IE590 Complex Systems: Theory and Applications (Fall 2017)
Project for Topic: Complex Networks

First Name: Nidhi
Last Name: Desai
Purdue ID: 0029214803

1. Description of the selected network:

The network chosen for this project is a network of function connections in a brain. The nodes are different regions of the brain and the edges are functional connectivity between these regions. The weight of the edges represents how strong the connection is between the two regions of the brain. Some edge weights can be negative as well, as some brain regions might be negatively functionally correlated. Although two regions in the brain may not have a direct neuronal connection, a functional connection may be inferred based on fMRI time signal correlations [1].

The original network had 164×164 nodes. The data was collected in resting state of the brain. Due to the inability of the current networks performance metrics to deal with negative values of weights, the adjacency matrix was thresholded above zero. Also, to remove insignificant connections from the network, the adjacency matrix was further thresholded at 0.2. After thresholding, five nodes were completely disconnected from any other nodes in the network since the weights for these five nodes were either negative or positive with value less than 0.2. Figure 1 shows the difference in adjacency matrix before and after thresholding. Finally, the network was left with 159×159 nodes.

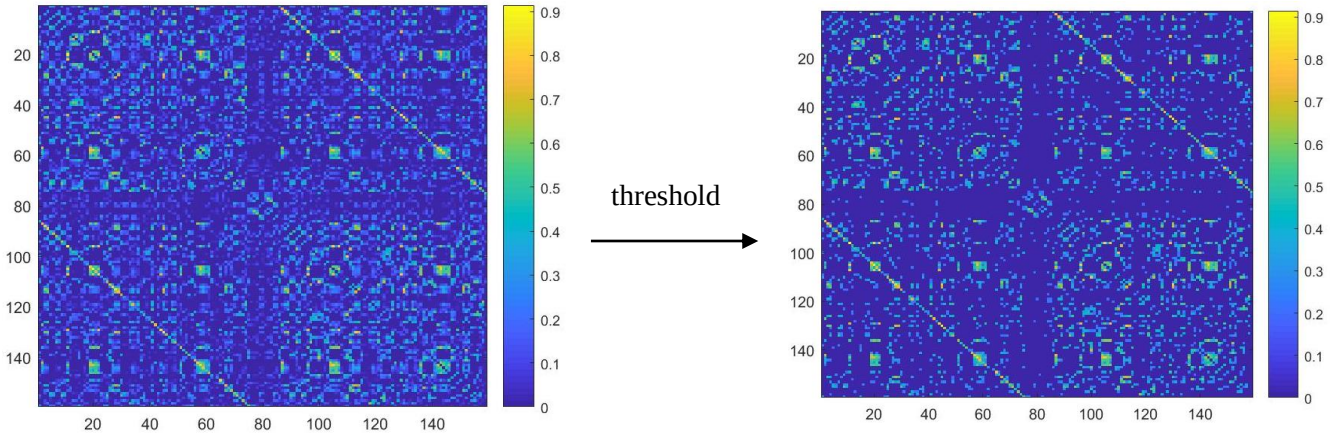


Figure 1: Effect of thresholding on the adjacency matrix

I personally choose this network as my research interest lies in computational neuroscience specifically using network science to study healthy versus disease brains and findings interesting differences between the two which can later be used for diagnosis of those diseases and tracking treatment of the patients.

The aim of this project is to study network properties like clustering coefficient, global efficiency, Q-score, characteristic path length, mean first passage time and leverage centrality with respect to randomized networks for the particular brain network used.

