Network Analysis on a Unified Human Brain Connectome

MSc Complex Networks and Web
Department of Computer Science, University College London

Abstract

This research paper will first discuss the processes involved in obtaining brain networks and importance of applying network analysis on them. It will then introduce a unified anatomical network from 477 individuals and outline the methodologies required to understand how the network varies as it becomes progressively generalised. The analysis will reveal a scale-free network for a rather generalised brain network and a small-world network for a relatively specific brain network. Finally, this paper will conclude by giving a few hypotheses on why the network topology changes when it becomes more generalised.

Table of Contents

1. Executive Summary

Brain connectomic analysis has become an extremely important and viable process in today's era owing to the advances in medical technology for mapping and estimating Gray Matter (GM) brain regions and White Matter (WM) fiber tracts respectively. I believe that a thorough analysis on the fundamental network structure of healthy (resting-state) brains will eventually reveal a model which can be used as a reliable comparison against other brain networks, such as that of 1) individuals with brain diseases and 2) individuals involved in various levels of cognitive tasks/action.

Hence, I have chosen a dataset taken from the Budapest Connectome Project where it contains a parameterizable and unified resting-state anatomical network from 477 healthy individuals. This project primarily uses the built-in Python library called Networkx to investigate how the network varies as it becomes progressively generalised. After doing a preliminary assessment of the dataset, these are the 2 questions which I will attempt to answer in this project:

1. How does the network change as we alter the percentage of brains used in the unified brain network?
2. Does the brain network resemble any model? If yes, is there any plausible reason for them to resemble the model?

The analysis reveals a scale-free network for a progressively generalised brain network and a small-world network for a progressively specific brain network. This is seen from a concurrent increase in small-world characteristics such as functional segregation and integration when the brain network becomes more specific.

My first hypothesis was that when the unified network becomes progressively specialised, appearing edges tend to form clusters for task specialization and yet preserve a low path-length between clusters to facilitate information transfer. My second hypothesis was that these new edges and clusters might be specific to individuals depending on factors such as upbringing and social life because they are not present in the relatively generalised networks.

2. Background

2.1. Building Connectomes using diffusion MRI

There are two main types of brain networks: anatomical and functional. In this project we will focus on the anatomical (or structural) network. The process of generating an anatomical connectome network can be broken down into two main steps: node delineation and edge mapping.

Nodes in the connectome represent spatially distinct **Gray Matter** (GM) regions and are obtained by a process called parcellation. Edges represent the **White Matter** (WM) fiber bundles that interconnect pairs of regions. Diffusion-MRI tractography is the process typically used to generate voxel curves and estimate these edges.[1]
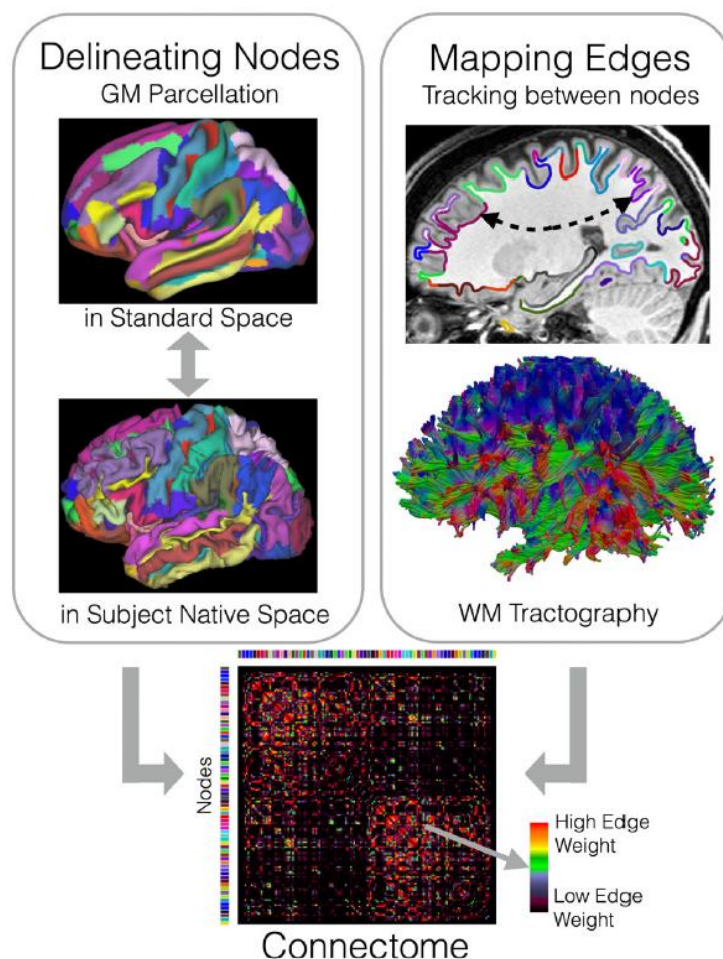

Figure 1: Process of obtaining connectome

2.2.   Motivation for Analysing Brain Networks

Since the word "Connectome" (comprehensive map of neural connections in the brain) was introduced in 2005, the number of scientific articles devoted to connectomes has risen continuously and exponentially. The emerging field of connectomics have and will continue to deliver insights into the principles that underlie network architecture and uncovers how these principles support network function.[2]

2.3.   Project Overview and Research Questions

This project will analyse merged brain networks and the common edges between them. The idea of a "Consensus Connectome" refers to the combination of connectome networks that share a common set of edges (refer to figure 7 in appendix).

