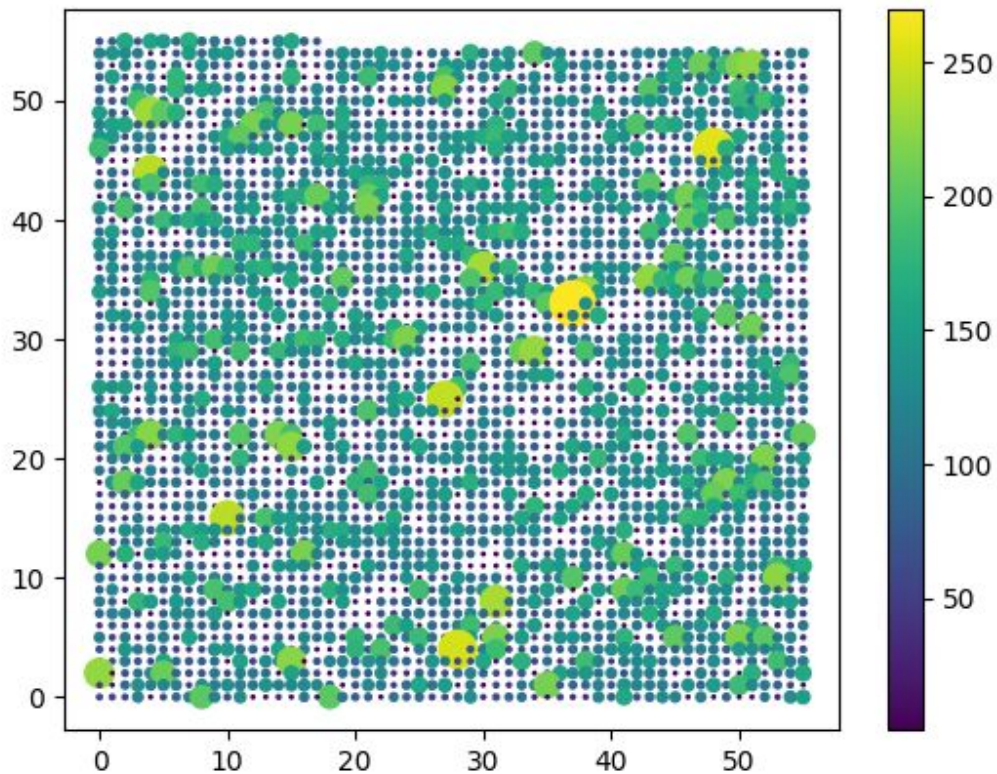


(i) Topological Analysis : Various topological features were analysed.

→ Degree Distribution : A power law degree distribution was obtained indicating a scale free nature as can be seen in the figure below. Further, in the next figure, the degree of node is represented by various colours and the size of the circle represents the size of the nodes. We can conclude that there are very few nodes of high degree by looking at the figure.