2. Visualization of network:

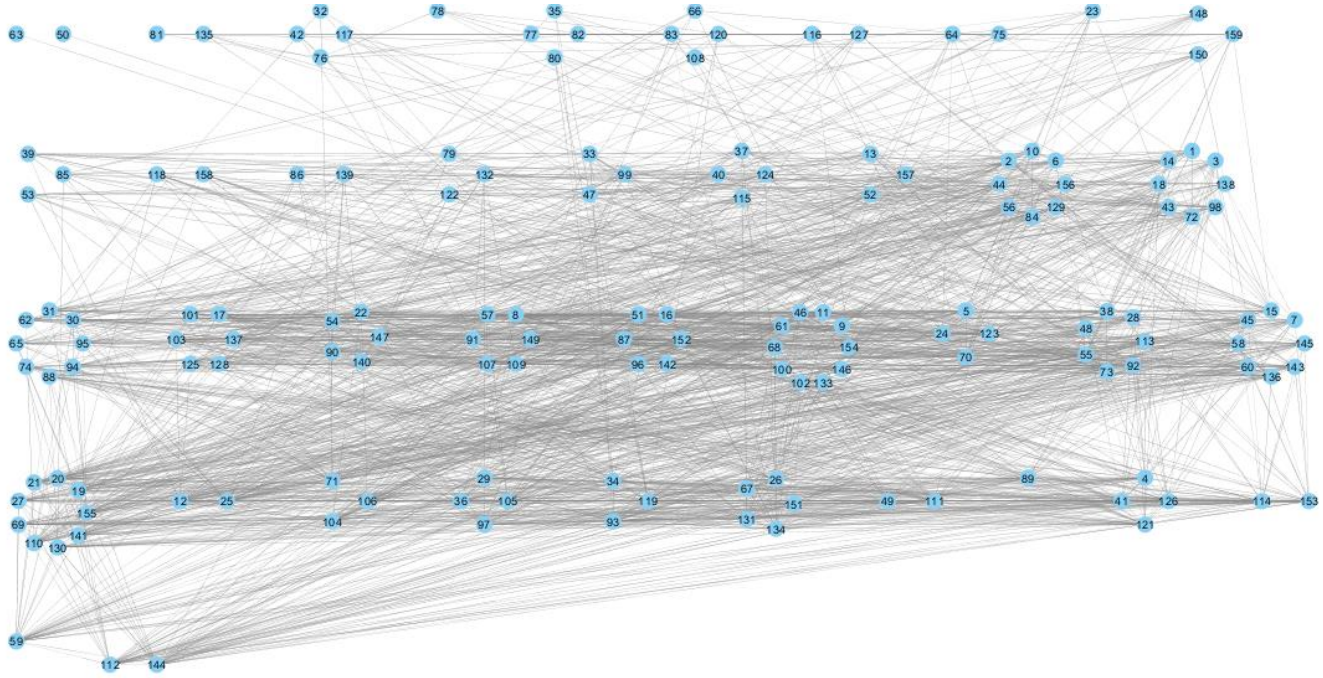


Figure 2: Group attribute layout by degree of nodes

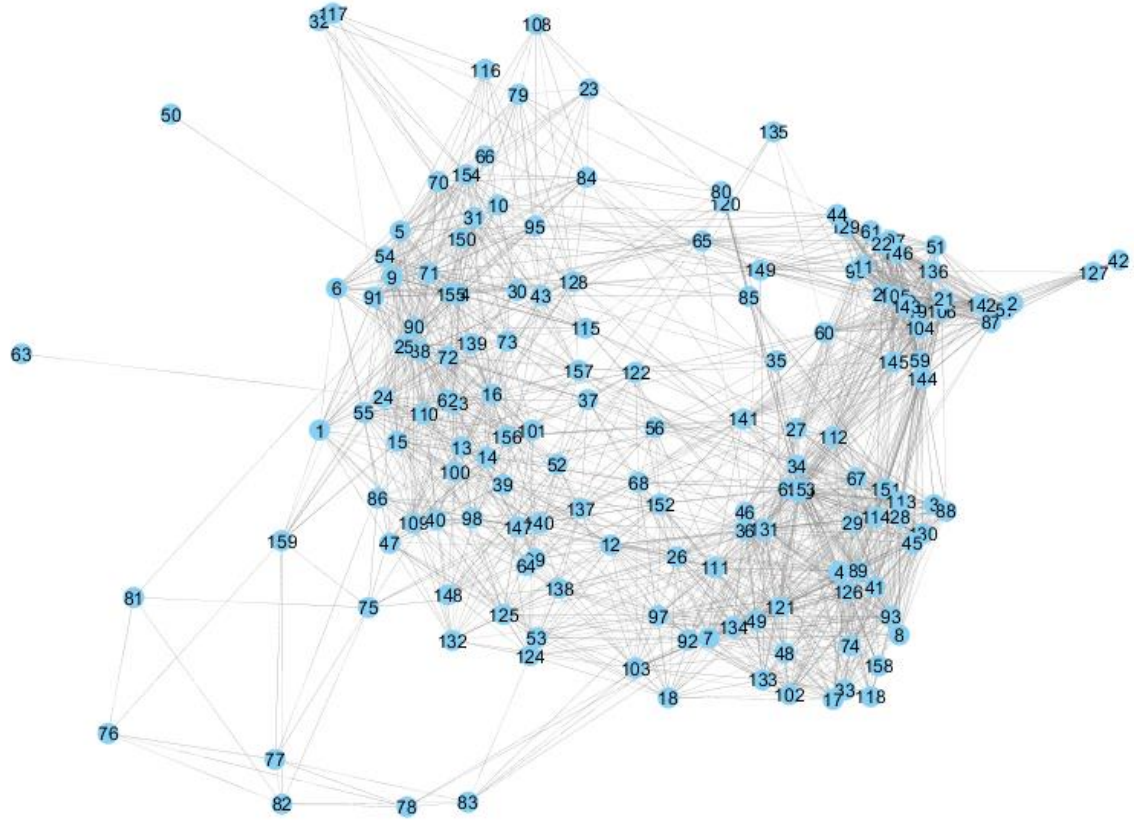


Figure 3: Edge-weighted spring embedded layout by weight of edges

Group attribute layout:

Figure 2 shows the network in the group attribute layout by degree of nodes. It shows the nodes arranged based on their degree. From the bottom of the graph as we move from up right to left the degree of the nodes decreases. The benefit of this layout is that we can easily point to nodes with highest and lowest degrees, which are nodes 112, 144 and 50, 63 respectively. Also, we can see which nodes have same degrees in each cluster.

Edge-weighted spring embedded layout:

Figure 3 shows the network in the edge-weighted spring embedded layout by weight of edges. The spring embedded layout is based on a "force-directed" paradigm. Network nodes are treated like physical objects that repel each other, such as electrons. The connections between nodes are treated like metal springs attached to the pair of nodes. These springs repel or attract their end points according to a force function. The layout algorithm sets the positions of the nodes in a way that minimizes the sum of forces in the network. Therefore, the distance of nodes would an indicator of the weight of that edge.

Degree sorted circular layout:

Figure 4 shows the network in degree sorted circular layout with size of nodes corresponding to their clustering coefficient. Nodes with the highest degree are in the same region and the degree decreases as you proceed counter-clockwise around the circle. The size of the nodes shows the clustering coefficient. It is useful for a quick view of the nodes with the highest clustering coefficient.

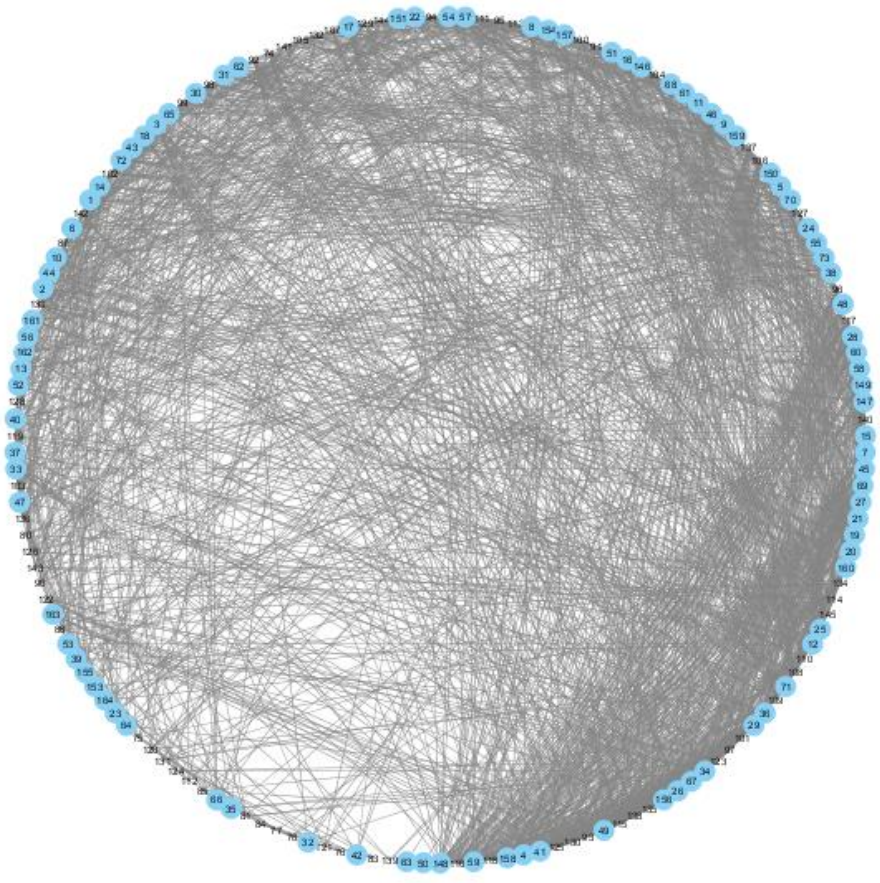
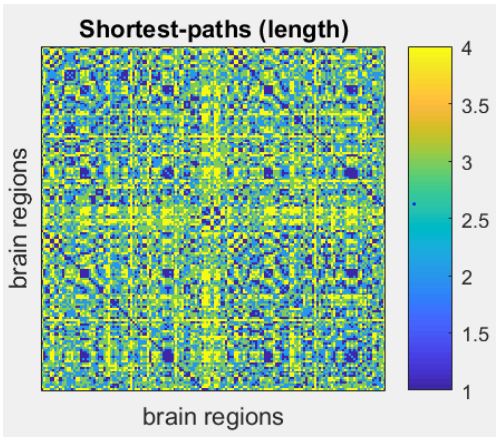
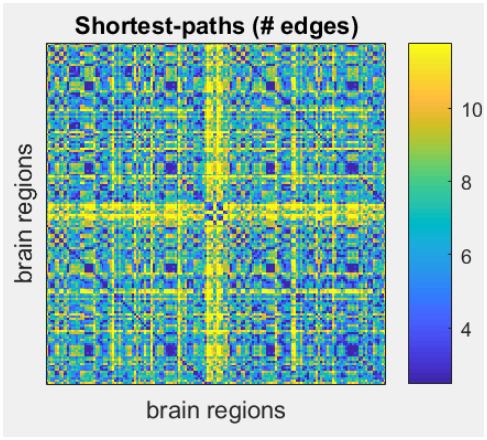