The term **"edge confidence"** or **"confidence level"** will be used frequently in this report. It refers to the percentage of connectomes that contains a set of edges (e.g. the consensus connectome network with edge confidence of 40% will show edges that are present in the connectomes of 40% of the subject pool). In other words, the higher the edge confidence, the more generalised the connectome becomes.

The research questions that this project aims to answer are listed below:

1. How does the network change as a function of confidence level?
2. Does the consensus connectome resemble any model? If yes, is there any plausible reason for them to resemble those models?

3. Literature Review

3.1. Complex Network Measures of Brain Connectivity

The author talks about several measures of the brain network. This project will focus on 3 of them, functional segregation, functional integration and centrality:

High functional segregation signifies a better ability for specialized processing to occur within densely interconnected groups of brain regions. It can be characterized by 1) large number of *triangles* in the network, 2) a high *average clustering coefficient*, 3) high *transitivity* and 4) distinct *communities*.

High functional integration signifies better communication between different groups of brain regions and can be indirectly measured by the concept of a path. This can be characterized by 1) *low characteristic path length* or 2) *high global efficiency*. The author goes on to emphasize that global efficiency would be a superior metric because it can be meaningfully computed on disconnected networks as well.

High Centrality can be characterized by 1) a high *degree centrality* and 2) a high *closeness centrality*. Many measures of centrality are based on the idea that central nodes participate in many short paths within a network, and consequently act as important controls of information flow.[3]

Here, the author explains how the metrics are correlated to the cognition processes in the brain. It would be worthwhile to investigate how these measures fair as a function of edge confidence.

3.2. Development of the Connections in the Human Brain

The authors have observed that as the minimum edge confidence decreases, the order of the appearance of the new edges is not random at all. Newly appearing edges are usually connected to the already existing edges.

The assumption is that connections that are present in almost all braingraphs developed first. When newer connections were developed, neurons whose connections were disconnected from these oldest neurons were eliminated. Only neurons which were connected to the building network survived. It was hypothesized that the newer the connections, the fewer the subjects have those edges.

This assumption is in line with the "competition hypothesis" of the brain development (if the axon of the cell does not contact the dendrites of the cell in its target area it will die).[4] This research paper inspires the project's direction to analyse whether the network topology of the consensus connectome changes as a function of edge confidence.

3.3. Doubly Preferential Attachment Simulation of the Consensus Connectome

The authors have shown that the doubly-preferentially attached model well-describes the Consensus Connectome Dynamics (CCD) phenomenon. This fact also strengthens their hypothesis that the CCD phenomenon mimics axonal developments on a macroscopic level, because new edges appear more probably between already high-degree nodes.[5] This project will investigate whether this pattern is apparent for metrics other than node degrees. (e.g. will new edges appear more probably between already high-centrality nodes)

4. Methodology

4.1. Data Acquisition and Evaluation

The datasets were acquired from the Budapest Connectome Project[6][7] which parameterized brain connectomes across 477 individuals in its latest version (v3.0). There are a few reasons for choosing this dataset for the purpose of this project:

1. Toggling the edge confidence parameter allows researchers to view only edges that are present in a certain percentage of the subject group. This greatly aids in the study of Consensus Connectome Dynamics (CCD)
2. The number of nodes represents Regions of Interest (ROI) of the brain instead of individual neurons. Networks with neurons as nodes are very temporal, providing very little insights.

File Type	CSV files
Data Type	Undirected and Weighted Graphs
Number of Connectomes	477
Nodes	841
Node Representation	GM Regions of Interest (ROI) within the brain
Edge Representation	Fiber tracts
Edge Weight	Average length of fiber
Population	All (including male and female)
Edge Confidence	[20%, 30%, 40%, 50%, 60%, 70%]

Table 1: Data Information

4.2. Data Processing

The Networkx library in python is used to analyse the Budapest Connectome datasets mentioned above. The library is capable of reading and writing various file types such as adjacency matrix, edge list and csv files, making it very flexible and robust. The library also has many built-in measurement functions and generative models which greatly eases the analysis.

The CSV files are first imported as Pandas dataframes first using *pd.read_csv('filename')*. The dataframes are then imported as undirected Networkx Graph objects using *nx.from_pandas_edgelist(dataframe, create_using=nx.Graph())*.

After obtaining the graph objects with different edge confidence, missing nodes are added to graphs with higher edge confidence such that all graphs have the same number of nodes before any analysis.