Degree Distribution

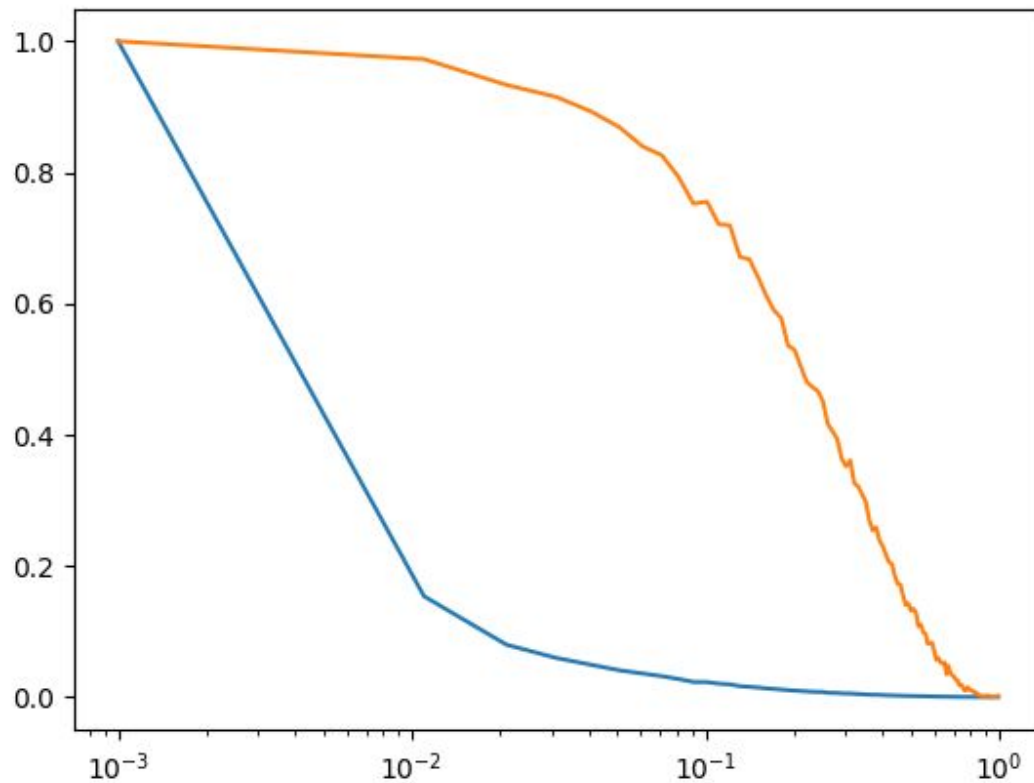




→ Small world analysis : Clustering coefficient of 0.64.

Characteristic path length of 3.590 was obtained

This highlights a small-world nature. For comparison, watz-strogatz model was implemented on another graph with the same number of nodes. The following graph was obtained.



So the clustering coeff.

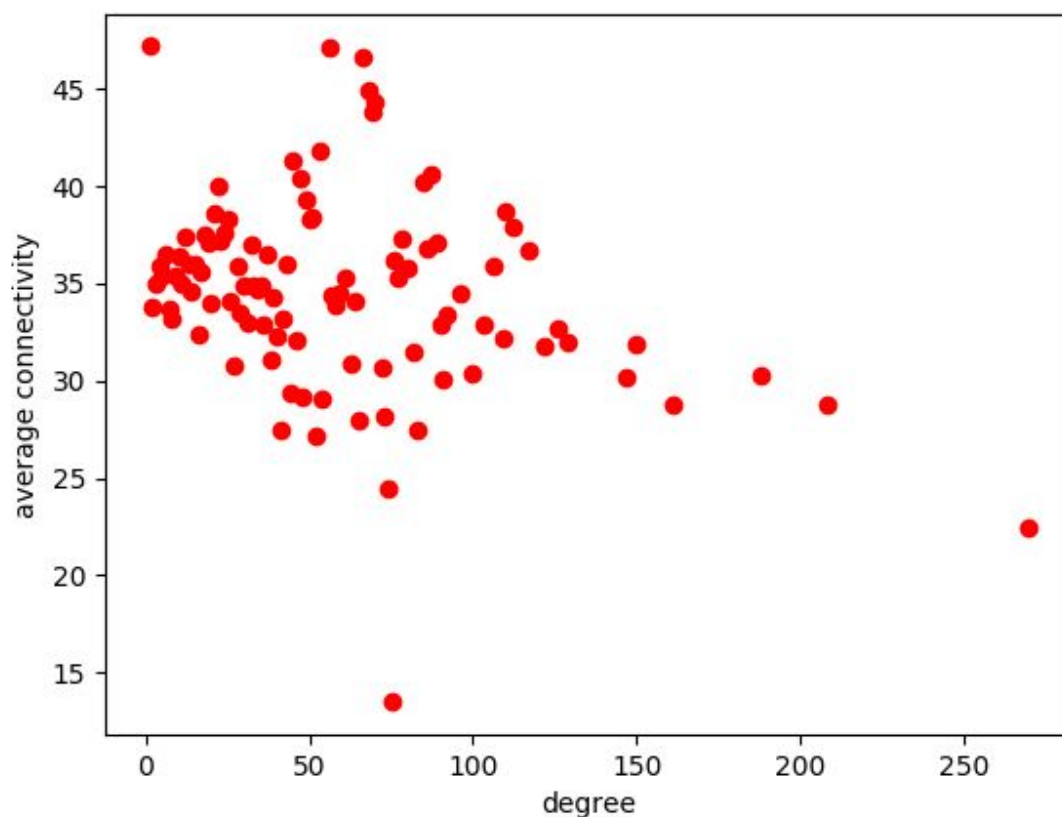
→ Clustering-Modularity: It indicates the degree of connectivity within the various modules of the network. Communities were detected using a networkx function and plotted according to colour. A value of 0.4804 was obtained indicating medium degree connectivity among the modules of a network also indicating there are inter-module as well as intra-module connections of roughly the same amount.