Figure 4: Degree Sorted Circular Layout

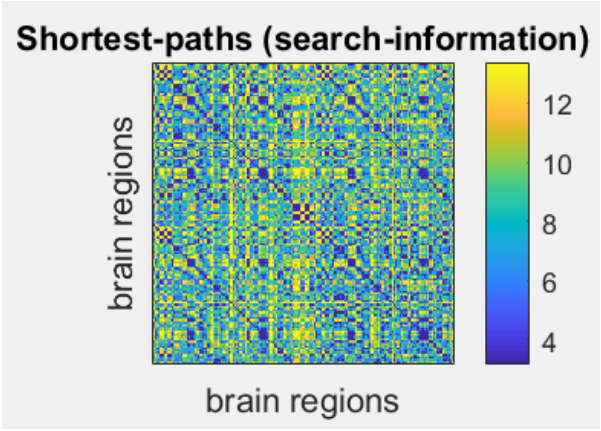
3. Network measurements of integration:

Network integration can be measured via shortest paths calculated in three ways, length, number of edges and search information required. Figure 5 shows the shortest paths between all the nodes. The nodes located in the middle yellow path in figure 5.b indicates that the they need to cross more than eight number of edges to reach each other. This might be the case because as we can see from the adjacency matrix given in figure 1, these nodes are very rarely connected to each other. So, they will be connected through 4 modules, which increase their shortest paths significantly.





Figuer 5: (a) Shortest path in terms of length (b) Shortest path in terms of number of edges



(c) Shortest paths in terms of search information

Figure 5.a gives the shortest path in terms of distance. The minimum distance found is while the maximum distance is 1.0937. We can see similar trend in the shortest path in terms of length and in terms of edges. Figure 6.c also tells us about the shortest path in terms of the information needed to reach from the source to the target.

Measures	Minimum	Maximum
Shortest path (length)	1.0937	17.8461
Shortest path (# edges)	1	7
Shortest path (search information)	1	159

Table 1: Shortest path lengths

4. Network measurements of segregation:

A useful measure in network analysis would be for detecting community structure in networks i.e. appearance of densely connected groups of nodes with only sparse connection between groups. Modularity was one of the measures designed to measure the strength of division of a network into modules. The modularity is, up to a multiplicative constant, the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random [7]. The value of modularity lies between -0.5 and 1 and is positive if the number of edges within groups exceeds the number expected by change.

The penalization factor, gamma, is an input parameter to modularity, which acts like a zooming lens to check the nodes in its view as one module i.e. determining the size of modules. By varying gamma, we change the resolution size while determining the modules in the network. Therefore, we will first study how modularity, denoted by Q-score, changes for different values of gamma as given in Figure 6.

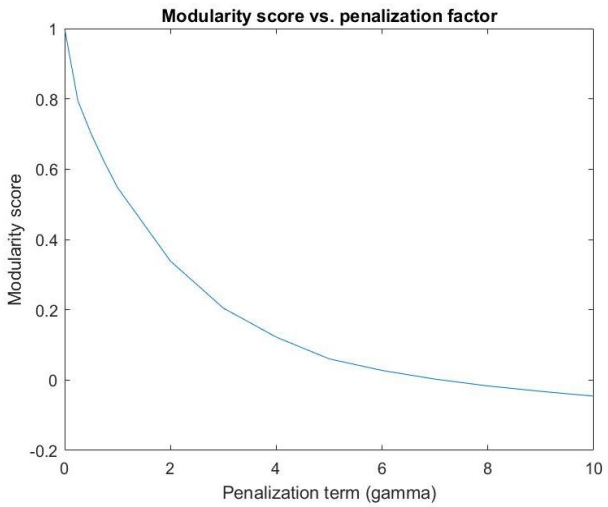


Figure 6: Modularity score versus of gamma

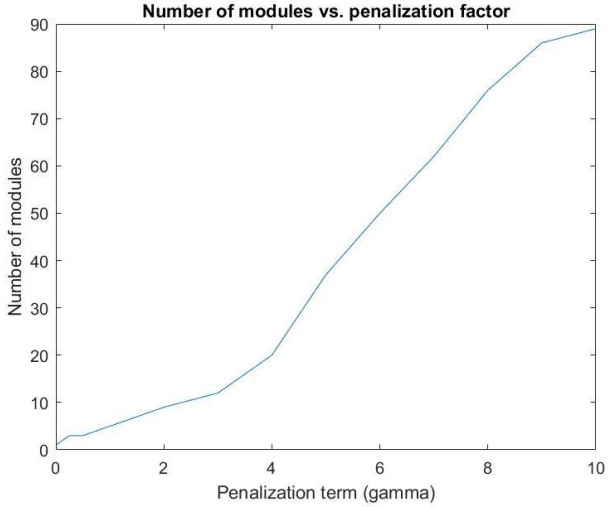


Figure 7: Number of modules versus of gamma

To study the change in modularity gamma was varied as [0;0.25;0.5;0.75;1;2;3;4;5;6;7;8;9;10]. From figures 6 &7, we can see that modularity score decreases with gamma and the number of modules increases with gamma.

Figures 8, 9 & 10 shows the adjacency matrix rearranged based on modules and network with various modules shown in different colors. As gamma was increased more module were formed. This is clear from the adjacency matrix as well as the graph with nodes with more colors for increased gamma.