4.2.1. Network Measurements

1. For functional segregation:
   o Measure average clustering coefficient for each graph using *nx.average_clustering_coefficient(G)*
   o Measure transitivity for each graph using *nx.transitivity(G)*
2. For network resilience:
   o Plot the degree distribution of each graph using a log-log graph
   o Measure degree assortative coefficient for each graph using *nx.degree_assortative_coefficient(G)*
3. For centrality:
   o Measure degree centrality for each node using *nx.degree_centrality(G)*
   o Measure closeness centrality for each node using *nx.closeness_centrality(G)*
   o Plot box plots for the degree centrality and closeness centrality values
   o Place all closeness centrality values for various edge confidences in a dataframe
   o Order the dataframe based on the closeness centrality values of the graph with 20% edge confidence
   o Plot the closeness centrality values of sorted nodes for all edge confidences (see figure 3)

4.2.2. Bitmap of Sorted Adjacency Matrix (BOSAM)

1. Calculate node degree of each node using *nx.degree(G)*
2. Obtain all the neighbors of each node using *nx.neighbors(node)* and append them into a python dictionary
3. Obtain the degree for each node's neighbors by referencing it from the python dictionary above
4. Obtain the maximum neighbor's degree for each node by using the Python's built-in function *max()*
5. Store nodes, node degrees and maximum degree of neighbors in a dataframe
6. Sort dataframe based on their degrees
7. For nodes with the same number of degrees, sort them based on their maximum degree of neighbors
8. Convert the graph to adjacency matrices with the pre-determined node sequence and plot the matrix (see figure 5)

5. Results

This section aims to outline some interesting observations regarding the consensus connectome as edge confidence level is altered. The table below shows the overview of the networks.

	confidence	edges	density	diameter	transitivity
0	20%	5624	0.016	9	0.313
1	30%	3056	0.009	11	0.226
2	40%	1730	0.005	13	0.164
3	50%	989	0.003	20	0.120
4	60%	574	0.002	14	0.093
5	70%	336	0.001	10	0.082

Table 2: Overview of graphs with different edge confidence (all the graphs are normalised with a node size of 841)

Results for various measurements suggest that as confidence level goes past 50%, the trend discontinues and inconsistent such that it is very difficult to derive any useful insights. One of the examples can be seen in the figure below where the median values for closeness centrality drop to 0 after 50%. There could be a few reasons for this phenomenon:

1. there is simply no observable fundamental anatomical structure for the consensus connectome in more than 50% of the subject population
2. the current diffusion-MRI techniques to estimate edges are not consistent enough in picking up the fiber tracts that are supposed to be present
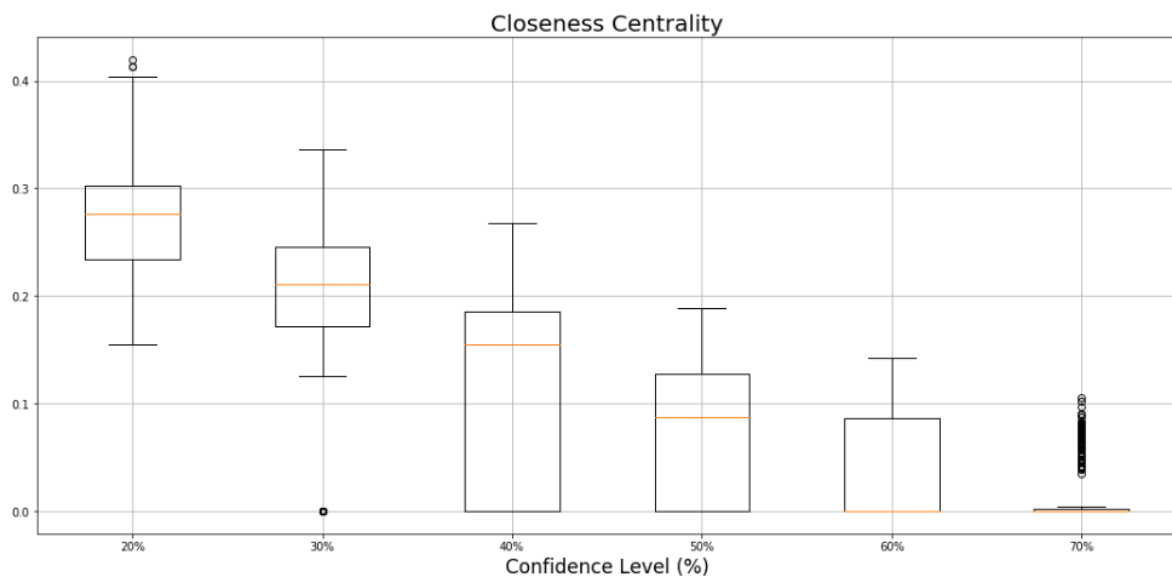

