

## **Distribution Agreement**

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Sriveena Chittamuri

Mar 28, 2019

# Using Hierarchical Random Graphs (HRGs) to Model Brain Networks

by

Sriveena Chittamuri

Dr. Shella Keilholz  
Adviser

Department of Physics

Dr. Shella Keilholz  
Adviser

Dr. Stefan Boettcher  
Committee Member

Dr. Ilya Nemenman  
Committee Member

2019

Using Hierarchical Random Graphs (HRGs) to Model Brain Networks

By

Veena Chittamuri

Dr.Shella Keilholz

Adviser

An abstract of  
a thesis submitted to the Faculty of Emory College of Arts and Sciences  
of Emory University in partial fulfillment  
of the requirements of the degree of  
Bachelor of Sciences with Honors

Department of Physics

2019

## Abstract

By Sriveena Chittamuri

The human cerebral cortex is functionally segregated with coactivating regions. These areas have been shown in literature to be organized across hierarchies from local to global networks<sup>11</sup>. Using fMRI to characterize whole brain activity, previous studies have shown the utility of functional connectivity graphs that represent the cross-correlation matrix of the fMRI activity at different brain regions as a measure to study network hierarchy<sup>13</sup>. However, functional connectivity estimates are known to be noisy- often requiring long timeseries to converge to a final value. We propose the utility of sampling the functional connectivity (FC) matrices from fMRIs and modelling the data as hierarchical random graphs (HRGs) that represent real data as dendrograms, since the hierarchical organization is theoretically independent of the noise. The HRG approach models the structure of the brain by clustering more connected brain regions together across increasing scales. We show that the random noise that exists in each individual scan as compared to the group average does not translate into this hierarchical representation and thus it can meaningfully predict missing edges from the ground truth more accurately than classical statistical methods can. By validating the use of HRGs through harnessing the hierarchical nature of brain networks to make predictions between FC matrices with known differences, we argue that this is a powerful approach to modelling the brain network since it can be translated to study unknown differences in varying populations including healthy versus disease brain studies.

Predicting Missing Edges on Network Graphs Using Monte Carlo Markov Chain Methods

By

Sriveena Chittamuri

Dr.Shella Keilholz

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences  
of Emory University in partial fulfillment  
of the requirements of the degree of  
Bachelor of Sciences with Honors

Department of Physics

2019

## Acknowledgements

I would like to thank Amrit Kashyap for his continued guidance throughout this project, for being the brains behind the algorithm, and for helping with the many drafts of this thesis. I would also like to thank Dr. Shella Keilholz for her help with finalizing on the approach and organization of the experiment.

Lastly, I am extremely appreciative of my academic advisor, Dr. Stefan Boettcher, for his continued guidance for the past 2 years as I navigated my eagerness and stubbornness to graduate early while taking on way more than I can handle. I wouldn't have been able to decide to do everything I did without his support.

## TABLE OF CONTENTS

<b>LIST OF FIGURES</b>	1
<b>1. INTRODUCTION</b>	3
1.1 Functional Architecture of the Brain	3
1.2 Functional Connectivity	4
1.3 Motivation for the Experiment	5
1.4 Networks and Hierarchies	6
<b>2. METHODS</b>	9
2.1 Experimental Paradigm	9
2.1.1 Preliminary Test	9
2.1.2 Optimizing Parameters	10
2.1.3 Analysis on Real Data	12
2.2 Dataset and Processing	14
2.3 Sampling Function	15
2.4 The Algorithm	17
<b>3. RESULTS</b>	22
3.1 Preliminary Test of the Algorithm	22
3.2 Positive and Negative Edges	24
3.3 Functional Connectivity graphs	25
3.4 Scanning Parameters	27

3.5 Comparing the Two Algorithms	34
<b>4. DISCUSSION</b>	36
4.1 Functional Connectivity Matrices	36
4.2 Positive and Negative Edges	37
4.3 Sampling Thresholds	38
4.4 Comparing the Methods	39
4.4.1 Missing Edges	39
4.4.2 Predicting Missing Edges	40
<b>5. CONCLUSION</b>	41
<b>REFERENCES</b>	43

## LIST OF FIGURES

Figure 1: Network model of the brain and its associated dendrogram <sup>16</sup>	8
Figure 2: Pipeline of the experiment on real data	14
Figure 3: Dendrograms <sup>1</sup>	19
Figure 4: Transitions of Dendrograms <sup>1</sup>	21
Figure 5 : Test Case 6x6 matrix	23
Figure 6 : Probability Graph of Test Case	23
Figure 7: Second Validation Test	24
Figure 8: The probability graph generated by the MCMC algorithm only on the negatively correlated edges	25
Figure 9: The probability graph generated by the MCMC algorithm only on the positively correlated edges	25
Figure 10: session 10 FC	26
Figure 11: All functional Connectivity Matrices	26
Figure 12: The average of the 10 FC. The FC values that the probability data is plotted against	27
Figure 13: The distribution of all the edges plotted with respect to their FC values	27
Figure 14: Figure 13: The scanning threshold parameters plotted against the distribution of all FC values.	28
Figure 15: the sweep of the different sampling coefficients.	29

Figure 16: The sweep of the different sampling coefficients.	30
Figure 17: The probability graphs of the gold standard at different sample size	32
Figure 18: The probability graphs of the predicted graph at different sample sizes	33
Figure 19: Predicted missing edges - Convention Method	34
Figure 20: Predicted missing edges - MCMC Method	34
Figure 21: Comparing the two methods on one sample against the mean FC	35
Figure 22: Comparing the two methods on one sample against the individual FC	36

## 1. INTRODUCTION

### *1.1 Functional Architecture of the Brain*

The brain is known to have specialized cortical regions that perform functions through segregated (localized) and integrated (distributed) processes<sup>7</sup>. The brain, therefore, can be viewed as adhering to two principles of functional organization: functional segregation and functional adherence. Functional segregation refers to spatial localization of neurons that perform similar roles, while adherence refers to the processes that involve cross-communication between these regions<sup>2</sup>. Neuroscience research is primarily focused on studying these two organizational structures in order to discover the complex processes that govern cognition.

The human cerebral cortex, in particular, has been explored in the recent years to reveal that its functionality can be described by distinct, yet intertwined, networks<sup>9</sup>. A set of structurally segregated neurons have similar functions, and a few sets of these neurons coactivate, and so on to create the several large-scale brain regions, that scale upwards.

Functional neuroimaging has been imperative in the past decade in establishing these as the principle organizations of the brain<sup>2</sup>. Functional Magnetic Resonance

Imaging (fMRIs) has been the primary and critical tool in isolating and analyzing different segments of the brain that change their activation based on a given stimulus or action. Unlike EEGs or MEGs, fMRI is an especially useful tool because it provides detailed whole brain scope of activity, with regional specificity. fMRI is sensitive to changes in the blood oxygen level dependent (BOLD) signals, which are indirectly dependent on neural activity<sup>4</sup>. fMRI's allow for ROI(region of interest) analysis, where changes in the BOLD signals at individual regions are studied. The average time signal (of changes in the BOLD signal) at these ROI's are observed and analyzed in both task and rest studies. Task based studies have given insight on the role of different regions of the brain, and allowed for thorough documentation of the functional segregation of the brain. However, until the past decade, little was known about the interactions between different regions of the brain, paving the way for resting state studies.

Resting state studies are performed with no explicit presence of stimuli or activity, and reveal details about the interactions between the different ROI, giving deep insight into the functional architecture of the brain. From resting state fMRIs, the average dependence (often a Pearson correlation) between each pair of ROI's is found from the time series, and a functional connectivity (FC) is created. These FC matrices are the primary tool in resting state whole brain analysis for various different applications in neuroscience<sup>10</sup>.

## ***1.2 Functional Connectivity***

As mentioned, functional connectivity (FC) graphs describe the correlation between ROI's, and the FC values refer to the quantification of the operational interactions of multiple spatially-separated regions of the brain<sup>4</sup>. Functional connectivity is agnostic, and so the value of the correlation describes a directionalless connection between two edges, where a higher values indicates a stronger correlation between the two segments of the brain. In terms of the fMRI collection, a higher FC value refers to a higher correspondence of the BOLD signal at these two regions at a given period of time. A positive correlation value refers to additive interactions between regions, where regions are activated, while a negative correlation refers to suppressive interactions between ROIs.

Similarly, this approach of studying fMRI's is also prevalent in studies involving neural diseases. Specifically, studies involving Parkinson's and alzheimer's have revealed functional connectivity differences in the brain of patients with these disorders<sup>14,15</sup>. Therefore there is a deep necessity for developing a tool to study these differences. However, the approach of using FC matrices to study human disease as well as cognition has its own shortcomings including the effects of noise surrounding the fMRI as well as the need for large amounts of data.