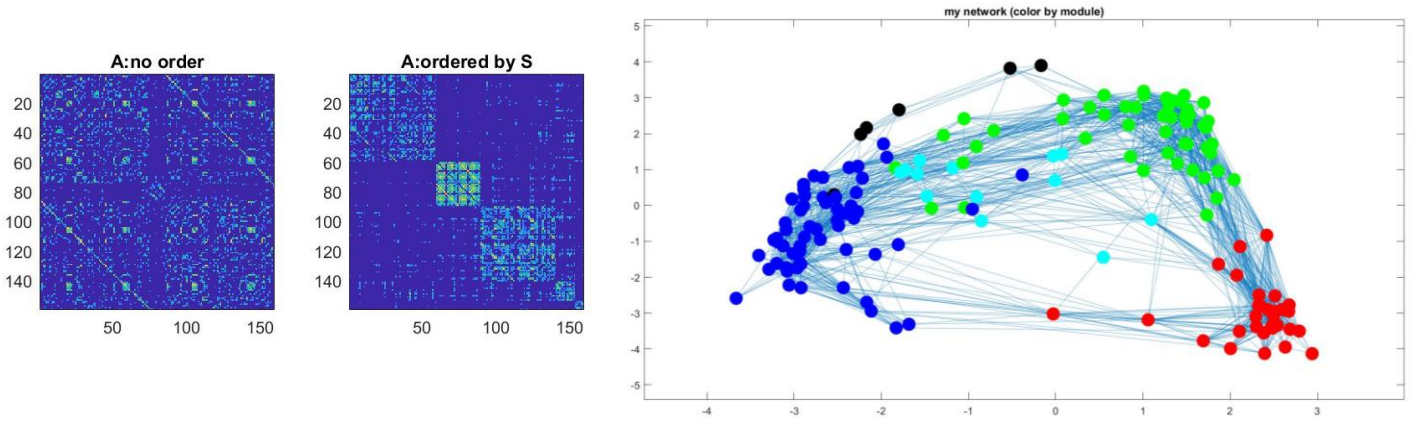


Figure 8: gamma = 3, Q-score = 0.5474, number of modules = 5 (a) Module wise arranged adjacency matrix (b) network visualization with nodes in different modules in different colors.

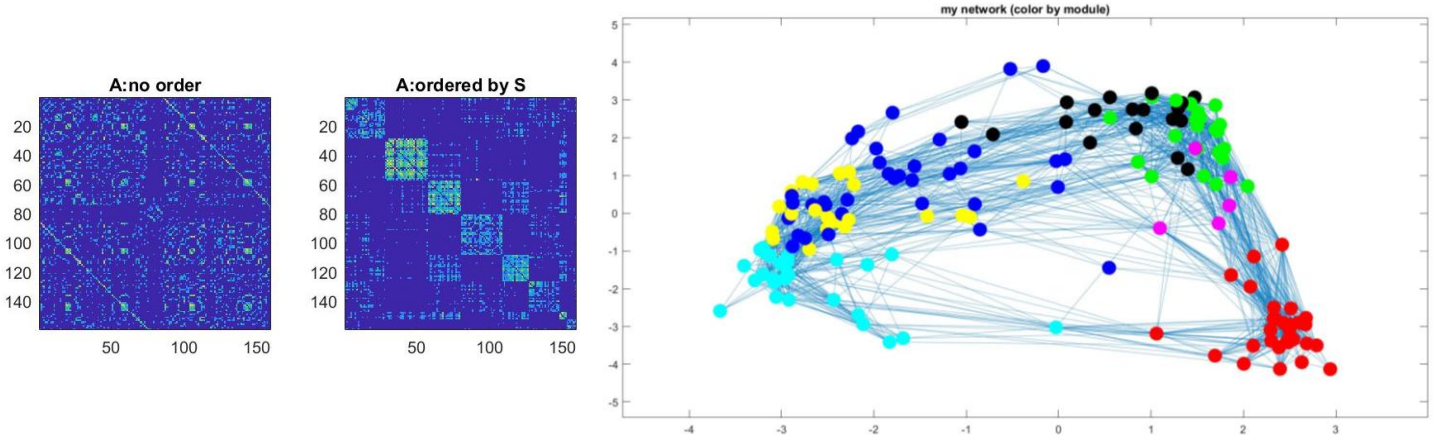


Figure 9: gamma = 4, Q-score = 0.3358, number of modules = 8 (a) Module wise arranged adjacency matrix (b) network visualization with nodes in different modules in different colors.

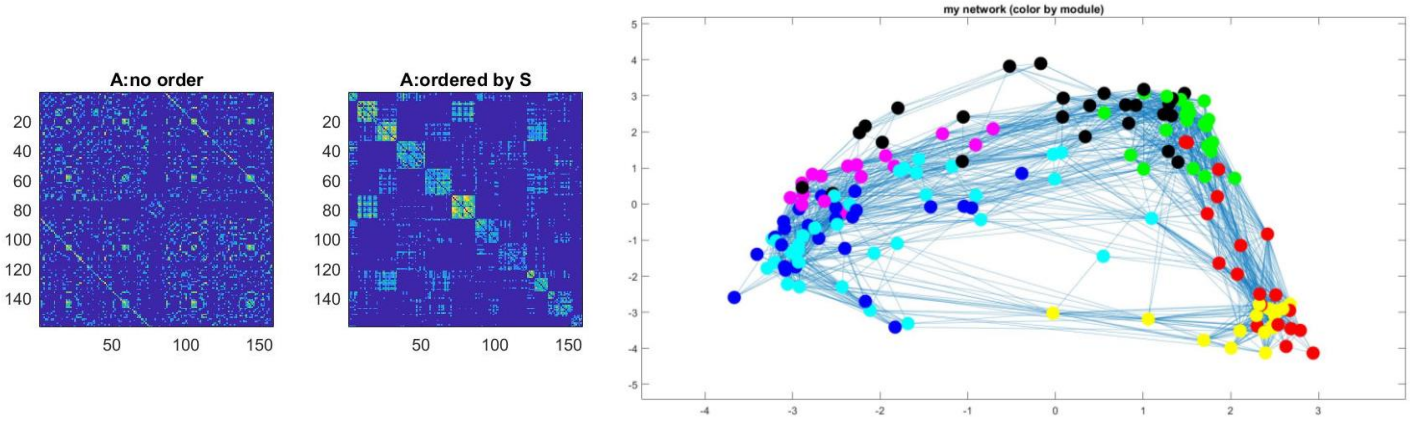


Figure 10: gamma = 5, Q-score = 0.2066, number of modules = 12 (a) Module wise arranged adjacency matrix (b) network visualization with nodes in different modules in different colors.