Figure 1: Statistics for Closeness Centrality

5.1.   Small World Nature of the Consensus Connectome

Clustering coefficient and global efficiency are used to indirectly measure the network's functional segregation and integration respectively. Both shows an upward trend as confidence level decreases. This suggests that as more edges start to appear, the network favours high clustering and shorter path lengths between nodes.
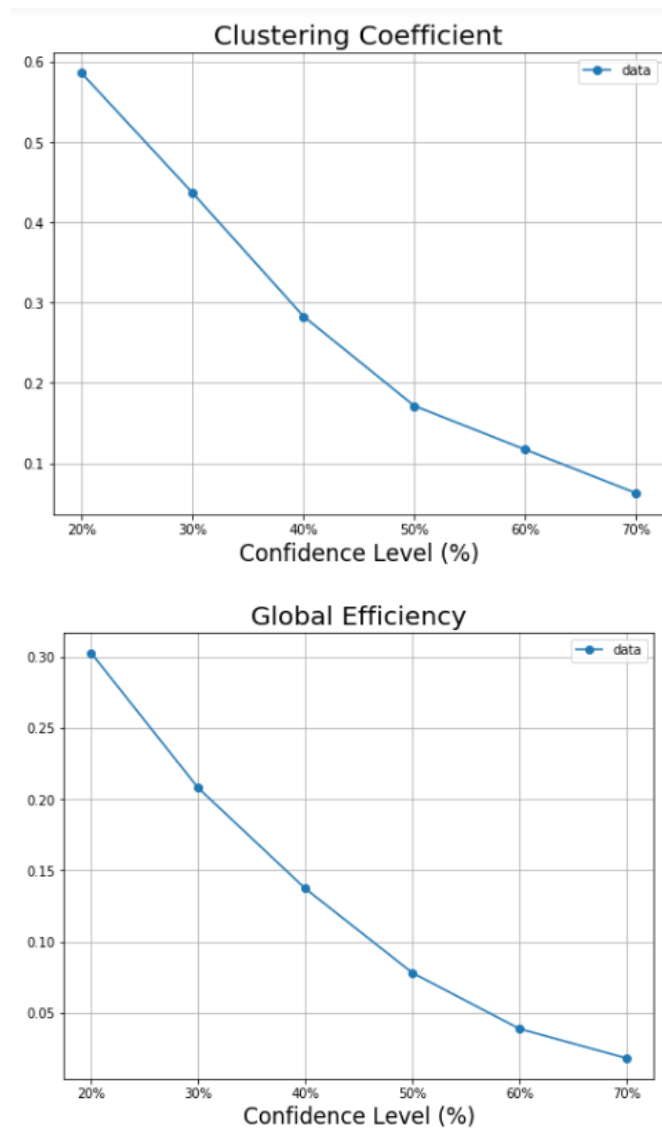

Figure 2: Clustering Coefficient and Global Efficiency

5.2.   Scale-Free Nature of the Consensus Connectome

A closer look at closeness centrality of nodes reveal that when the minimum edge confidence is reduced and more edges appear in the consensus connectome, nodes that originally already have high closeness centrality values see their values increase even more. This means that new edges not only prefer nodes with high degrees (as mentioned above in the literature review), but also nodes with high closeness centrality.
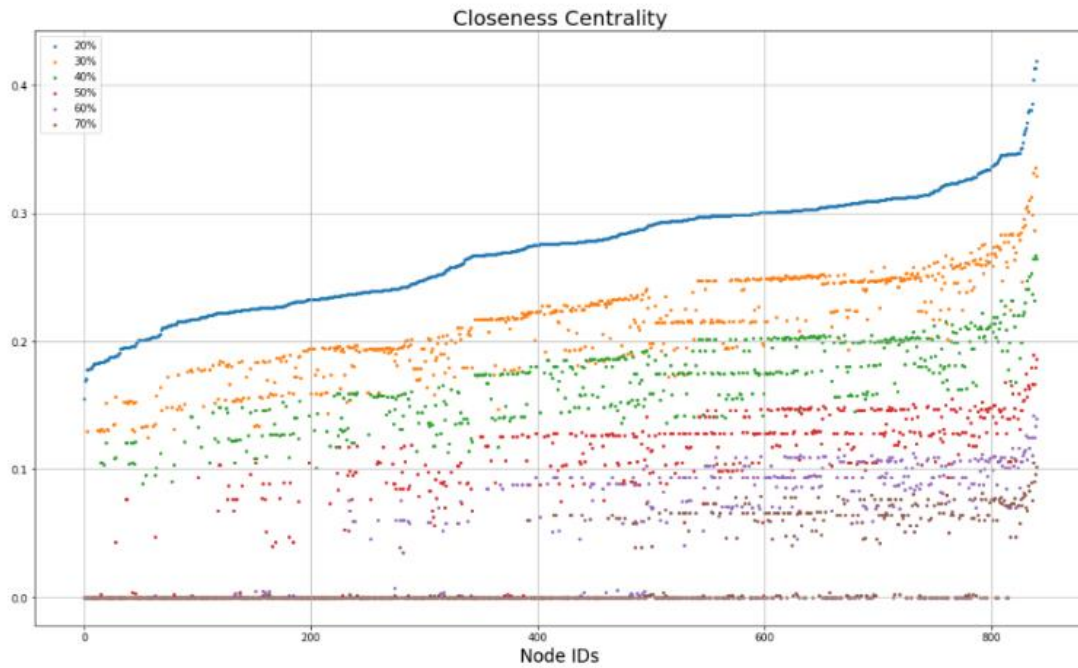