### ***1.3 Motivation for the Experiment***

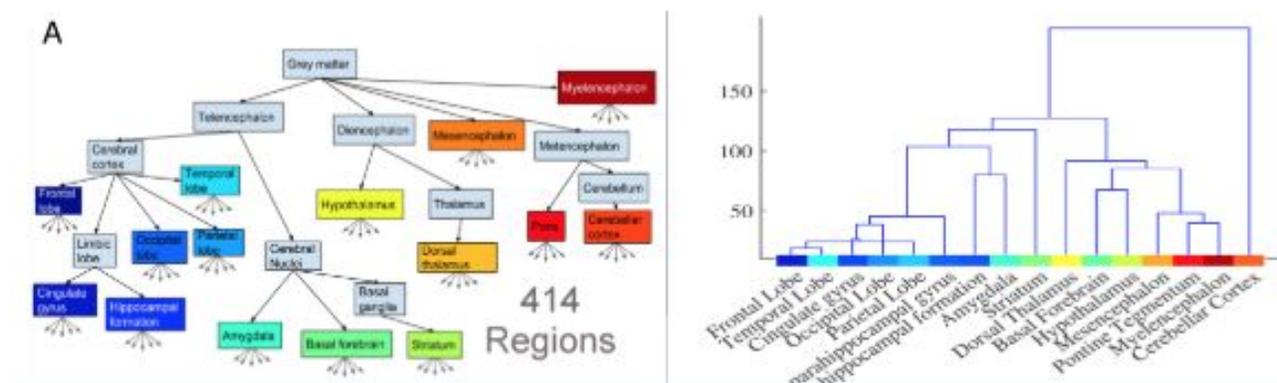
Resting State fMRI measures whole brain data in low frequency ( $<0.1\text{Hz}$ ) BOLD signals but is a noisy measurement, accumulating scanner and nuisance noise. The noise

is attributed to illegitimate signals interfering with the BOLD signal from ventricles and white matter, as well as head motion. Raw data is often corrected for this noise, but the noise impacts the functional connectivity matrix and can produce spurious connections between ROI<sup>11</sup>. In order to settle this issue, FC matrices are thresholded to ensure accuracy across the dataset<sup>11</sup>. However, the thresholding can lead to individual differences between the FCs including missing edges. So, proportional thresholding is often done to equalize the density of the edges in the dataset<sup>11</sup>. Another approach to resolve the issue is to perform longer scans or over a large enough population to gain enough data to make up for these differences. This can be both expensive and even impossible in situations that require data from people with rare disorders. Therefore, there is a strong necessity to identify the differences in FC in different populations even with noisy data and to do so with minimal amount of data present. Here, we present an approach to studying FC matrices that is unaffected by the noise in the RS-fMRI and requires far shorter segments of fMRI scans.

#### ***1.4 Networks and Hierarchies***

Different segments of the brain coactivate together to perform specific functions. This functional organization allows the brain to be modelled as a hierarchical random graph, where brain regions that are more functionally active are clustered in different levels<sup>9</sup>. At the lowest levels, these are local networks- which are highly connected voxels or sets of neurons with the same function. These connected neurons are connected at a

higher level based on the region of the brain that they constitute (eg: amygdala, basal forebrain, and basal ganglia). These parts of the brain are then connected into regions with similar function (eg.cerebral nuclei), which are connected on a higher level and so on, with the top of the hierarchy constituting the entire grey matter of the brain. Figure 1 depicts the schematic of the brain(right of the figure) and it translated into a hierarchical structure (left of the figure)<sup>16</sup>. Evidently, the architecture, or the functional connectivity of the brain, can be modelled through a dendrogram, where the highest level of organization describes global networks, and scales down to local, highly correlated networks. Each of these local networks, therefore, are tied together in the hierarchy, which accounts for the overlapping. Our argument here is that that functional connectivity graphs can be described using these dendrograms. There is evidence of this in literature<sup>9</sup> but also in the clear clustering of the functional connectivity matrices (see figure 9), where correlations are grouped in boxes-suggesting segregated functionality. These dendrograms are used to build Hierarchical Random Graphs (HRGs) that describe this hierarchical organization of the brain, and therefore act as a strong representation of the FC data.



**Figure 1:** The left figure reveals a schematic of the different organizational levels of the brain. The figure on the right reveals the same organization in the form of a dendrogram<sup>16</sup>

The integral property of HRGs is that they have naturally existing cliques, or clusters, that are inherent to this structure. The cerebral cortex, as mentioned, has similar clustering and so HRGs act as a powerful and accurate tool in modelling the data. The mathematical formalization of creating the HRGs and choosing the best possible one is detailed in the methods section and is in accordance to the paper by Aaron Clauset, Christopher Moore and M. E. J. Newman<sup>1</sup>. The method of producing a series of dendrograms from the network graph is called Monte Carlo Markov Chain (MCMC). The MCMC method utilizes a chain of transitions that swap out edge connections to create new dendrograms. The processes of choosing the best fitting hierarchical graph from the dendrograms is through minimizing the Shannon entropy. Shannon entropy refers to the number of bits it takes to create each dendrogram, the fewer bits it takes to create a dendrogram, the more likely the hierarchical random graph is to truly model the network because it requires the fewest stems to accurately generate the network. The paper by Clauset et al, recounts the pipeline to create the best fitting random hierarchical graph from a network graph in order to give insight into various network phenomena including

clustering coefficients, right-skewed distributions, and predicting missing edges. Here, we present a similar approach of creating an HRG but by generalizing it to take in several different graphs.

In order to reveal the strength of using HRGs to describe the functional connectivity matrices and give powerful insight into the differences in FC, we test our algorithm by predicting known missing edges and comparing it to a conventional method of finding difference in FCs. Provided the algorithm's validity, it can be generalized to study differences in a variety of experiments, including disease and task based studies.

## 2. **METHODS**

### *2.1 Experimental Paradigm*

In order to validate the MCMC algorithm and motivate using HRGs to model brain networks, a missing edges experiment is performed. The accuracy of the missing edges experiment is intended to provide support for generalizing the use of this method to model FC graphs to reveal differences between the fMRI data in different populations.

#### *2.1.1 Preliminary Test*

Firstly, the algorithm is run on a simple 6x6 matrix. Noise is added to the matrix (random edges are changed to equal 1 and others are removed) and then

sampled 10 times. A 3D matrix of all the sampled graphs is then fed into the MCMC algorithm with 1000 iterations (transitions in the MCMC chain) where the Markov chain is initiated. The best fitting hierarchical graph is created, from which the probability graph is generated.

Another validation test is done on an artificial 66x66 matrix that has clear clustering across the matrix. Then 20% noise is added to the data, where edges are randomly swapped and added. Then, 10 samples for this is fed to the MCMC algorithm to generate a probability matrix.

### ***2.1.2 Optimizing Parameters***

Firstly, FC matrices are oversampled at a certain threshold in order to construct a set of graphs that we feed into generating HRGs for the given network data. Then using the MCMC algorithm we estimate the HRG that best represents the network data. The FC matrix is first separated in to two networks for the positive FC and the negative FC edges. These are separated because the positive and negative correlation values represent additive and suppressive activity respectively. It, therefore, benefits to run them as separate graphs since the algorithm is better equipped to handle them as such in its predictions. In order to ensure that only positive and then only negative edges are extracted, the matrix is multiplied by +1 or -1 (for positive and negative edges respectively). The sampling function (described in detail in section 2.3) extracts all the edges above a certain positive threshold. When the negative edges experiment is run, the entire

matrix is multiplied by -1 so that the negative edges will be extracted rather than the positive ones. The multiplied FC is sampled 10 times (10 different graphs are generated) and processed by the algorithm. This sub experiment is done for an individual graph (session 10).

In order to find the optimal thresholding value and the sample size, first the sample size is held constant while the sample parameters are changed. The three different parameters swept include: 20%, 25%, 30%, i.e. the FC values are multiplied by 1.2, 1.3, and 1.4 in order to increase their probability of being selected by the sampling function (explained in detail in section 2.3). This parameter is set and a leave one out cross validation approach is done, where 9 of the samples ( session 1-9) are run together to represent the gold standard, and the remaining graph (session 10) acts as the predicted graph. The difference between the predicted sample and the nine is found for each different parameter, and then the three are correlated to the global gold-standard FC ( the mean of all 10 samples) to decide the optimal parameter. In this part of the experiment, only the positive edges are extracted.