5. Randomized network:

Randomly generated networks with same topological invariants, namely, degree-distribution, degree, number of edges and number of nodes. This can be done by performing random xswaps of 2 edges. Figure 11 shows the graph of dissimilarity versus number of x-swaps made in the original graph. We can see that dissimilarity almost plateaus after 4000 number of changes, which means that the graph becomes completely randomized after doing 4000 x-swaps in the original graph. Any further x-swaps doesn't lead to changes and hence dissimilarity remains almost constant. Therefore, we choose 4000 number of randomization steps to ensure a fully randomized graph.

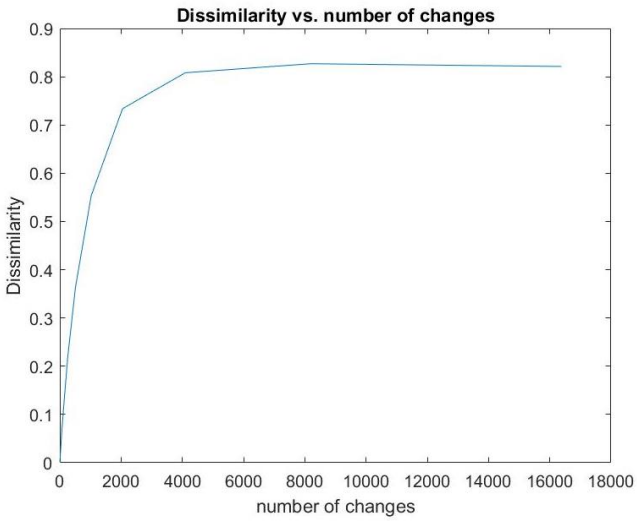


Figure 11: Dissimilarity vs. Number of changes

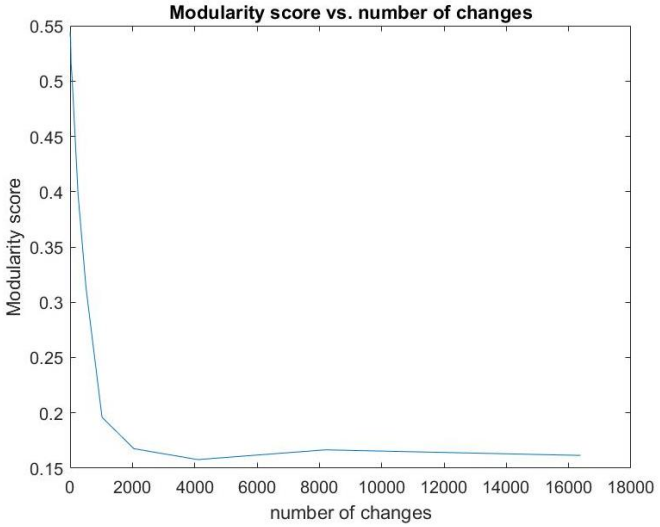


Figure 12: Modularity score vs. Number of changes

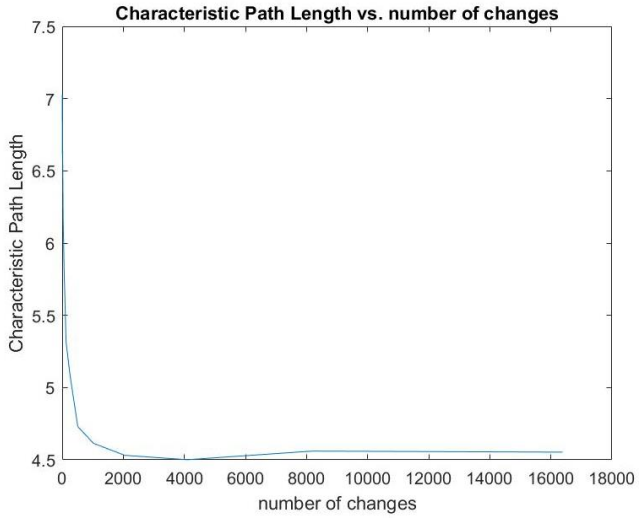


Figure 13: Characteristic path length vs. number of changes

Figure 12 shows the graph of modularity score versus number of changes. It shows that Q-score becomes almost constant after 4000 iterations. There are some fluctuations but not very drastic changes since the network properties doesn't change significantly after certain number of x-swaps. Similar explanation applied to characteristic path length, but an important thing to notice is that it becomes near constant after 2000 iterations itself.

Figure 14.a shows the adjacency matrix of the original network. We can see small clusters of nodes which are highly connected. The main diagonal has zero values, since self-looping was removed. There is a diagonal with highly weighted connections, which is shown by the yellow line. Since the adjacency matrix is symmetric, the pattern in the adjacency matrix suggest the brain is divided into three regions which are connected through a few inter module connections. Inside each region, there seems to be local modules of highly connected nodes.

Figures 14.b, 14.c and 141.d shows the adjacency matrix of randomized networks. We can observe that the local strongly connected nodes in 1.a cannot be observed in the randomized networks. The three main modules can still be observed but there are now more connections between these modules. We can also observe that the three matrices have different patterns of connections, which suggest that different randomizations took place while creating the ensemble of randomized networks.