→ degree correlation: It indicates the preference for a network's nodes to attach to others that are similar in some way. For example hubs (greater degree nodes) tend to attach to greater degree nodes in case of social media networks whereas in case of biological networks , they tend to show dissortativity, that is greater degree nodes tend to attach themselves to lower degree nodes. Even in our case, CAD protein-protein interaction network, a value of -0.033 was obtained which indicated the above mentioned result.

→ identification of hub nodes: Hub nodes were identified by their degree. Greater degree nodes tend to be hub nodes. The following hub nodes were obtained along with there degree. (top 10)

(u'TP53', 270), (u'SRC', 208), (u'ESR1', 188), (u'EGFR', 161), (u'AR', 150), (u'GRB2', 147), (u'EP300', 129), (u'CREBBP', 126), (u'FYN', 122), (u'SHC1', 117)

→ average degree connectivity: The average degree connectivity is the average nearest neighbor degree of nodes with degree k.



(ii)Node Deletion Studies: carried out to study the effects of random failure as well as targeted attack on the network. The graphs are attached in the folder.

→ Characteristic path length was found to increase substantially from 3.59 to 4.3 as fraction of deleted nodes were increased from 0 - 0.04 in steps of 0.001 in case of targeted attacks whereas in case of random failures it increased only a little from 3.59 to 3.61 in the same conditions.

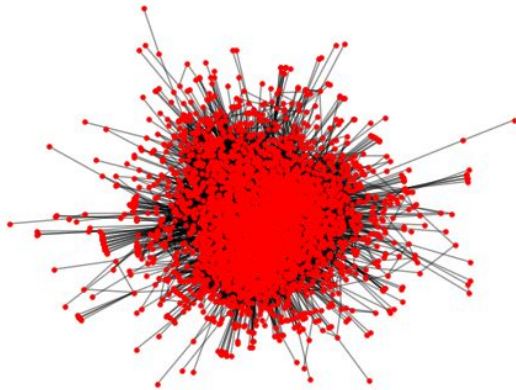
→ The size of giant cluster also decreased, but this decrease was faster and more as compared to random failures as can be seen in the graphs.

→ size of isolated components : showed an overall decrease in targeted attacks and little less decrease in random attacks although I was not able to capture the initial increase in size of isolated components in targeted attacks as in case of “Error and attack tolerance” paper.