Figure 3: Nodes are sorted based on closeness centrality values of the graph with 20% edge confidence

This probably suggests that fiber tracts in the connectome tend to exhibit preferential attachment. This can be further testified when the degree distribution of the graph (at 50% confidence level) is comparable with that of a scale-free power law graph.
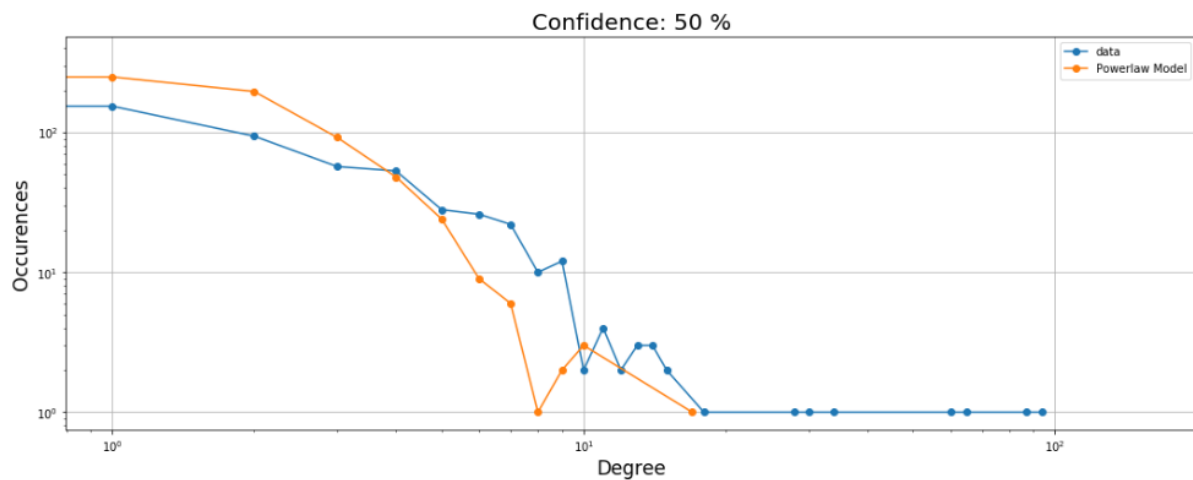

Figure 4: Degree Distribution plotted on a log-log scale (50% edge confidence level)

5.3.   BOSAM Visualisation

A zoomed in view of the BOSAM graph shown below reaffirms the hypothesis that the consensus connectome resembles that of a scale-free model at a reasonably high confidence level (50%). The BOSAM Visualisations (figure 8) also showed that this scale-free model slowly disintegrates as confidence level decreases and more edges start to appear. This could be owing to the network favouring a high clustering coefficient and high global efficiency, tending towards a small-world network.
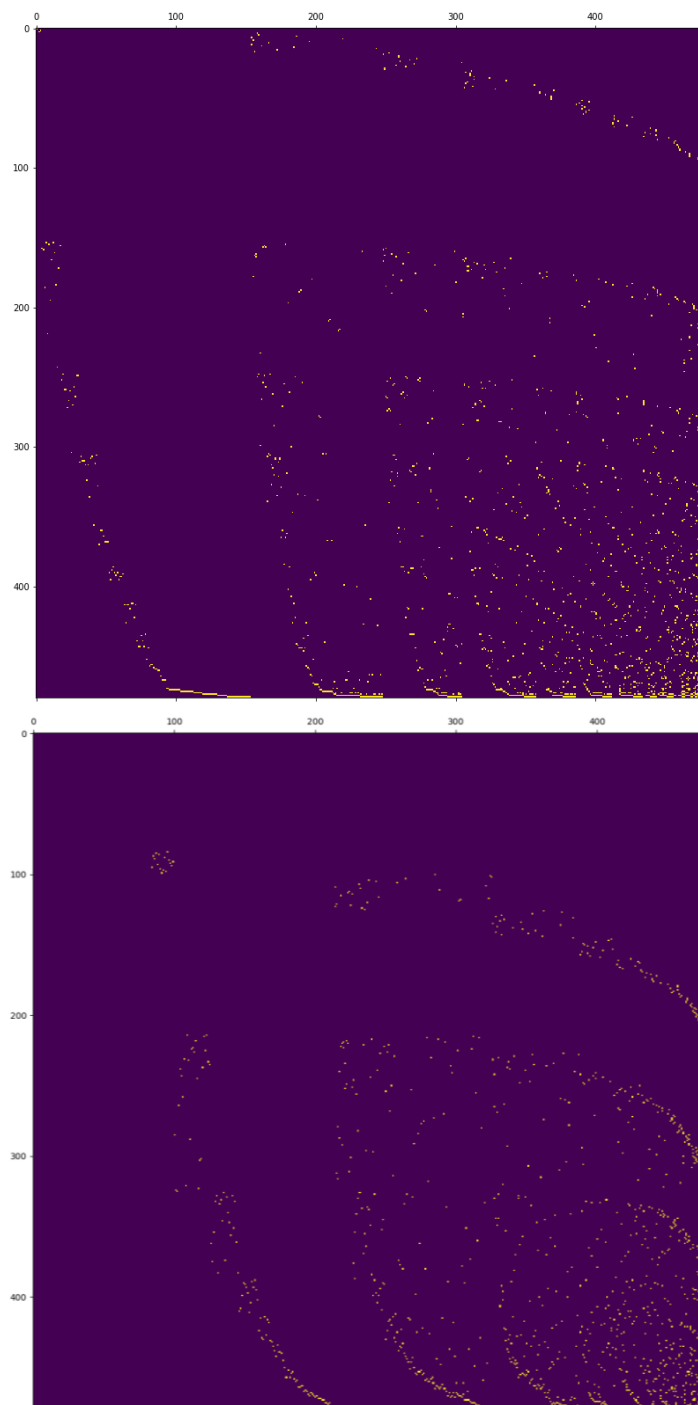
Figure 5: Top image shows BOSAM visualisation for consensus connectome with edge confidence of 50%. Bottom image shows BOSAM visualisation for network generated by power law sequence