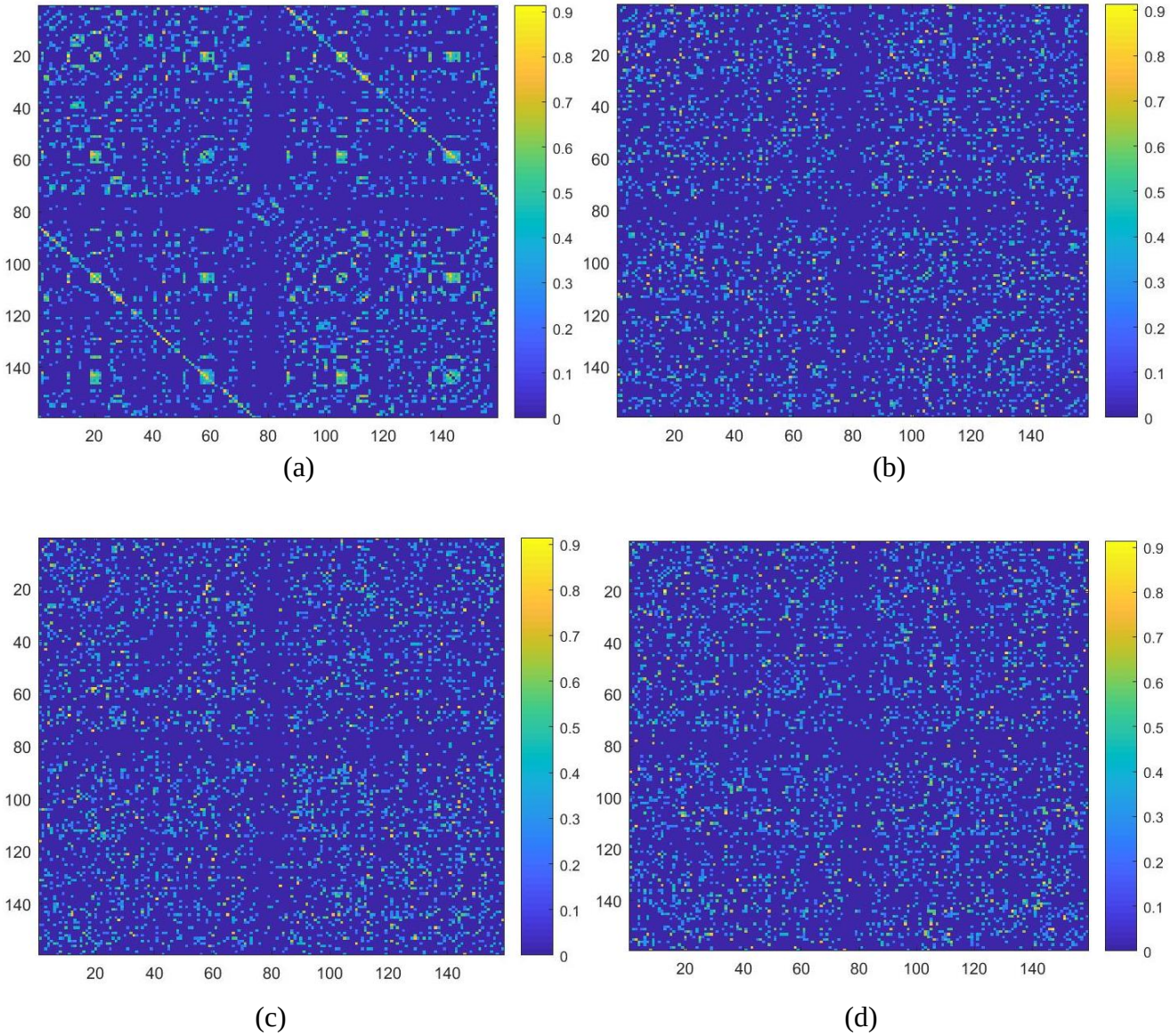


Figure 14: (a) Adjacency matrix of original network compared to (b, c, d) three random networks

6. Network measurements of randomized network:

In this section, we wish to check if our network belongs to a set of random networks i.e. whether these network connections are expected by chance. In order to check this, we will compare our network with 100 randomly generated networks in the previous section. For comparison we will use 5 network measurements. Two measures of segregation, clustering coefficient and Q-score, and three measures of integration, global efficiency, characteristic path length and mean first passage time.

Table 2 gives the values of whole network measurements of the original network and the mean and standard deviation of the ensemble. The z-score is for the null hypothesis that the original network belongs to the ensemble of random networks. The z-score for all the measures is significantly higher than 1.96, which shows that the probability of getting networks with same measures as the original network is less than 5%. Therefore, we can conclude that the original network is statistically different than randomly generated ensemble of networks. Here, we can also notice that the global efficiency of the original network is less than the randomized ensemble.

Whole network measures		**Original network**	**Mean of ensemble**	**Standard Deviation of ensemble**	**Z-score**
Measure of segregation	Clustering coefficient	0.225	0.0686	0.0015	104.2667
	Q-score	0.0617	0.011	0.0016	31.6875
Measure of integration	Global efficiency	0.1729	0.2414	6.1887e-4	-110.6856
	Characteristic path length	7.1456	4.526	0.0243	107.8025
	Mean first passage time	357.049	276.0624	13.4117	6.0385

Table 2: Whole network measures of original network and ensemble

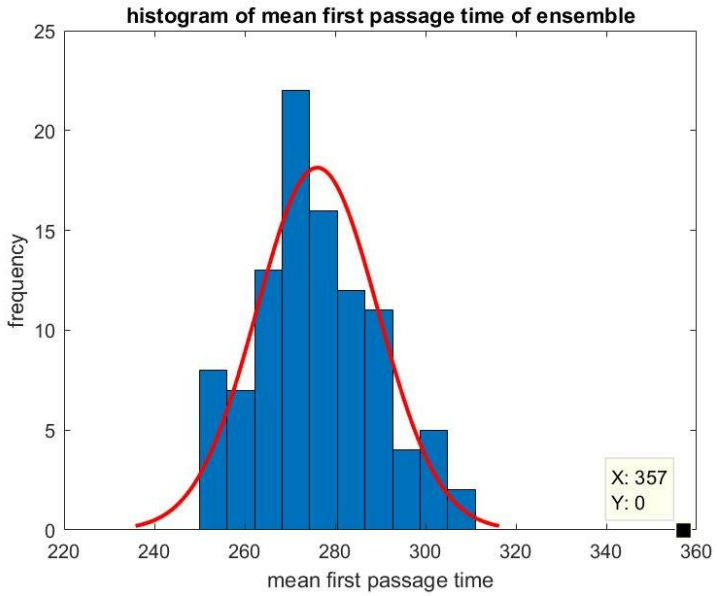


Figure 15: Histogram of mean first passage time of ensemble. The label data point at x=357 shows the value of original network. It is more than two standard deviations away from the mean of the ensemble. It signifies that the original network has network measures statistically different that the ensemble of randomized networks.

7. Network degree distribution follows power law?

Some networks have degree distributions in the form of a power law, i.e. the probability that a node has degree k is given by $Prob(k) \sim k^{-\lambda}$ [6], where λ is called the degree exponent. That is the data points on a log-log scale form an approximate straight line. If we take log of the above equation, we get $\log p_k = -\lambda \log k$. Figure _ shows this log-log plot. We can see that the data points don't lie on a straight line. The fitted equation is as follows:

$$\log(p_k) = 0.2739 + 0.1917 \times \log(k)$$

The R^2 is 0.0638 which is quite low and shows that the log-log data points do not follow a straight line. Therefore, the network doesn't follow a power law distribution.

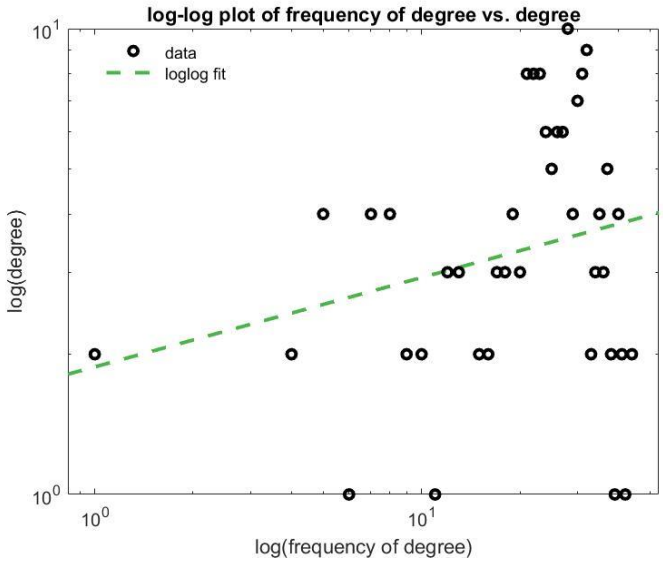


Figure 16: log-log plot of frequency of degree vs. degree. Green line is the fitted power law line.