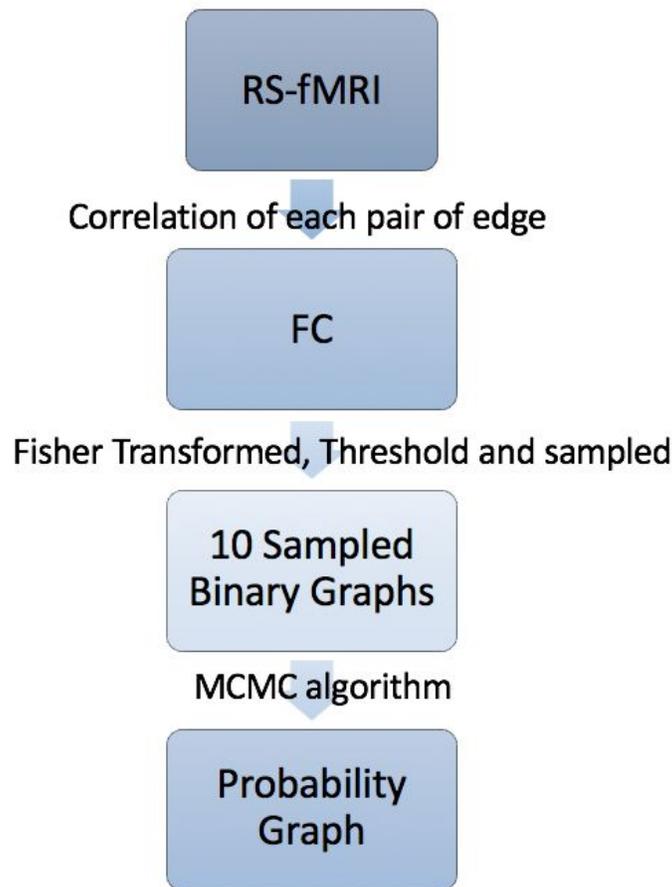
Then, the optimal scanning parameter is kept constant while the sample size is altered. The three sample sizes explored are 10, 15, and 20. In each instance, the probability graph for the first nine sessions is produced and the last sample acts as the prediction set and has its own probability graph. The prediction set is subtracted from the nine in order to observe the number of difference between the two to choose the optimal parameter.

### ***2.1.3 Analysis on Real Data***

The missing edges experiment is done on resting state data of different sessions from the same subject. 9 of the 10 samples are combined to represent the gold standard. Since the dataset is large enough, we work under the assumption that all of the existing edges are present in these 9 graphs and so it can be cross validated to test the accuracy of the algorithm.

Resting state time series data from one subject from the Midnight Scan Club (MSC) Project is preprocessed and projected into atlas space. The RS-fMRI is not functionally segregated and so this is an important step in extracting the functional connectivity. It also ensures that the role of each of the different segments (ROIs) is known, which is advantageous for further analysis of the data. Then, a functional connectivity matrix is created for each of the 10 sessions of the data collection. The functional connectivity matrix is merely a correlation graph of the correlation between each ROI(edge) of the brain. This acts as the network graph given to the MCMC algorithm. Providing it with multiple graphs from the same subject or same experiment results in a probability matrix with high accuracy. In order to predict the missing edges, a leave one out cross validation approach is used again. Nine of the functional connectivity graphs are then sampled and fed (according to the optimized parameters) into the algorithm to find the probability of each edge existing. This is then compared to the one predicting set that receives its own probability graph. The MCMC method is

compared to a brute force (referred to as the conventional) method of finding the p values from a one sample student t-test statistic that observes the difference between an individual sample and the average of the rest of the graphs. The t-test gives a p value for each edge of the prediction graph existing in the training set. A high p value suggests a higher probability of it existing in the set, and this constitutes a missing edge. The two are then cross validated. If the missing edges predicted by the MCMC algorithm are edges with higher functional connectivity than those predicted by the conventional method, it suggests that the algorithm succeeds in accurately predicting missing edges since these are the ones with high connectivity with other edges in the graph. This process is then repeated for all 10 permutations of the leave one out method for both the MCMC and the conventional method and cross validated to account for any variation in individual data and any possible outliers. Figure 2 shows the general pipeline of the experimental design for the real dataset.



*Figure 2: Pipeline of the experiment on real data for an individual sample. In the case of the nine graphs, a 3D matrix with each sample sampled 10 times into binary graphs is sent to the algorithm*

## ***2.2 Dataset And Processing***

The data is obtained from the Openneuro database. The resting state data is part of the Midnight Scan Club (MSC) Project which collected data from ten different subjects for five hours. Each subject is scanned for 30 minute scans over ten different sessions. Each of these sessions acts as an individual sample in this experiment. The MSC provides

derivatives of the raw dataset of the RS-fMRI data which includes the minimally preprocessed, motion-sensored, and confound regressed resting state data in CIFTI(cortical: fs\_LR32K;subcortical: Talarich) space. Resting State fMRI explains the functional architecture of the brain by measuring spontaneous, low frequency ( $< 0.1\text{Hz}$ ) BOLD signals<sup>3</sup>. Bold signal preprocessing requires correction for slice-dependent time shifts as well as intensity differences. It also needs to be corrected for noise due to head movements as well as nuisance signals from white matter and ventricles. Following that, spatial smoothing and low pass filtering for low frequencies is done to cancel out the noise due to non-neuronal signals. These are the pre-processing steps done on the dataset when they are extracted from the project. At this stage the resting state time series are then projected onto an atlas space so that they are grouped in accordance to their functional segregation<sup>3</sup>. The timeseries of each session is averaged to a Desikan-Killiany atlas<sup>12</sup>. Then, we further pre-processed the data using the following processing pipeline: z-scoring each time series, then band passing filtering the signal from 0.01 to 0.25 Hz, then global signal regression using a linear regression model, and finally, applying a final z-score step. These steps were selected in accordance with Cabral et al.<sup>1</sup> The correlation is extracted for each sample. This is the functional connectivity matrix of each session and acts as an individual sample in our experiment.

### ***2.3 Sampling Function***

Thresholding the connectivity matrix is an important step in theoretical graph analysis, and there is increasing evidence that proportional thresholding is a better tool for doing so than absolute thresholding<sup>11</sup>. Proportional thresholding chooses edges to represent the graph based on their probability of existing- so edges with a higher probability are more likely to be chosen to be in the sample than edges with lower probabilities<sup>11</sup>. When working with functional connectivity matrices, the thresholding function is meant to grab edges with higher FC more often than edges with lower FC since these on average have a higher probability of being spurious and thus introduce a higher degree of randomness to the generated network<sup>11</sup>.

Each FC in the experiment is sampled before being fed into the MCMC algorithm. The sampling function chooses an edge to exist in the sampled graph based on its probability. So an edge with a higher probability has a higher chance of being sampled. The function randomly generates a floating point number between 0 and 1, and if the FC has a probability greater than that value, the edge is sampled and set to equal 1. In this way, the function generates a binary matrix with edges with high FC existing in the sampled graph.

In terms of choosing an optimal scanning parameter, the goal is to sweep the area of different values and choose one that samples the most number of edges with high probability while still producing accurate probability graphs. So, multiplying the FC matrix with 1.2, 1.25, and 1.3, as described in section 2.1.2, increases the probability of an edge being selected by 20%, 25%, or 30%.

Since the sampling method only samples when a value is greater than a floating point number between 0 and 1, it is only able to sample positive edges. For the positive and negative edges graph sub-experiment, the FC matrix is first multiplied by -1, so that the edges that are positive are now negative and vice versa. Additionally, for this sub-experiment, no sampling threshold parameter is added to the matrix, so the probability of an edge being sampled is equal to its functional connectivity. The real dataset is also processed by separating the positive and the negative FC edges in order to increase the predictive accuracy of the algorithm.

#### ***2.4 The Algorithm***

In order to predict the existence of connections between two edges, the hierarchical property of the network of the fMRI data is exploited using the Monte Carlo Markov Chain (MCMC) methods to sample the data. The algorithm then produces the probability of the existence of an edge in the network.

Firstly, from the sampled FC graphs, the algorithm produces an initial dendrogram. The dendrogram (see figure 1) is essentially a tree, where each node represents a hierarchical network. The tree goes from global connectedness to local. Each node represents clusters of edges that have high connectivity. So, at the lowest level, are several highly connected edges.

Each edge of a graph  $G$  (the functional connectivity matrix) is represented by the variable  $n$ . A dendrogram  $D$  is constructed with  $n$  leaves. Each  $n-1$  internal nodes of  $D$  is associated with the group of nodes it is descended from. Each node  $r$  has a probability  $p_r$  associated with it.  $p_r$  represents the probability of there existing a connection between 2 specific nodes. Then,  $\{D, p_r\}$  defines the hierarchical random graph. The probability of each node  $p_r$  is averaged over all given graph  $G$ s, to produce a dendrogram  $D$ .