6. Conclusion

6.1. Discoveries

The optimal range of edge confidence to analyse the consensus connectome is 20% - 50%. A graph with confidence level lower than 20% would be too specific towards a few individuals and not be such a good representation of the whole subject pool. On the other hand, a graph with confidence level higher than 50% is too restrictive to derive any useful insights

At higher edge confidence level, the network is seen to resemble a power law network, exhibiting scale-free characteristics. At lower edge confidence level, the network is seen to favour a higher functional segregation and integration, exhibiting more small-world characteristics to facilitate quick information transfer

6.2. Hypotheses

When new edges develop from the generalised connectome, they can be added across low-degree nodes to form clusters for **task specialization** and yet preserve a low average path-length between clusters to **facilitate instant information transfer** (figure 7 in the appendix will clearly show how the outer low-degree nodes in the 50% graph become interconnected in the 20% graph, favouring high functional segregation and integration)

Foundational fiber tracts that are originally present in graphs with higher edge confidence are probably attached in a scale-free manner (high-degree nodes might represent regions that are heavily involved in data processing and low-degree nodes might represent sensory regions). More fiber tracts are then attached to form clusters around specific brain regions which we utilise more often in our individual lives. This could be pivotal in explaining how our brains might be similar and yet different from one another at the same time.

These hypotheses need to be tested and solidified more vigorously by understanding more about the nature of the subjects whose connectomes are used in this project's analysis (e.g. upbringing, social life)

7. Reference

[1]   S. N. Sotiropoulos and A. Zalesky, "Building connectomes using diffusion MRI : why , how and but," no. June 2016, pp. 1–23, 2019.

[2]   O. Sporns and D. S. Bassett, "Editorial : New Trends in Connectomics," pp. 2–4, 2018.

[3]   M. Rubinov and O. Sporns, "NeuroImage Complex network measures of brain connectivity : Uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

[4]   C. Kerepesi and V. Grolmusz, "Does the Budapest Reference Connectome Server Shed Light to the How to Direct the Edges of the Connectomes :," no. October 2016, pp. 0–9, 2015.

[5]   B. Szalkai, B. Varga, and V. Grolmusz, "The Robustness and the Doubly- Preferential Attachment Simulation of the Consensus Connectome Dynamics of the Human Brain," no. July, pp. 1–6, 2017.

[6]   PIT Bioinformatics Group, "Budapest Connectome Web Portal."