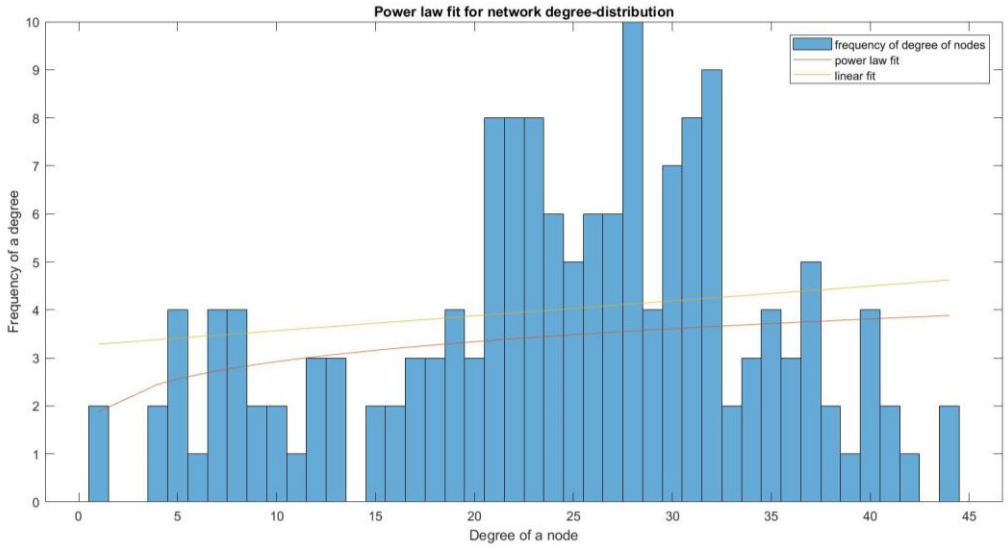


Figure 17: Power law and linear fit on the data points of frequency of nodes versus degree of a node

8. Network shows small world properties?

Small-world networks are defined as networks which are significantly more clustered than random networks but have almost same characteristic path length as random network.

$$S = \frac{C/C_{rand}}{L/L_{rand}}$$

The above formula calculates the measure of small-worldness. A network is defined to be a small world if this measure follows the following rule:

$S > 1\ (C \gg C_{rand}\ and\ L \approx\ L_{rand})$

For the network used in this study and one of the random networks, the values are as follows:

$C = 0.2225\ \ C_{rand} = 0.069\ L = 7.1456\ L_{rand} = 4.5177\ S = 2.0616$

This shows that *S* is greater than 1, and $C \gg C_{rand}\ and\ L \approx\ L_{rand}$. Also, figure _ shows a box plot of *S* calculated with respect to the 100 random networks generated. Therefore, our brain network shows small world properties.

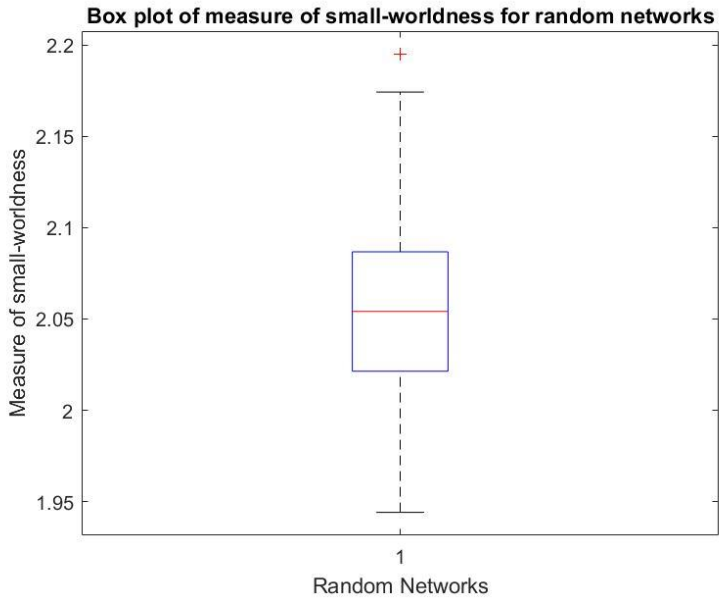


Figure 18: Box plot of measure of small worldness calculated with respect to 100 different random networks.

Small world topology is characterized by dense local clustering of connections between neighboring nodes yet a short path length between distinct pair of nodes due to the existence of relatively few long-range connections. This is a good model for the organization of functional brain networks because a small world network can support both segregated and integrated information processing [5]. Also small-world networks are economical, minimizing wiring costs while supporting high dynamical complexity.

9. <u>Network Measurement – Leverage centrality:</u>

Some of the nodes in the network play a crucial role in mediating a vast number of network connections. Such nodes are central in network organization and are likely to be highly influential over the behavior of the network. Therefore, centrality of the nodes is an important measure while studying brain networks [2]. Central nodes can be found using different centrality measures such as degree centrality, betweenness centrality, closeness centrality, subgraph centrality, page rank centrality and eigenvector centrality.

Degree centrality is the number of edges connecting to a node. This metric assumes that the importance of a node in the network is directly proportional to the number of other nodes it is connected to. An essential node may not necessarily have high number of connections to other nodes in the network [3]. Betweenness centrality considers nodes along the shortest paths to be the most central in the network. It is a good measure for social networks as it can find middlemen. Brain networks most likely process information via multiple neighbors (parallel routes) and is not restricted along shortest paths (unrestricted walks) [2]. Therefore, betweenness centrality is not the best measure for brain networks. Some other measures used are closeness centrality, which is the mean distance between a node and all other nodes in a graph, and subgraph centrality, which measures the number of closed walks beginning and ending at a node. Eigen vector centrality is given by the sum of the values with the principle eigenvector corresponding to the direct neighbors. In contrast to degree centrality, eigenvector centrality not just favors nodes with high correlations but also specifically favors nodes that are connected to nodes that are themselves central [4].

Here we will study the centrality metric called leverage centrality. Leverage centrality is based on the degree of a node relative to its neighbors. The assumption here is that a node is central if its immediate neighbors rely on it for information [2]. Leverage centrality captures nodes which are more connected than its neighbors i.e. it controls the amount of information received by its neighbors. It is different than degree centrality in the sense that a high degree node is not highly central according to leverage centrality if its neighbors also have high degree [2].