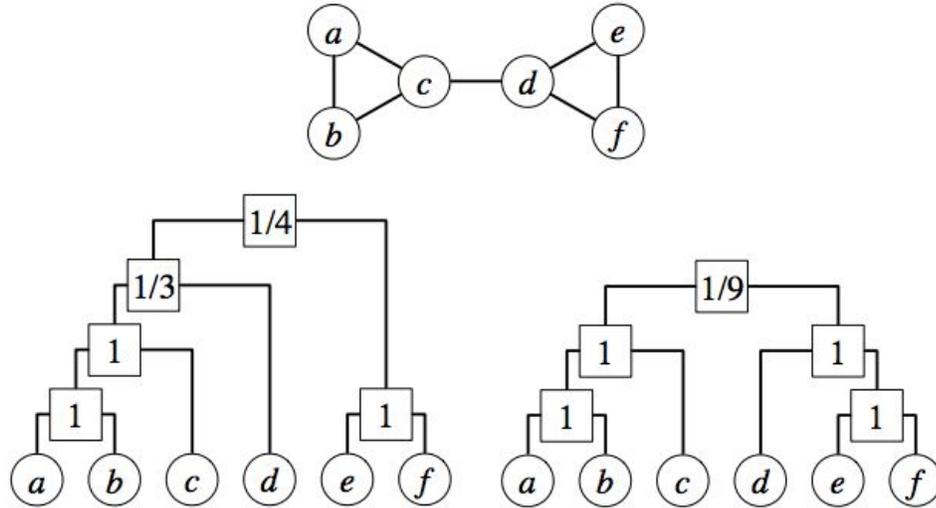
However, there are several different possible dendrograms that can be created given a set of graphs  $\{G\}$ , so in order to find the best hierarchical random graph, the *likelihood*  $L$  with which the algorithm generates the dendrogram is studied. This is under the assumption that every hierarchical random graph is equally as likely to be produced. This  $L$  is then maximized - the space is sampled for all models with probability proportional to  $L$ , in order to find the hierarchical graph that best fits the given set of functional connectivity matrices.

$E_r$  is the number of edges in  $G$  whose endpoints have  $r$  as their lowest common ancestor in  $D$ , and  $L_r$  and  $R_r$  represent the number of leaves to the left and right of the subroots at  $r$ . The likelihood of producing a dendrogram given a set of graphs  $G$  is

$$L(D) = \prod_{D \in G} \prod_{r \in D} p_r^{E_r} (1 - p_r)^{L_r R_r - E_r} \quad (1)$$

Different dendrograms have different likelihood values. The goal is to maximize the likelihood in order to discover the best fitting hierarchical graph (see figure 3). We do

so by sampling several graphs and finding the average probability of each edge and using that in the likelihood function.



**Figure 3:** The left random hierarchical graph has a likelihood of 0.00165 and the right has a likelihood of 0.0433 according to equation (1). The left one is clearly the best fitting hierarchical graph

When a dendrogram is fixed with the probabilities  $\{\bar{p}_r\}$  that maximize equation (1), the

$\bar{p}_r$  is then

$$\bar{p}_r = \langle \frac{E_r}{L_r R_r} \rangle \quad (2)$$

which is the fraction of potential edges between the two subtrees of  $r$  that actually exists.

The  $\bar{p}_r$  is found for an edge in one graph, and then averaged over all graphs  $G$  given in order to find the average set of maximized probabilities  $\{\langle \bar{p}_r \rangle\}$  to work with. The

likelihood at this maximum is then

$$L(D) = \prod_{D \in G} \prod_{r \in D} [\bar{p}_r^{\bar{p}_r} (1 - \bar{p}_r)^{(1 - \bar{p}_r)}]^{L_r R_r} \quad (3)$$

,where  $\overline{p_r}$  refers to the average of the probabilities over all the graphs. The logarithm of this function is easier to work with

$$\log L(D) = - \sum_{D \in \mathcal{G}} \sum_{r \in D} L_r R_r h(\overline{p_r}) \quad (4)$$

where

$$h(\overline{p_r}) = -p \log p - (1-p) \log(1-p) \quad (5)$$

is the Gibbs Shannon Entropy Function. Equation (4) is then maximized when  $p_r$  is close to 0, or the entropy is minimized. The shannon entropy refers to the the number of bits required to encode something. So, the best fitting hierarchical graph is one in which the entropy is minimized and where the connections between two edges is most common.

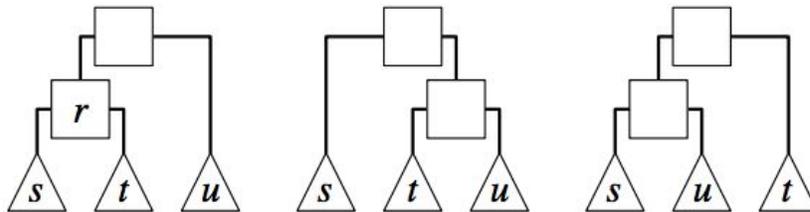
The Markov chain Monte Carlo (MCMC) method is used to sample the dendrograms  $D$  of the graphs  $G$  with the probability proportional to their likelihood  $L(D)$ . This method is used to sample the surrounding space to find the best fitting dendrogram. The majority of the code follows this. The Markov chain is created by first allowing a new dendrogram  $D'$  to be created. This is done so by selecting a node  $n$  uniformly at random, which has three subtrees  $s$ ,  $t$ , and  $u$  (see figure 4) with  $s$  and  $t$  descending from its daughter and  $u$  from its sibling, and then selecting one of the two configurations to switch uniformly at random. These transitions are *ergodic*: any set of finite dendrograms can be connected through a finite number of these transitions. This then generates  $D'$ , then in order to verify which of the dendrograms to carry on the chain with, the Metropolis-Hastings rule is used.  $D'$  is chosen if

$$\Delta \log L = \log L(D') - \log L(D) \quad (6)$$

is nonnegative ( $D'$  is at least as likely to occur as  $D$ ). Otherwise  $D'$  is accepted with the probability

$$\exp(\log \Delta L) = L(D')/L(D) \quad (7)$$

If  $D'$  is not accepted then  $D$  is continued in the chain. The Markov Chain converges relatively quickly with the likelihood reaching a plateau at around  $O(n^2)$ .



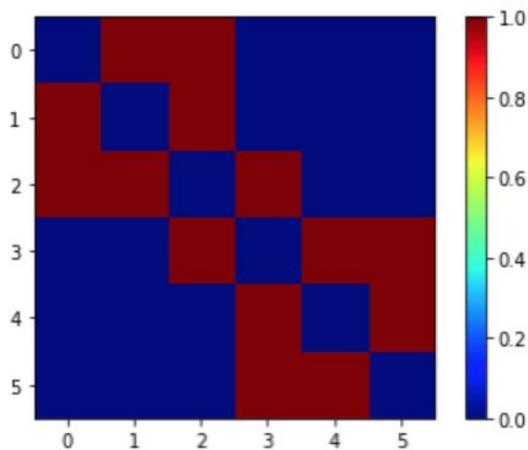
**Figure 4:** The first represents a dendrogram  $D$ , while the following two are two different  $D'$  generated with different transitions in accordance to the Monte Carlo Chain

In order to predict the missing connections, the Markov chain is initiated with a random dendrogram and ran until it equilibrates, and generates the best random hierarchical graph. For each pair of edges  $i, j$  that do not have a known connection, we calculate the mean probability  $\langle p_{ij} \rangle$  that they are connected by averaging over the corresponding probability  $p_{ij}^l$  in each of the sampled dendrograms<sup>1</sup>. Then, a probability graph is generated where each edge has an associated probability of it existing in the network.

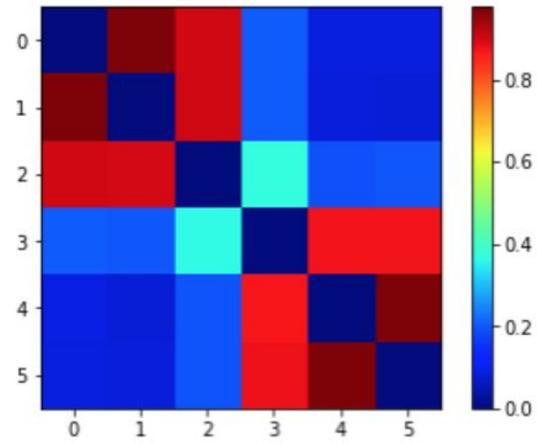
### 3. RESULTS

### 3.1 Preliminary Test of the Algorithm

The original 6x6 matrix generated to test the MCMC algorithm can be seen in figure 5. Figure 6 represents the probability graph generated by the MCMC algorithm. It is clear that the edges with the highest probability (red) are edges present in the original graph. This validates the performance of the algorithm on a small dataset.

