Study of the Brain Functional Network Using Synthetic Data

Somayeh Sojoudi* and John Doyle+

* Langone Medical Center, New York University
+ Computing and Mathematical Sciences, California Institute of Technology

Abstract—The brain functional connectivity is usually assessed with the correlation coefficients of certain signals. The partial correlation matrix can reveal direct interactions between brain regions. However, computing this matrix is usually challenging due to the availability of only a limited number of samples. As an alternative, thresholding the sample correlation matrix is a common technique for the identification of the direct interactions. In this work, we investigate the performance of this method in addition to some other well-known techniques, namely graphical lasso and Chow-Liu algorithm. Our analysis is performed on some synthetic data produced by an electrical circuit model with certain structural properties. We show that the simple method of thresholding the correlation matrix and the graphical lasso algorithm would both create false positives and negatives that wrongly imply some network properties such as small-worldness. We also apply these techniques to some resting-state functional MRI (fMRI) data and show that similar observations can be made.

I. INTRODUCTION

The human brain is organized into complex networks to support various cognitive functions. The study of brain connectivity has been an active area of research in the past few decades, with the goal of understanding the communication networks associated with disjoint brain regions during cognition or at rest [1]–[6]. Brain connectivity is defined based on different measures of statistical dependencies such as correlation, coherence, and phase locking value. The main purpose of this work is to investigate the role of correlation coefficients in the brain functional connectivity studies.

In most fMRI studies, the brain functional connectivity is assessed with the correlation coefficients among the BOLD signals of regions of interest (ROIs), followed by a thresholding at some pre-specified level [3], [4], [7]–[10]. The thresholded correlation matrix represents the strongest connections/correlations and it can be mapped to a graph showing the connectivity between ROI’s or selected voxels. Based on this technique, some previous studies have shown that the human brain follows a small-world topology [11]–[13]. However, the relationship between the functional networks obtained from various techniques, such as thresholding the correlation matrix, and brain anatomical connections is not well understood. In this work, the aim is to study this problem via a physical electrical network consisting of capacitors and resistors. In particular, we demonstrate that the thresholding technique may not fully describe the physical structure of a network.

In probability theory and statistics, partial correlation is the measure of dependence/independence between two random variables conditioned on the rest of the variables. As an example, consider a simple network of three nodes labeled as $A$, $B$ and $C$, where each node represents a variable component. Assume that node $B$ is connected to nodes $A$ and $C$, and that nodes $A$ and $C$ are not connected to each other. Although there is no direct connection between $A$ and $C$, there is an indirect correlation between them through the common node $B$. Partial correlation, however, partials out the effect of the third variable (i.e., $B$ in this case) and correctly indicates that there is no direct link between $A$ and $C$. The partial correlation matrix can simply be computed by first inverting the correlation matrix and then normalizing it, provided the correlation matrix is positive definite. However, in many practical situations, the sample correlation matrix is not always invertible due to an insufficient number of samples. As an alternative, one might threshold the sample correlation matrix, with the hope that the thresholded matrix has the same structure as the partial correlation matrix.

In this work, we study the performance of the simple method of thresholding the correlation matrix, in addition to some other methods that are commonly used in the literature for estimating the partial correlation matrix. In the next section, we first design a physical network of electrical components and show the relationship between the physical structure of the network and the sparsity pattern of the partial correlation matrix. We then study two network structures, namely tree and mesh grid, and compare the results of four different techniques—thresholding the correlation matrix, graphical lasso algorithm, Chow-Liu algorithm and partial correlation—on synthetic data generated by our circuit models. Finally, we apply these techniques to the resting-state fMRI data of 20 healthy subjects and compare their networks.

II. PHYSICAL NETWORK

Consider the resistor-capacitor (RC) circuit drawn in Figure 1(a). Suppose that the nodes of the circuit, $n_i$ for $i = 1, 2, \ldots, n$, play the role of some brain regions that are connected through some physical links modeled by RC elements. Figure 1(b) is a graph representation of the RC circuit depicted in Figure 1(a). Assume that the physical
structure of the circuit is unknown and only the nodal voltages are available for measurement. It is desired to find the structural connectivity of the circuit from the measured signals. Given a time instance \( t \), let \( V(t) \) denote the vector of the voltages for nodes \( 1, \ldots, n \) at time \( t \). Similarly, let \( I(t) \) denote the vector of the currents injected to the nodes from some external devices at time \( t \). It follows from the Kirchhoff’s current law that

\[
CV(t) = -GV(t) + I(t)
\]

where \( C \) and \( G \) are \( n \times n \) capacitance and conductance matrices, respectively. Suppose that \( E = \{e_1, \ldots, e_q\} \) denotes the edge set of the graph representation of the circuit model. With no loss of generality, assume that the values of the resistors and capacitors are equal to 1. Therefore,

- If \( i \neq j \),
  
  \[
  G_{ij} = C_{ij} = \begin{cases} 
  -1 & \text{if } (i, j) \in E \\
  0 & \text{otherwise}
  \end{cases}
  \]

- If \( i = j \),
  
  \[
  G_{ii} = C_{ii} = \alpha_{ii} - \sum_{(i,j) \in E} C_{ij}
  \]

where \( \alpha_{ii} \) denotes the value of both the capacitance and the conductance of the parallel RC circuit that connects node \( i \) to the ground. Note that at least one of the nodes in the circuit must be connected to ground in order to make the matrices \( C \) and \( G \) invertible.