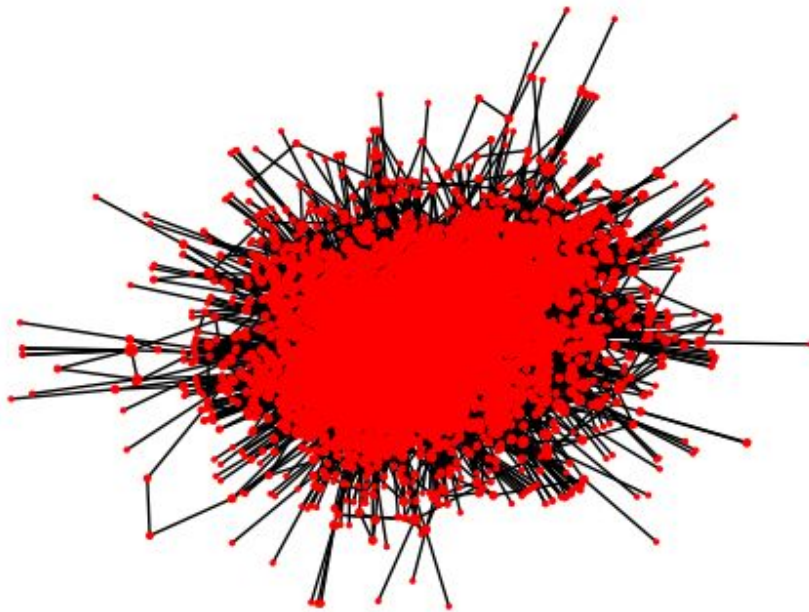


(iii) Network visualisation : The size of the node is proportional to its degree. On zooming