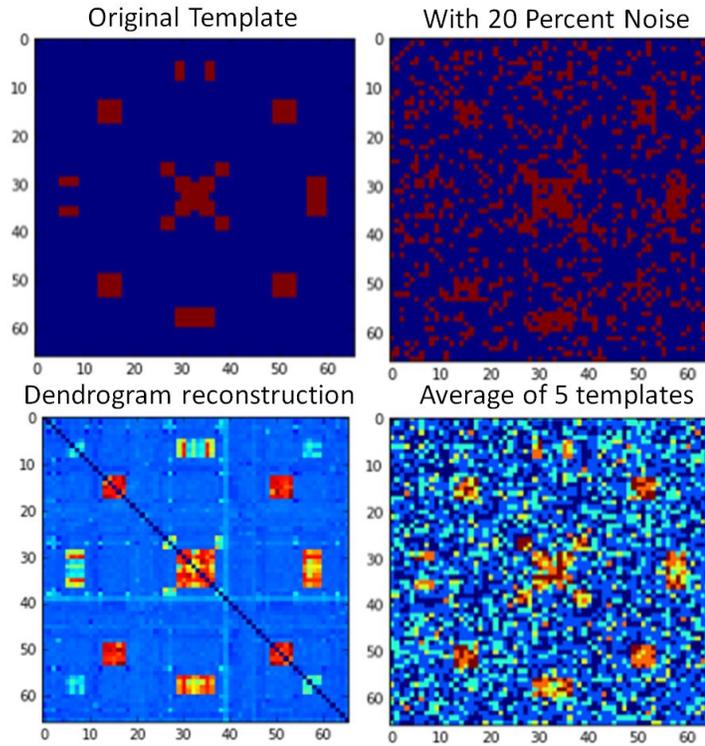


**Figure 5: The test case 6x6 matrix**



**Figure 6: The probability graph generated from the MCMC**

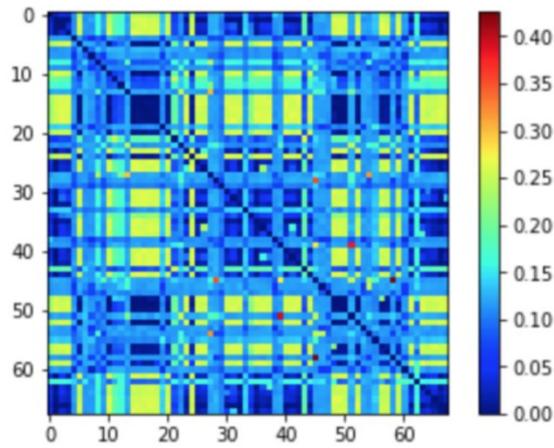
The second test is performed on the 66x66 matrix and can be seen in figure 7. The figure reveals that the probability graph in the bottom right retains the same clusters as in the original FC matrix provided to the algorithm. This is done for five different runs and overlaid in the figure on the bottom right of figure 7 to reveal the accuracy of the algorithm



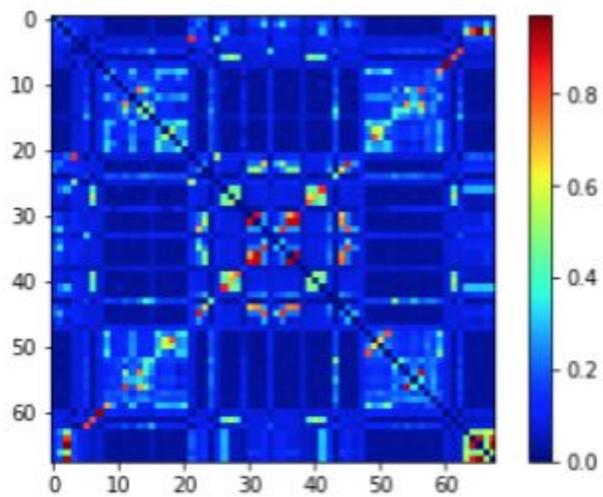
**Figure 7: Second validation test. Top Left: Original artificial FC, Top Right: FC with noise added, Bottom Left: The probability graph generated, Bottom Right: Average of all 5 templates.**

### **3.2 Positive and Negative Edges**

In order to show the differences between the positively and negatively correlated edges in the FC graphs, a separate sampling is done on both. Figure 8 and 9 show the two different probability graphs produced when sampled separately.



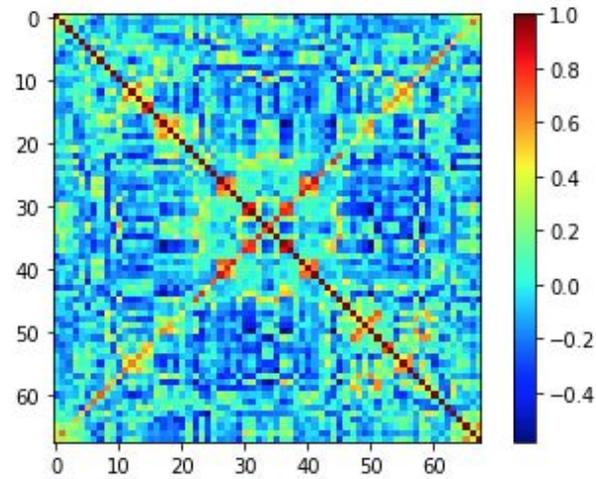
*Figure 8: The probability graph generated by the MCMC algorithm only on the negatively correlated edges of one sample*



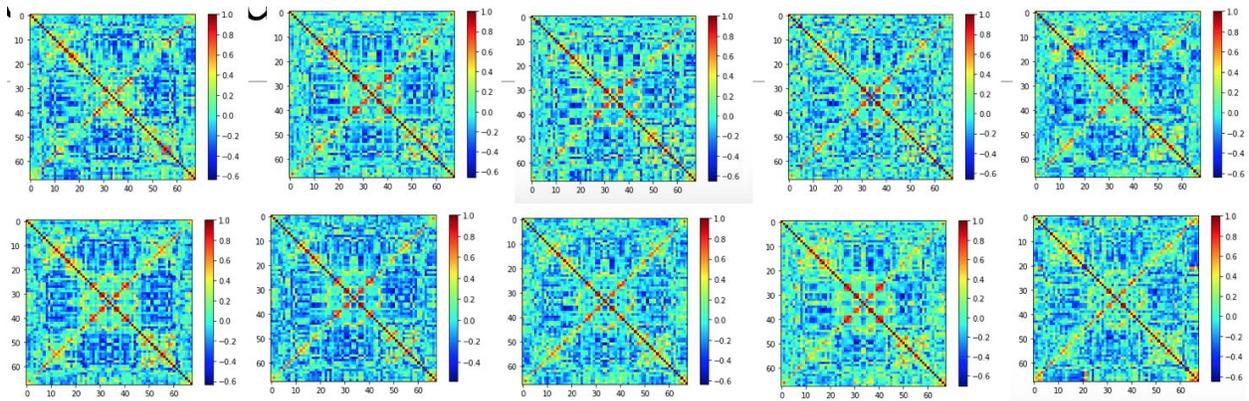
*Figure 9: The probability graph generated by the MCMC algorithm only on the positively correlated edges of one sample*

### 3.3 Functional Connectivity Graphs

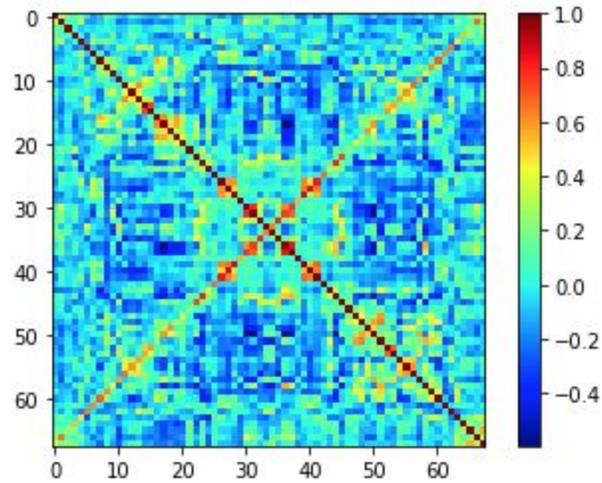
The functional connectivity matrix for one of the samples is visible in figure 10. The FC of the others looks fairly similar but varies in some edge connections (figure 11). The diagonals are all ones because of self connections. Figure 12 then is the average of all 10 of the graphs. This is the FC graph that the final results are plotted against.