Assume that the circuit elements are subject to white thermal noise \( W(t) \), namely Johnson-Nyquist noise, such that 

\[
E(W(t)W(t)^T) = 2 \times 1_n \text{ for every } t \geq 0,
\]

where \( 1_n \) is the \( n \times n \) identity matrix. Assume also that the external current vector \( I \) corresponds to the stochastic currents coming from the thermal noise. In this case, one can write:

\[
CV(t) = -GV(t) - G^{\frac{3}{2}}W(t)
\]

Let \( \Sigma \) denote the “steady-state” covariance of the voltage measurements, which satisfies the Lyapunov equation

\[
-(C^{-1}G)\Sigma - \Sigma(C^{-1}G)^T + 2(C^{-1}G^{\frac{3}{2}})(C^{-1}G^{\frac{3}{2}})^T = 0
\]

Recall that \( G = C \). Hence, it follows from the above equation that \( \Sigma = C^{-1} \). Note that the sparsity pattern of \( C \) is consistent with the topology of the circuit. Therefore the inverse covariance matrix \( \Sigma^{-1} \) and consequently the partial correlation matrix (i.e., the normalized inverse covariance or inverse correlation matrix) has the same sparse structure as \( C \). Therefore, the partial correlation matrix can reveal the physical connectivity of the circuit. Assuming that the circuit under study has a sparse structure, it can be concluded that

- \( \Sigma \) is generically a dense matrix, being the inverse of the sparse matrix \( C \).
- \( \Sigma^{-1} \), known also as concentration matrix (and equivalently the partial correlation matrix), is sparse, and more importantly its sparsity conforms with the circuit topology.

The above physical model illustrates the fact that the topology of a system may have been encoded in the partial correlation matrix. Now, consider the problem of recovering the circuit topology from the voltage vector \( V(t) \). To address this problem, one can sample the vector \( V(t) \) at different times \( t_1, t_2, \ldots, t_N \) and construct a sample covariance matrix

\[
\Sigma_s = \frac{1}{N} \sum_{i=1}^N V_i(t_i)V_i(t_i)^T,
\]

where \( N \) denotes the number of samples. Note that \( \Sigma_s \) converges to the population covariance \( \Sigma \) as \( N \to \infty \). When \( N \) is finite, two possible scenarios are as follows: i) \( \Sigma_s \) is invertible but the inverse matrix needs to be thresholded to some level due to the error \( \Sigma_s - \Sigma \), and ii) \( \Sigma_s \) is not invertible and therefore alternative methods are required to estimate the inverse matrix. In this work, we study the following four methods using the synthetic data generated by the abovementioned circuit model:

**Thresholding the correlation matrix:** In this method, we first compute the sample correlation matrix, denoted as \( \Sigma^*_s \), and then threshold the matrix to some pre-specified level to make it sparse.

**Graphical-Lasso algorithm:** For a given sample covariance matrix \( \Sigma_s \), the graphical lasso estimates a sparse inverse covariance matrix by minimizing the negative log-likelihood of the data distribution over the space of positive definite matrices while imposing an \( L_1 \) penalty on the matrix solution. This optimization is as follows:

\[
\min_S \text{ trace}(S\Sigma) - \log(\det(S)) + \lambda \|S\|_1
\]

subject to \( S \succeq 0 \)
To plot the error, we form an error vector discussed above to some synthetic data derived from certain maximum likelihood function of the concentration matrix $\Sigma^{-1}$ of the concentration matrix $\Sigma$, whose graph is plotted in Figure 2. With no loss of generality, assume that $B_{ij} = G_{ij} = -1$ for all $(i, j) \in \mathcal{E}$, where $\mathcal{E}$ is the edge set of the graph representation of the circuit. Furthermore, assume that node 5 of the circuit is connected to ground through a parallel RC circuit with $\alpha_{55} = 4$. The capacitance matrix of this circuit, i.e., $C$, is shown as a colored matrix in Figure 3, where each entry is colored according to its value. Assume that the sample covariance matrix $\Sigma_e$ is equal to the population covariance matrix $\Sigma$. In other words, assume for now that there are infinity many samples of the voltages available (i.e., $N = \infty$). The covariance, correlation and partial correlation matrices are plotted in Figure 4.