[7]   C. Kerepesi and V. Grolmusz, "Parameterizable consensus connectomes from the Human Connectome Project : the Budapest Reference Connectome Server," pp. 113–116, 2017.
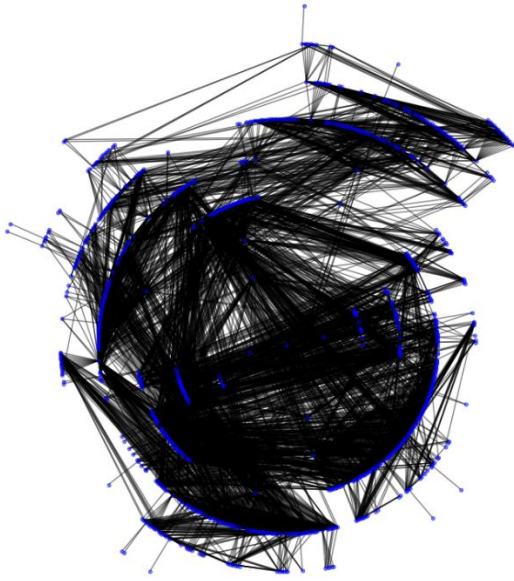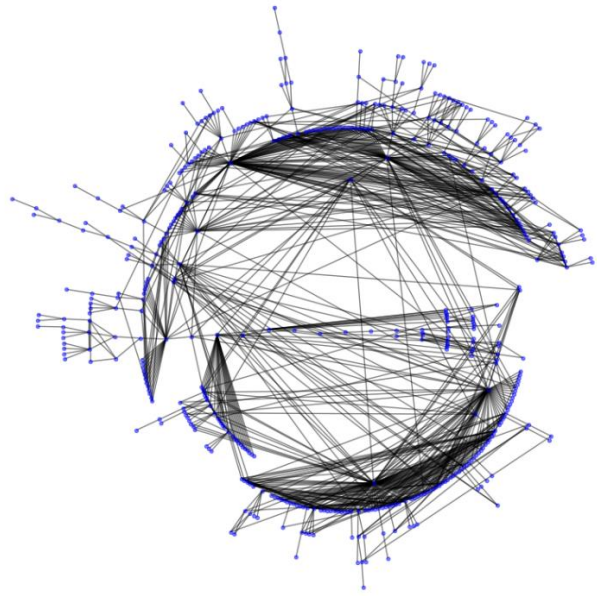
8. Appendix
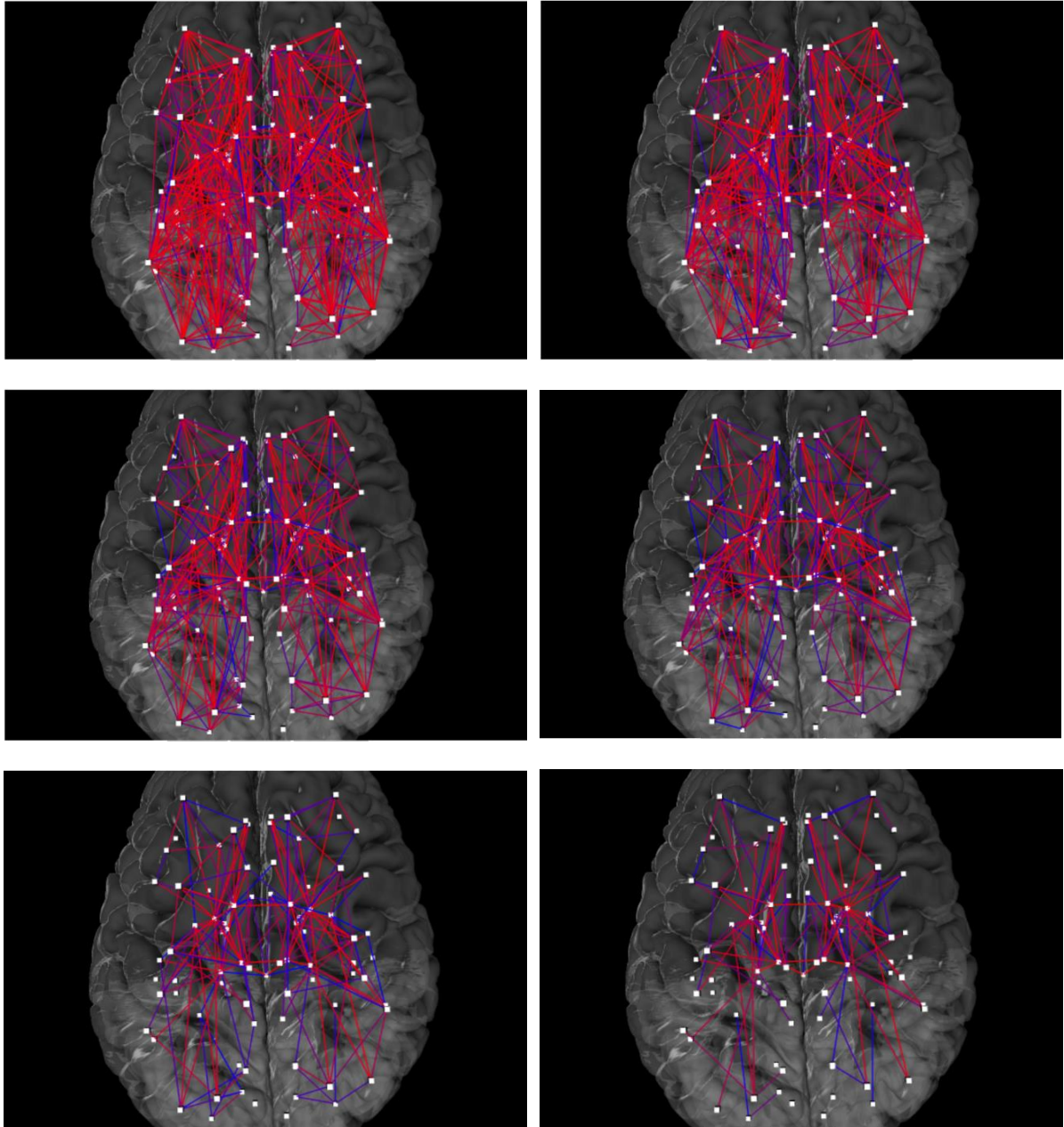
Figure 6: Network Visualisation on the Budapest Connectome Project (top-left image shows edges present in 20% of the