*Figure 10: Session 10 FC*



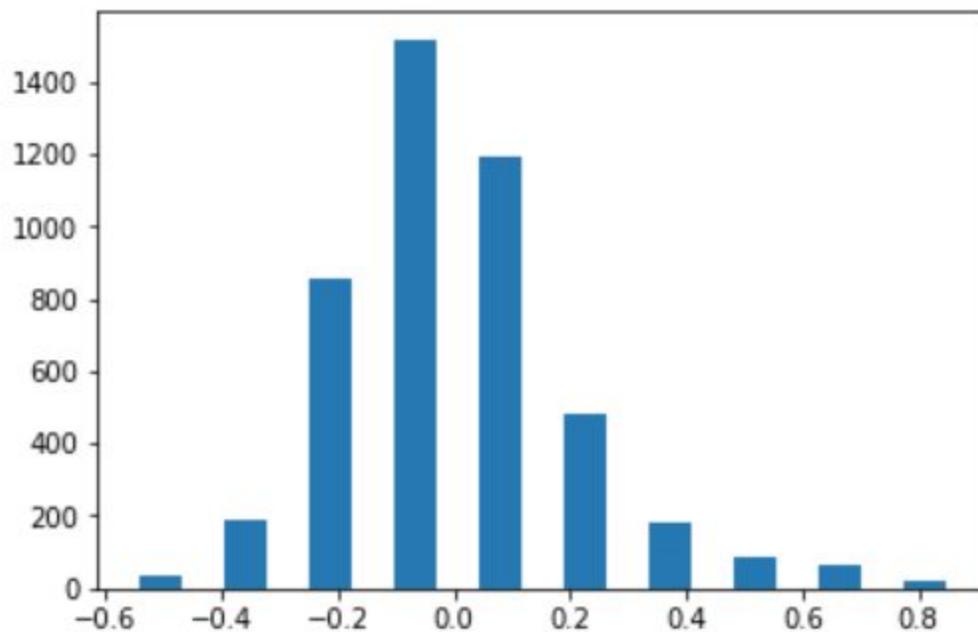
*Figure 11: All functional connectivity matrices*



*Figure 12: The average of the 10 FC. The FC values that the probability data is plotted against.*

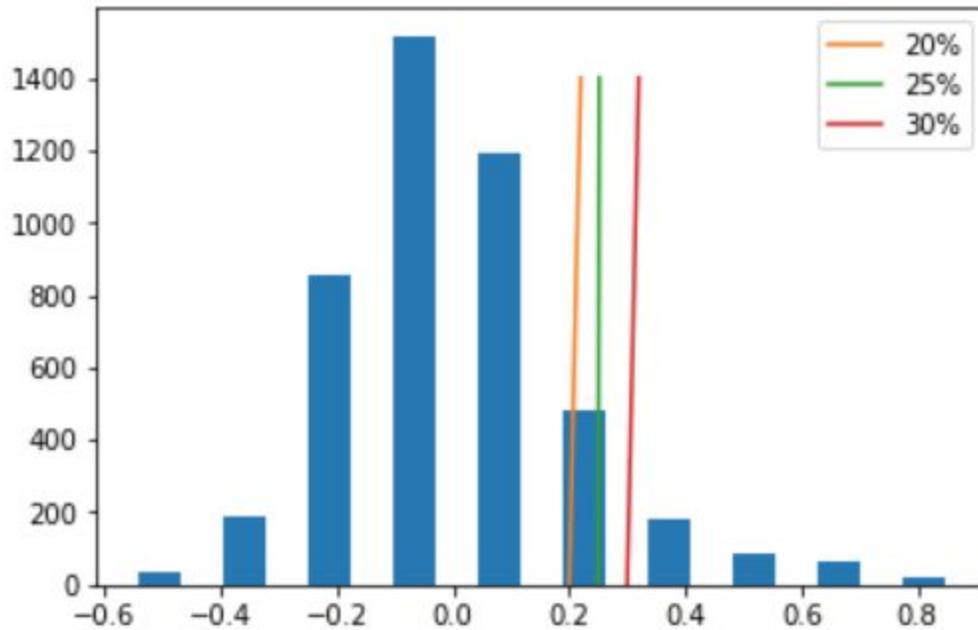
### 3.4 Scanning Parameters

The distribution of all the edges in the mean FC graph can be seen in figure 13.



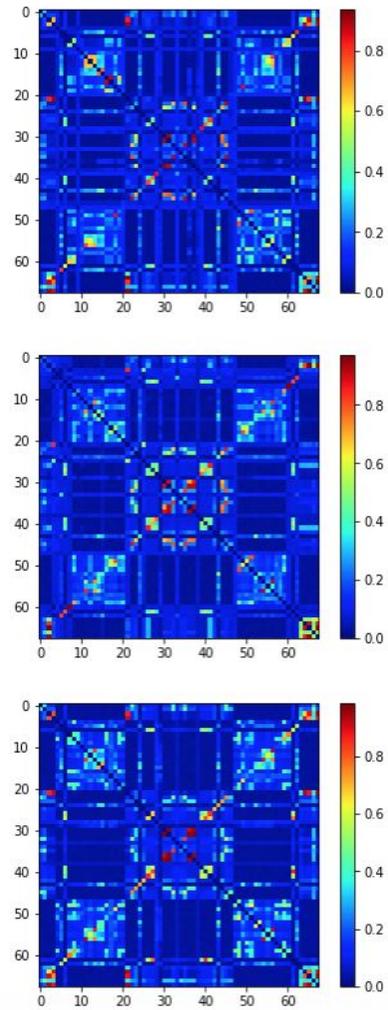
*Figure 13: The distribution of all the edges plotted with respect to their FC values. The self-connections are removed from this plot.*

In order to ensure that the sampling function collects the edges with the highest FC values, three different scanning thresholds are tested. Figure 14 shows the three different parameters: 20%, 25%, 30%. The FC values are multiplied by 1.2, 1.25 and 1.3 in order to increase their probability of being selected by the sampling function. In this part of the experiment, only the positive edges are extracted.

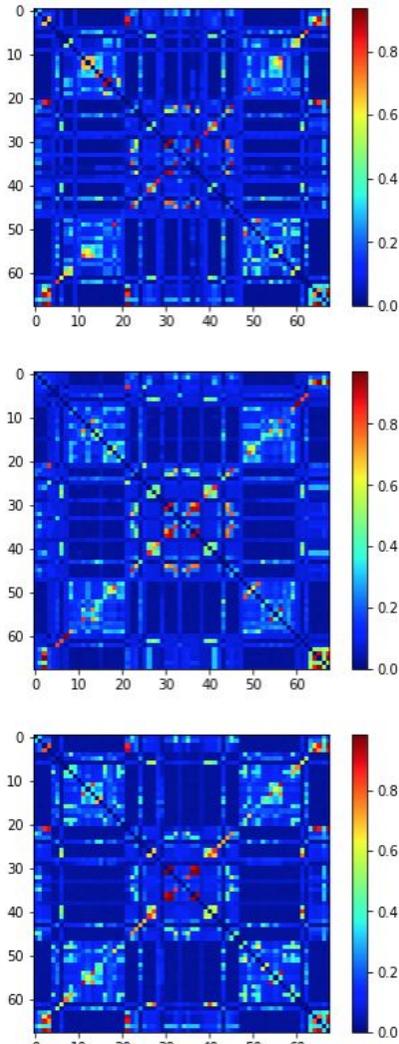


**Figure 14: The scanning threshold parameters plotted against the distribution of all FC values. The equation used to set up the probability of being sampled:  $FC \cdot k$ , where  $k = 1.2, 1.25, \text{ and } 1.3$**

Figure 15 and 16 show the three different probability graphs generated for each scanning parameter. Based on the figures alone, it appears that the 25% threshold has the most structural clustering.



*Figure 15: The sweep of the different sampling coefficients. Each probability graph is generated from running 10 samples of 9 different RSFC sessions. Top: 20% threshold, middle: 25%, Bottom: 30%*



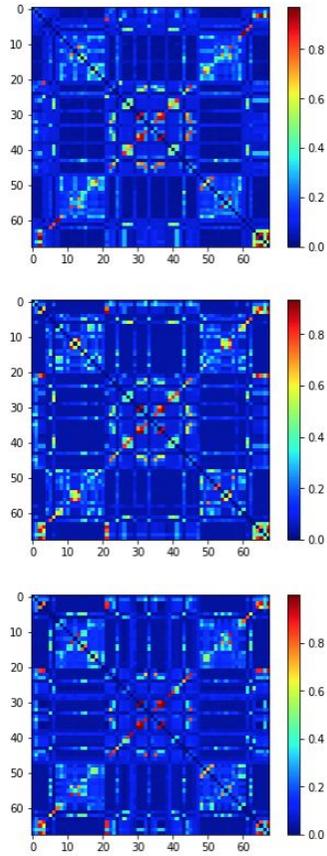
*Figure 16: The sweep of the different sampling coefficients. Each probability graph is generated from the FC of session 10.*

*Top: 20% threshold, middle: 25%,  
Bottom: 30%*

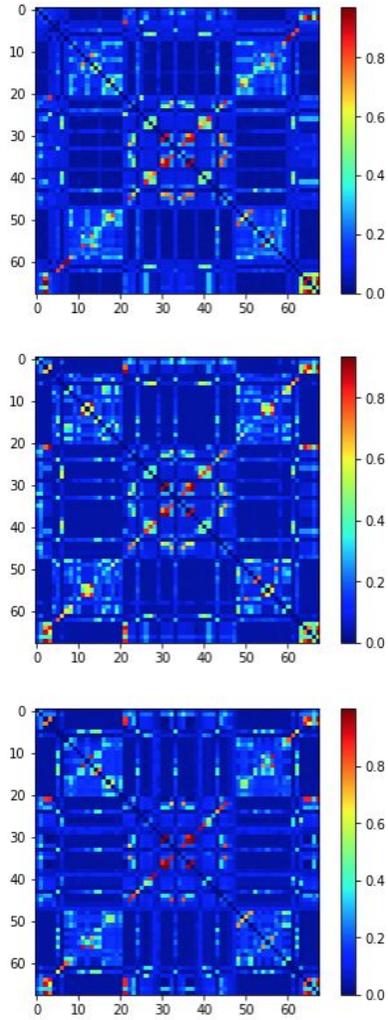
The predicted probability graph of session 10 (figure 16) is subtracted from the gold standard (figure 15) to count the number of differences. The higher the number, the more different edges that the sampling function is able to grab. The 20% sampling revealed 4550 edges with non-zero probabilities, while the 25% produced 4555 and the

30% produced 4554. When the probability graph of the nine are correlated with the original FC values (the mean FC of the 10 samples (figure 12)), the correlation values for the 20%, 25%, and 30% were 66.423, 72.059, and 65.976. The 25% sampling threshold is set as the optimal parameter.