Leverage centrality is a measure of the relationship between the degree of a given node and the degree of each its neighbors averaged over all neighbors. Leverage is defined on the interval (-1, 1) and is given by the following equation:

$$l_i = \frac{1}{k_i} \sum_{N_i} \frac{k_i - k_j}{k_i + k_j}$$

where,

l_i = Leverage centrality,

k_i = degree of a given node

k_i = degree of its neighbors

N_i = Number of neighbors

A node with negative centrality is influenced by its neighbors since the neighbors interact with far more nodes. A node with positive leverage centrality influences its neighbors since the neighbors have far fewer connections. Since leverage centrality is a node wise measure, its value for all the nodes for our network cannot be mentioned here. Therefore, the following network graph shows each node's size varying with its value of leverage centrality.

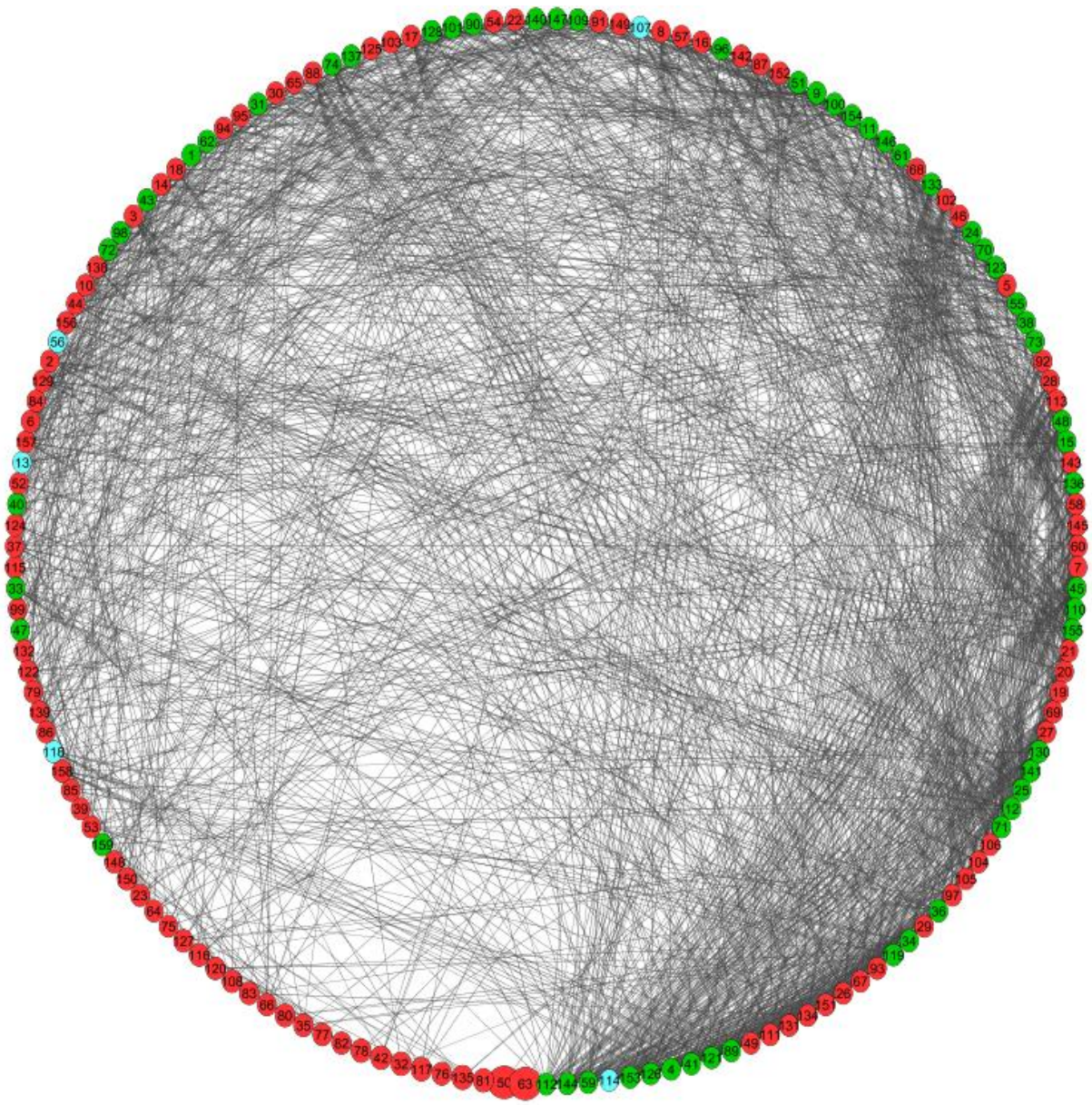


Figure 19: Degree sorted graph of network with size of nodes corresponding to the values of leverage centrality and color corresponding to the sign of leverage centrality. Red color of the node means that the leverage is negative, and they are influenced by their neighbors. Green color of the node means that the leverage is positive, and they influence their neighbors. Blue color of the node means zero leverage.

From the graph of the network, we can observe that the nodes with the highest values of absolute leverage centrality have negative leverage i.e. they are influenced by their neighbors and have the lowest degree. For nodes with large degrees, leverage values are both negative and positive, which shows that even nodes with high degree can not be an influencer. Therefore, leverage centrality, even though comes from degree of nodes, gives more information of most influential nodes in the network. Leverage centrality relates to degree in the sense that nodes with small degrees are more likely to be influenced by neighbor nodes than nodes with higher degrees.

Refences:

1. Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, et al. (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proceedings of the National Academy of Sciences 102: 9673–9678.

2. Joyce KE, Laurienti PJ, Burdette JH, Hayasaka S. A New Measure of Centrality for Brain Networks. Sporns O, ed. PLoS ONE. 2010;5(8):e12200. doi:10.1371/journal.pone.0012200.

3. Xi-Nian Zuo, Ross Ehmke, Maarten Mennes, Davide Imperati, F. Xavier Castellanos, Olaf Sporns, Michael P. Milham; Network Centrality in the Human Functional Connectome, *Cerebral Cortex*, Volume 22, Issue 8, 1 August 2012, Pages 1862–1875, https://doi.org/10.1093/cercor/bhr269

4. Lohmann, Gabriele, et al. "Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain." PloS one 5.4 (2010): e10232.

5. Bassett, Danielle Smith, and E. D. Bullmore. "Small-world brain networks." The neuroscientist 12.6 (2006): 512-523.

6. Bullmore, Ed, and Olaf Sporns. "Complex brain networks: graph theoretical analysis of structural and functional systems." Nature Reviews Neuroscience 10.3 (2009): 186-198.

7. Newman, Mark EJ. "Modularity and community structure in networks." Proceedings of the national academy of sciences103.23 (2006): 8577-8582.