participants; bottom-right image shows edges present in 70% of the participants)
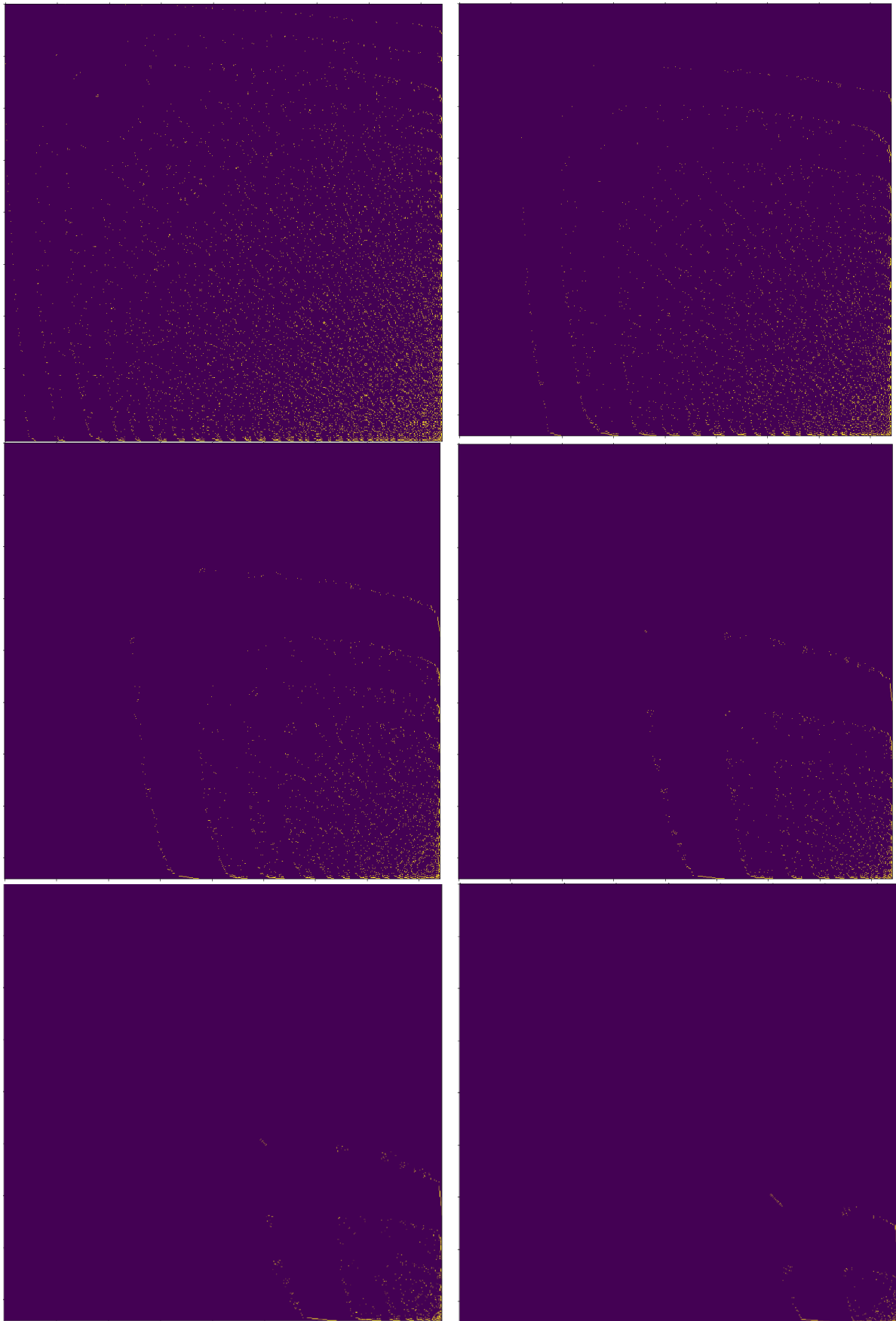
Confidence: 20 %
Confidence: 50 %

Figure 7: Circular plots for networks with edge confidence levels of 20% and 50%

Figure 8: BOSAM Visualisations for Consensus Connectomes (top-left image shows edges present in 20% of the participants; bottom-right image shows edges present in 70% of the participants)