Once the sampling threshold is set at 25, the number of samples for each FC is swept. The first is set at 10, then 15, then 20. Figure 17 and 18 reveal the gold standard and the prediction graphs at each parameter. Again, the differences between the two are counted. The lower the number, the better the algorithm is at predicting missing edges since there is less variance between the average probability graph and the individual one. There are 4555 different edges with the 10 samples and 20 samples, but 4448 with the 15 samples and so the sample size in the algorithm is set to 15.



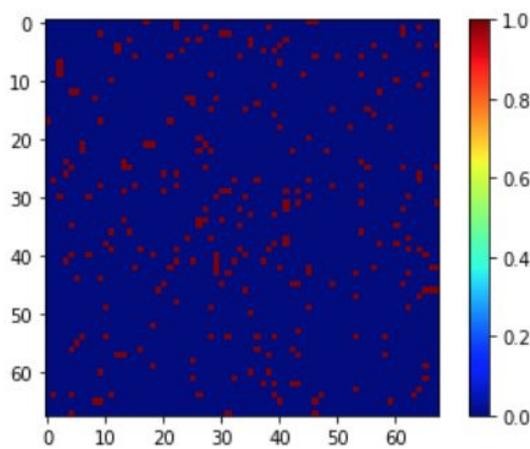
***Figure 17: The probability graphs of the gold standard at different sample size  
Top: 10 samples, middle: 15,  
Bottom: 20***



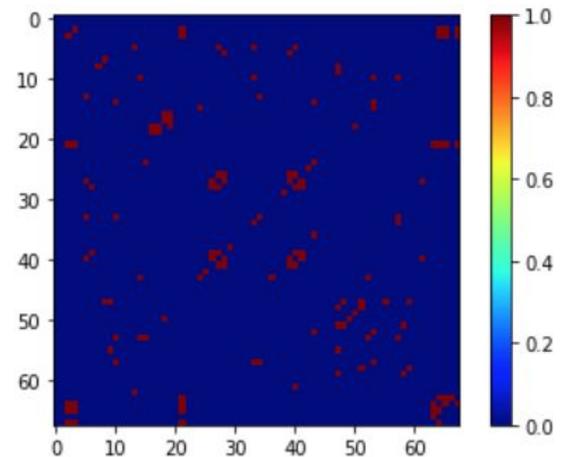
***Figure 18: The probability graphs of the predicted graph at different sample sizes  
Top: 10 samples, middle: 15,  
Bottom: 20***

### 3.5 Comparing the two algorithms

In order to verify that the MCMC is better at predicting truly missing edges, the algorithm is compared to a brute force method of performing a one sample t-test on the mean of 9 sessions and the remaining (prediction) session. The t-test produces a matrix where each edge has an associated p-value of it existing in the average of the 9 sessions. These p-values are thresholded above 0.95 (  $\alpha = 0.95$ ). Figure 19 shows all the edges predicted as missing according to the conventional method ( edges with p values  $> 0.95$ ). This can be compared to figure 20, which shows all of the predicted missing edges from the MCMC algorithm.



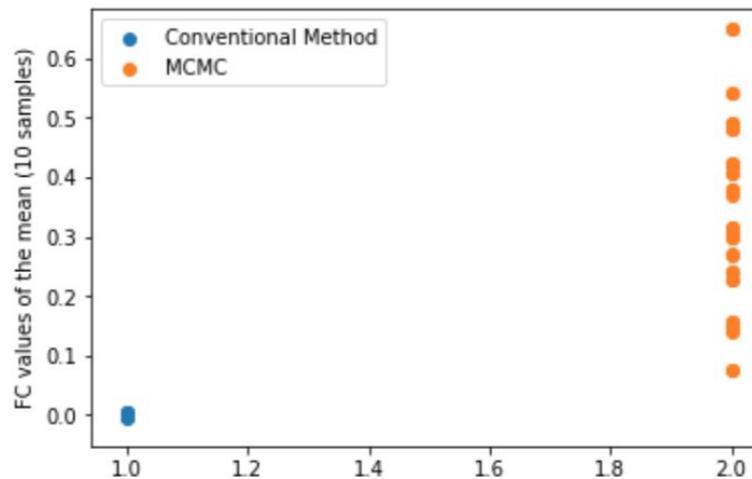
***Figure 19: The predicted missing edges across all permutations in accordance to the convention method.***



***Figure 20: The predicting missing edges across all permutations through the MCMC Method***

Then, the probability graph from the individual session is subtracted from the gold-standard probability graph. The differences between the two are then sorted in order

of decreasing probability. The same number of edges as in the thresholded p-values are extracted from the sorted data. These represent the predicted missing edges based on the conventional method and the MCMC algorithm respectfully. These edges are then plotted against the original FC values. Figure 21 reveals that the edges that the MCMC algorithm predicted are ones with higher functional connectivity, while those predicted by the conventional method have FCs 0.0-0.2 range. Evidently, the MCMC algorithm is better at predicting edges that truly exist since these are the ones with the highest correlation.

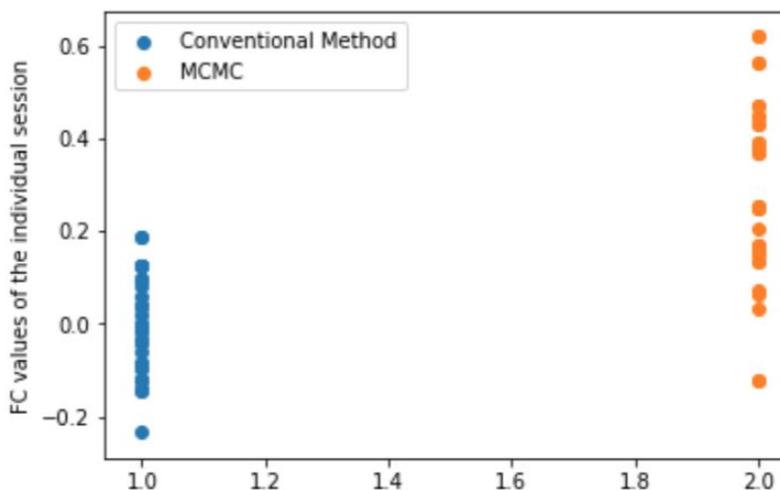


**Figure 21: The FC values of the predicted missing edges based on the MCMC algorithm (x-axis = 2) and the conventional method (x axis =1)**

Figure 21 is the graph for one permutation of the data, where the gold standard includes sessions 2-10 and the prediction graph is session 1. 71.4% of the edges predicted by the MCMC algorithm have a functional connectivity of greater than 0.4, while 0% of the conventional method predicts edges with the same FC.

This process is run through all permutations of nine versus one. Across all permutations, the MCMC algorithm predicts missing edges with higher functional

connectivities than the conventional method. On average, 61.6% of the predicted edges (across all permutations) have functional connectivities that are greater than 0.4, while the conventional method has 0% of the predicted edges with this high of an FC value, suggesting that the conventional method predicts missing edges randomly and that the differences are due to random variability, rather than statistically relevant edges. The mean FC values predicted by the MCMC method is 0.0038, while that of the conventional method is  $3.41e-05$ . The standard deviation of the predicted missing edge's FC's for the MCMC method and the conventional method were 0.046 and 0.00076 respectively. Figure 22, is the MCMC method and the conventional approach's predicted missing edges plotted against the FC of the individual for the same dataset.



*Figure 22: The MCMC method vs conventional method on one permutation. The missing edges plotted against the FC of the mean of an individual sample*

## 4. DISCUSSION

### *4.1 Functional Connectivity Matrices*

The FC matrices show a clear clustering of the ROI. The distinct boxes and the symmetry of the graphs not only validate our preprocessing and correlation methods but also the fact that organization of the brain is in distinct cliques which can be observed in the FC matrices. These unique groupings are also consistent over all the sessions of the data, reconfirming that the functional segregation of the brain is indistinct, and unique cliques<sup>16</sup>, and thus allows us to use the HRG method.