Degree rank plot

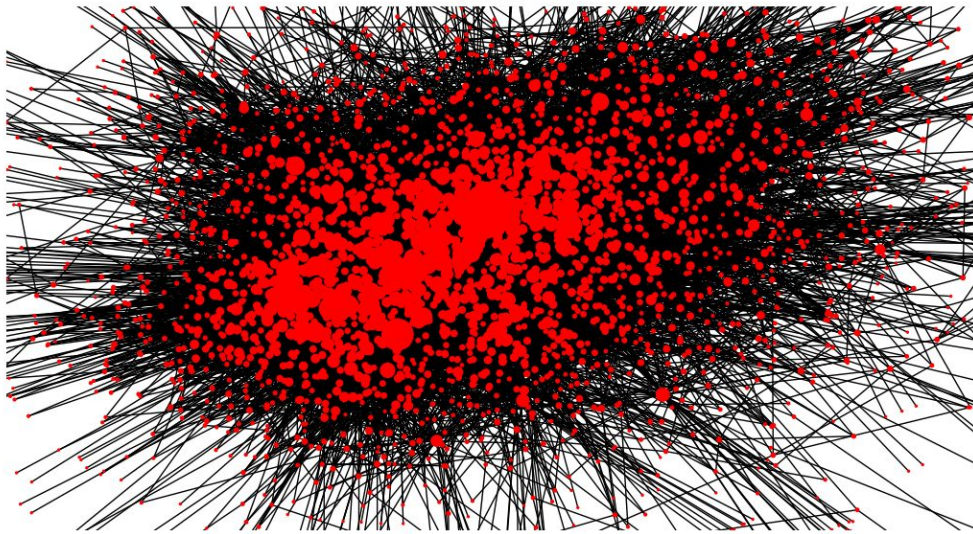


Network

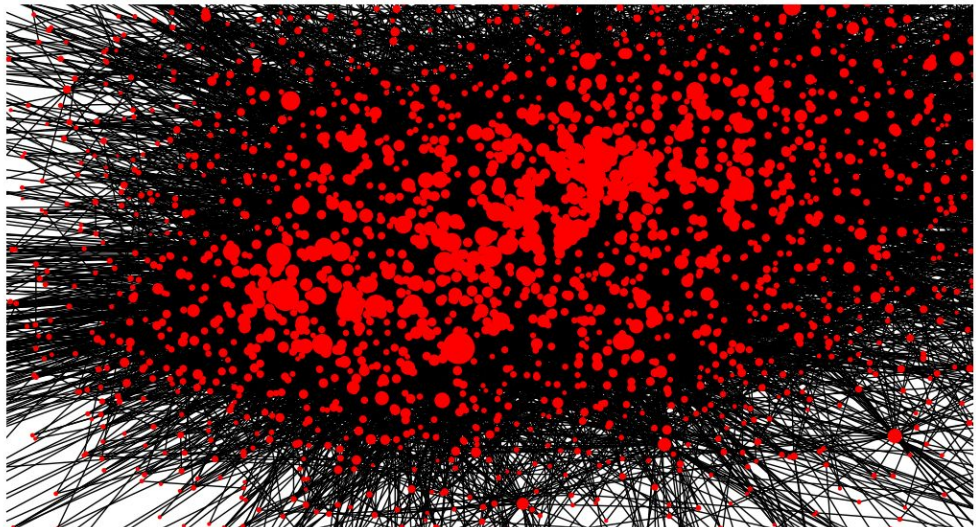




Network

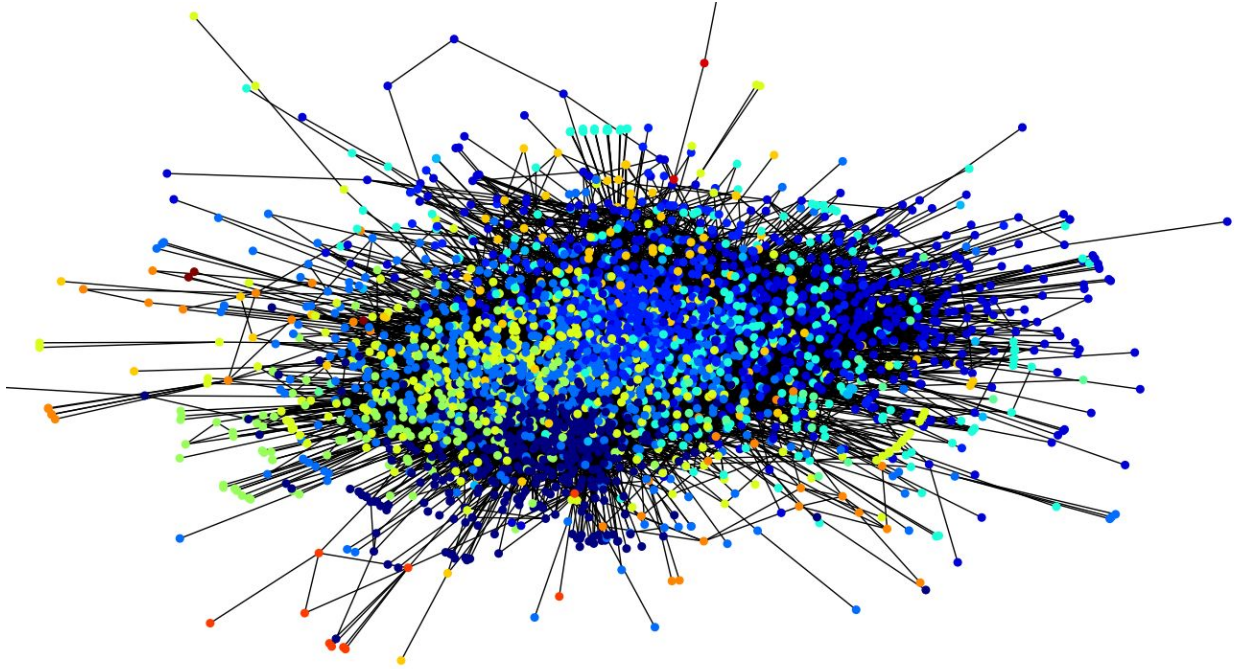


Network

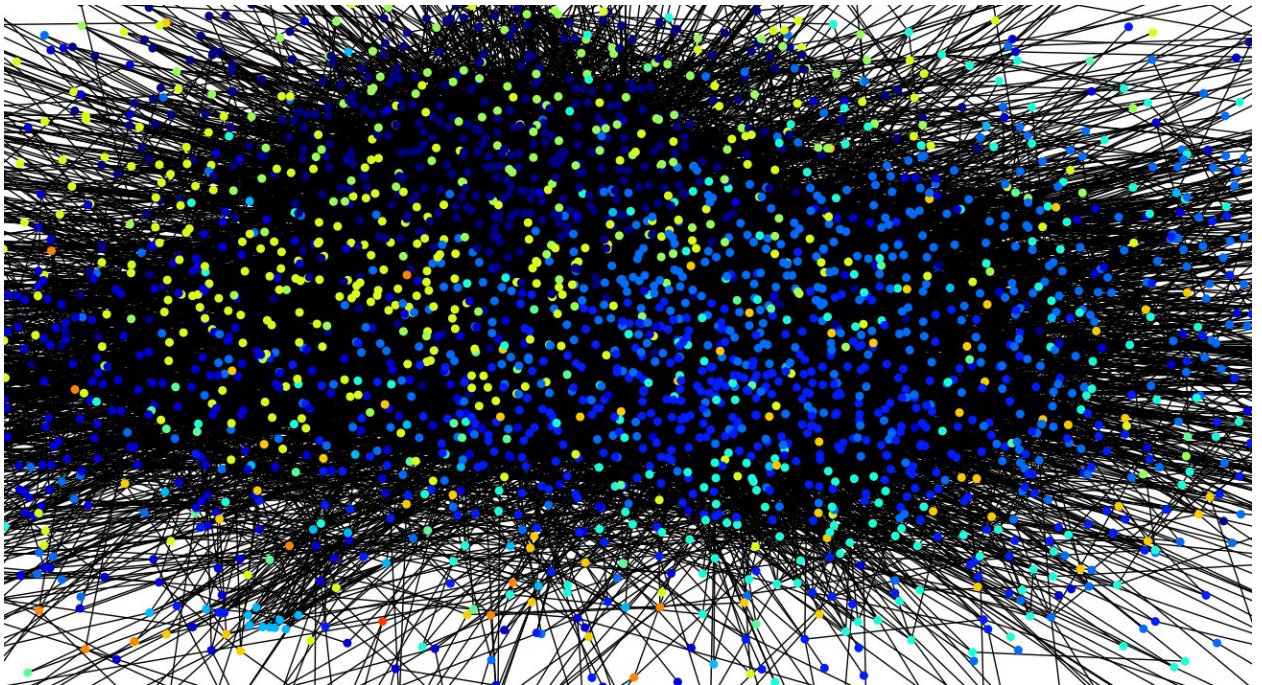




Network modules found as part of clustering modularity : Here the colour of the nodes indicate the module it belongs to. This represents the community structure of the network. Also saved in file modules.txt in the folder. The modules with the nodes that belong to them were further used in gene ontological enrichment analysis. The modules can be seen below.



On zooming further...



Gene ontological enrichment analysis: The file module.txt contains the modules found during module detection while calculating clustering modularity. Each module has a set of genes, by doing this analysis of gene enrichment, we wish to find whether there is a relationship or common bond that binds the genes represented by a particular module or community. We examine 3 things : molecular function, cellular components, biological process. GOrilla, which is an online gene enrichment tool was used for the same. One of the examples is explained below, the other results are attached in the folder.

Suppose we use the genes belonging to a module 10 in module.txt file. It is a set of 9 genes. On using gorilla following results were found.

Set of genes used: 37

Set of genes found common to a biological process : 9

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
<a href="#">GO:0007188</a>	adenylate cyclase-modulating G-protein coupled receptor signaling pathway	2.84E-5	2.95E-2	3.60 (37,8,9,7)	<a href="#">[-] Hide genes</a> GHRHR - growth hormone releasing hormone receptor NPY1R - neuropeptide y receptor y1 NPY2R - neuropeptide y receptor y2 CALCL - calcitonin receptor-like CALCA - calcitonin-related polypeptide alpha CORT - cortistatin MC4R - melanocortin 4 receptor

Here p-value is a null hypothesis test. For eg : if we were to test the hypothesis that a gene does not exist in a module then a lower p-value would imply rejecting the null hypothesis test, that is there are more chances of that gene belonging to that module.

N : Number of input genes

B: is the total number of genes associated with a specific GO term

n:is the number of genes in the top of the user's input list or in the target set when appropriate

b: number of genes in the intersection of N

The common genes to the biological process are mentioned in the last column and the diagram involving them is shown as follows :