Note that the partial correlation matrix is sparse and its corresponding graph representation has the same structure as the graph in Figure 2. However, the covariance matrix (and likewise the correlation matrix) is dense. The four methods introduced in the previous section will be applied to the dense sample correlation matrix $\Sigma_e$. Since the tree graph under study has 9 edges, the level of thresholds and the regularization parameter of the graphical lasso algorithm are chosen such that total number of links in each graph becomes equal to 9. In general, since the network structure is unknown, choosing right values of threshold and regularization parameter is non-trivial. The graphs obtained from the aforementioned four methods are shown in Figure 5.

As shown in Figure 5(a), both thresholding technique and graphical lasso result in the same graph with 5 false positives and negatives. The false positives and negatives are shown in red and dashed blue lines, respectively. This is due to connecting one of the central nodes (i.e., node 5 in this...
The covariance, correlation and partial correlation matrices of the tree circuit under study are shown in Figures (a), (b) and (c) for $N = \infty$. The color of each block shows the value of its corresponding entry in the matrix.

Fig. 4: The covariance, correlation and partial correlation matrices of the tree circuit under study are shown in Figures (a), (b) and (c) for $N = \infty$. The color of each block shows the value of its corresponding entry in the matrix.

Fig. 5: a) The graph corresponding to both thresholding the correlation matrix and the graphical lasso algorithm for the tree circuit (the red edges are false positive and the dashed blue lines are false negatives), b) the graph corresponding to both Chow-Liu algorithm and the partial correlation matrix for the tree circuit.

Fig. 6: Error curves for the tree circuit. Green dotted line: Chow-Liu algorithm, blue line: partial correlation, black circles: thresholding the correlation matrix, and dashed red line: graphical lasso algorithm.

Fig. 7: Sample correlation matrix of the tree circuit with $N = 80$. example) to ground via an RC element with relatively high admittance/capacitance. In this case, the grounded node and its neighbors will have low variance and low correlation even though connected, whereas the ungrounded nodes will tend to float with large variance and correlation even though they are not connected. Therefore, thresholding the correlation matrix results in false positives and false negatives. The same line of argument is also valid for graphical lasso. We have shown in [16] that graphical lasso and thresholding correlation have the same support graph under some conditions. This means that their graphs have the same structure although the weights on the edges could be different due to the $L_1$ norm term in the graphical lasso algorithm. Note that Chow-Liu algorithm correctly identifies the true network structure as shown in Figure 5(b). Note that this algorithm is adopted to find the maximum weight spanning tree and therefore it avoids drawing an edge between some of the highly correlated nodes to prevent any loops in the graph. Since there is no sampling limitation in this example and the partial correlation matrix is already sparse, its corresponding graph has exactly $n - 1$ edges which perfectly captures the structure of the network under study.

The error vector $e$ is calculated for each method and plotted in Figure 6. Since the tree graph under study has 10 nodes, each error matrix $E$ is a $10 \times 10$ symmetric matrix and therefore each error vector $e$ has $\frac{10 \times 9}{2} = 45$ elements which are ranked in ascending order.

So far, we have assumed that $\Sigma_c = \Sigma^c$. To illustrate the effect of the number of samples on the performance of the previous four different techniques, we consider again the same 10-node tree circuit but sample each nodal voltage only 80 times, i.e., $N = 80$. The sample correlation, the graph representations of the four methods under study, and their corresponding error curves are plotted in Figures 7, 8, and 9, respectively. For this data set, thresholding the sample correlation matrix results in 5 false positives/negatives. The graphical lasso algorithm creates 4 false positives/negatives, whereas Chow-Liu algorithm and partial sample correlation create 1 false positives/negatives.

Remark 1: Note that although the tree network has short path lengths, it does not have clustering and therefore it fails to be a small-world network. However, the above example
Fig. 8: The graph representations for the tree circuit with \( N = 80 \): a) thresholding the sample correlation matrix, b) graphical lasso algorithm, c) Chow-Liu algorithm, and d) partial sample correlation. The red edges are false positives and the dashed blue lines are false negatives.

Fig. 9: Error curves for the tree circuit with \( N = 80 \). Green dotted line: Chow-Liu algorithm, blue line: partial sample correlation, black circles: thresholding the sample correlation matrix, and dashed red line: graphical lasso algorithm.

Fig. 10: The mesh grid studied in Section II-B.

Fig. 11: The capacitance matrix of the mesh grid network under study.

Fig. 12: The correlation matrix of the mesh grid network under study.