#### ***4.2 Positive and Negative Edges***

One of the problems with the sampling function is that it only grabs the positive edges. Consequently, we compare the two probability graphs generated when only positive and negative edges are sampled separately in order to observe the differences between the two. Figures 8 and 9 reveal that there is an inherent structural difference between the networks that govern negative and positive correlations. This is because the negative correlation values in the FC matrix refer to suppressive processes between the regions of the interest. The positive edges graph (figure 9), despite its structural groupings, does not have the same all-square organization that figure 8 has. This could be because the mechanisms that govern additive relations between these edges- mechanisms that cause these ROIs to activate at the same time- are perhaps more complex in their organization. At the very least, the organization of the positively correlated edges are different from that of the negatively correlated ones. We chose to go forth with performing the rest of the experiment by separating the positive and negative edges since

the positive graph is more highly correlated with the mean FC ( correlation value of 0.523 for the positive edges graph versus -0.37 for the negative edges graph).

### ***4.3 Sampling Thresholds***

The issue of finding the optimal sampling threshold is a complicated one that needs to be addressed to ensure that the sampled graphs are accurate representations of the network as a whole while still providing optimal data to produce the HRGs from. This means that the sampling function should grab edges with higher FC since these are the ones with highest probability of existing. And so, a proportional thresholding method is used. To further ensure that only the top percent of a given distribution of edges is sampled, the three different thresholds are tested. When the matrix is multiplied by 1.25 after being fisher transformed, it has the highest correlation with the original FC values, indicating that it is the closest in extracting the relevant edges.

When searching for the optimal sampling size for this experiment, it is clear that 15 samples produced the probability graph with the fewest differences, indicating that it produced probability graphs closest to the global standard and so 15 is chosen as the optimal sample size.

However, all three graphs generated with all the different thresholds revealed probability graphs that looked similar, and with similar correlations to the original FC. This suggests the robustness of the algorithm: there is minimal differences between the probability graphs generated with the different sampling thresholds and sample sizes and

so the algorithm is able to converge to the most likely HRG as long as the graph is oversampled, irrespective of the actual amount of oversampling.

#### ***4.4 Comparing the Methods***

##### ***4.4.1 Missing Edges***

Figures 19 and 20 reveal the edges that the two methods predict as missing from the individual predicted sample. It is clear based on the figures that the conventional method predicts missing edges randomly. Essentially, here, the noise is randomly distributed and so when the edges are predicted as missing, they are predicted at random ( without structure).

The MCMC algorithm however, predicts edges with structural segregations. There is a clear structural organization of the edges that it predicts as missing. This is because by using HRGs, we are better able to separate the noise and the data. The algorithm, through its test of the likelihood of the HRG it produces, predicts missing edges based on the probability that it should exist due to its relation to the surrounding edges. The noise surrounding the data, therefore does not impact the hierarchical structure of the data. Moreover, the sampling function chooses edges with high functional connectivity under the assumptions that the lower FC values suggest spurious connections. Hence, since the algorithm predicts missing edges based on this hierarchy, and thresholds it while sampling for the hierarchy, it is unaffected by noise unlike the convention method.

#### ***4.4.2 Predicting Missing Edges***

The missing edges plotted against their FC reveal that the MCMC method predicts missing edges with functional connectivity greater than 0.4, on average, 61.6% of the time, while the conventional method predicts missing edges with the same functional connectivity 0% of the time. The mean FC of the predicted edges is roughly 0.0038 for the MCMC method, and  $3.41e^{-5}$  for the conventional method, while the standard deviation for the MCMC and the conventional method is 0.046 and 0.00076 respectively. Despite the difference between individual samples, the MCMC predicts missing edges with significantly higher functional connectivity in the mean FC than the conventional method. The low mean FC value for the MCMC method is attributed to the outliers in some individual graphs that have a large difference between two “top” edges. Nonetheless, the mean FC predicted by significantly greater than that of the conventional method. This, again, is attributed to the separation of noise and data in the HRG space that does not exist in conventional methods.

When the missing edges are plotted against the individual FC matrix, the validity of the MCMC algorithm is further emphasized since it predicts edges with low FC values in the individual, i.e. edges that do not exist in the individual matrix, but should exist due to their high FC value in the mean. Whereas, the conventional method predicts edges as missing based on the random differences

between the individual and the mean, and so it extracts edges that have a low FC value in both the mean and the individual since it only accounts for relative differences.

## 5. CONCLUSION

The issue of noisy fMRIs is a significant and daunting one in neuroscience. Most conventional methods of studying differences between fMRIs in different populations assume that the noise is randomly distributed. However, this produces wide variability when predicting these differences, and leads to compounding of the differences as the number of ROI increase and the data available decreases. To tackle this issue, we argue that brain networks should be studied in HRG space - or modelled as hierarchical graphs. The structural segregation and overlapping network of these segregated regions is a known and accepted property of the human cerebral cortex. So, creating HRGs out of the FC data from fMRIs in order to study the differences between different populations not only minimizes the need for large amounts of data from these populations, but also consequently makes it more cost effective. This is especially important when it comes to studying diseases when large amounts of data for people with neurological disorders are unavailable or costly to collect. A method that can provide detailed insight into these brain structures with less data is an appealing feat.

Here, we prove that the MCMC algorithm can harness the hierarchical structure of the brain and accurately predict missing edges. Therefore, it is clear that the HRG method accurately models the brain network while being unaffected by gaussian noise in

most FC data. This method can now be used to study the differences between disease brains and healthy brains, where instead of comparing 9 samples to 1, we can study the differences between brains with neurological diseases. This can also be translated to studying differences in task vs rest studies. The scope of the algorithm is vast and important, and through the missing edges experiment, we prove that it is also feasible.

## REFERENCES

1. Clauset, Aaron, et al. “Hierarchical Structure and the Prediction of Missing Links in Networks.” *Nature*, vol. 453, 2008, pp. 98–101.
2. Friston, Karl J. “Functional and Effective Connectivity in Neuroimaging: A Synthesis.” *Human Brain Mapping*, vol. 2, no. 1-2, 1994, pp. 56–78.
3. Lee, M H et al. “Resting-state fMRI: a review of methods and clinical applications” *AJNR. American journal of neuroradiology* vol. 34,10 (2012): 1866-72.
4. Rogers, Baxter P et al. “Assessing functional connectivity in the human brain by fMRI” *Magnetic resonance imaging* vol. 25,10 (2007): 1347-57.
5. Kashyap, Amrit, and Shella Keilholz. “Dynamic Properties of Simulated Brain Network Models and Empirical Resting-State Data.” *Network Neuroscience*, vol. 3, no. 2, 2019, pp. 405–426.
6. Vaden, Kenneth I et al. “Multiple imputation of missing fMRI data in whole brain analysis” *NeuroImage* vol. 60,3 (2012): 1843-55.
7. Yan, Bowen, and Steve Gregory. “Finding Missing Edges and Communities in Incomplete Networks.” *Journal of Physics A Mathematical and Theoretical*, vol. 44, no. 49, Sept. 2011.

8. Churchill , NW, et al. “The Functional Segregation and Integration Model: Mixture Model Representations of Consistent and Variable Group-Level Connectivity in FMRI.” *Neural Computation*, vol. 28, no. 10, Oct. 2016, pp. 2250–2290.
9. Yeo, B.t. Thomas, et al. “Estimates of Segregation and Overlap of Functional Connectivity Networks in the Human Cerebral Cortex.” *NeuroImage*, vol. 88, 2014, pp. 212–227., doi:10.1016/j.neuroimage.2013.10.046.
10. “Thresholding Functional Connectomes by Means of Mixture Modeling.” *NeuroImage*, Academic Press, 5 Jan. 2018
11. “Proportional Thresholding in Resting-State FMRI Functional Connectivity Networks and Consequences for Patient-Control Connectome Studies: Issues and Recommendations.” *NeuroImage*, Academic Press, 3 Feb. 2017
12. Desikan R. S., Ségonne F., Fischl B., Quinn B. T., Dickerson B. C., Blacker D., ... Killiany R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3), 968–980.
13. Bassett, Danielle S et al. “Dynamic reconfiguration of human brain networks during learning” *Proceedings of the National Academy of Sciences of the United States of America* vol. 108,18 (2011): 7641-6.
14. Damoiseaux, Jessica S. “Resting-state fMRI as a biomarker for Alzheimer's disease?” *Alzheimer's research & therapy* vol. 4,2 8. 15 Mar. 2012, doi:10.1186/alzrt106
15. Schipper, Laura J. De, et al. “Altered Whole-Brain and Network-Based Functional Connectivity in Parkinsons Disease.” *Frontiers in Neurology*, vol. 9, 2018, doi:10.3389/fneur.2018.00419.

16. Kirsch L, Chechik G (2016) On Expression Patterns and Developmental Origin of Human Brain Regions. *PLoS Comput Biol* 12(8): e1005064.