B. Mesh Grid Network

Consider a mesh grid circuit with \( n = 24 \) nodes that are connected to one another through 38 links. The graphical model of this grid circuit is shown in Figure 10. With no loss of generality, assume that \( B_{ij} = G_{ij} = -1 \) for all \((i, j) \in \mathcal{E}\). Furthermore, assume that nodes 8 and 9 of the circuit are connected to ground through parallel RC circuits with \( \alpha_{88} = \alpha_{99} = 2 \). The capacitance matrix \( C \) of this circuit is shown in Figure 11. Assume that \( N = \infty \) and that the sample covariance matrix \( \Sigma_s \) is equal to the population covariance matrix \( \Sigma \). The correlation matrix of this network is given in Figure 12.

Similar to the tree network studied in the previous subsection, we apply the four techniques of interest to this mesh network. The thresholding levels as well as the regularization parameter \( \lambda \) needed for the graphical lasso algorithm are chosen such that the total number of edges for each graph becomes equal to 38 (similar to the mesh network under study). The graphs obtained from these techniques and their corresponding error curves are plotted in Figures 13 and 14.

For this circuit example, the Chow-Liu algorithm creates 15 false negatives due to avoiding the edges that create loops in the graph. Thresholding the correlation matrix and the graphical lasso algorithms result in 12 and 10 false positives/negatives, respectively. The partial correlation graph has
Fig. 13: The graph representations for the mesh grid network under study: a) thresholding the correlation matrix, b) graphical lasso algorithm, c) Chow-Liu algorithm, and d) thresholding the partial correlation matrix. The red and dotted blue lines show the false positives and negatives, respectively.

Fig. 14: Error curves for the mesh grid network. Green dotted line: Chow-Liu algorithm, blue line: partial correlation, black circles: thresholding the correlation matrix, and dashed red line: graphical lasso algorithm.

Fig. 15: The Sample correlation matrix of the resting-state fMRI data of one participant.

Brain connectivity based on resting-state fMRI data of one subject: The sample correlation matrix obtained from the resting-state fMRI data of one of the subjects is given in Figure 15. Three techniques of thresholding the correlation matrix, graphical lasso and Chow-Liu algorithms are applied to the sample correlation matrix and the graphs are plotted in Figures 16(a), (b) and (c), respectively. The nodes of these graphs represent the brain regions (projected to 2D space) and the size of each node reflects its degree, which is the number of other nodes or brain regions it is connected to in the network. Note that all these graphs have $n - 1 = 139$ edges. We should emphasize that this technique might not be physiologically meaningful and is used here for illustration purposes only.

The relationship between the number of false positives/negatives and the number of samples is considered as a future work.

Similar to the tree network studied earlier, thresholding the correlation matrix creates false positives and negatives in light of low variance in grounded nodes and their neighbors as well as high variance and correlation in the remaining ungrounded nodes of the network.

So far, we studied four different techniques using synthetically generated data derived from our circuit model. In the following section, we study these techniques for resting-state fMRI data of 20 healthy subjects.

III. BRAIN FUNCTIONAL CONNECTIVITY

In this section, we apply the four techniques under study to the resting-state fMRI data of 20 healthy subjects. These fMRI data sets are borrowed form [10]. Each data set includes 134 samples of the low frequency neurophysiological oscillations, taken at 140 cortical brain regions (nodes) in the right hemisphere. Using the time series data, the $140 \times 140$ sample correlation matrix for each subject can be computed. Note that the number of samples is smaller than the number of variables (brain regions), and therefore the sample correlation matrices are not invertible. As a result, it is not possible to estimate a partial correlation matrix by taking the inverse of the sample correlation matrix. To deal with this issue, we perform our analysis for the following two scenarios:

I) We first apply the three methods of thresholding the correlation matrix, Chow-Liu and graphical lasso algorithms to the fMRI data of one of the subjects. To simplify the analysis and comparisons, we set the thresholding level and the regularization parameter such that the resulting graphs will have $n - 1 = 139$ edges.

II) We combine all the 20 individual time series data and form a $140 \times (134 \times 20) = 140 \times 2680$ data set. The sample correlation matrix formed for this data set is well-conditioned and invertible. This would make computing the partial correlation matrix plausible. Similar to the above case, the parameters of different techniques are chosen to obtain graphs with $n - 1$ edges. We should emphasize that this technique might not be physiologically meaningful and is used here for illustration purposes only.
neighboring nodes. Comparing this pattern of connectivity with the false positives created by the two techniques of thresholding the correlation matrix and graphical lasso for the synthetic data studied earlier in the paper suggests that the high connectivity in that area could be false positives created by the phenomenon seen in the circuit examples. The similarity in the pattern of highly correlated areas in the correlation matrices of synthetic data (both mesh and tree networks) and the fMRI data is also interesting and supports this observation. This phenomenon, however, cannot be observed in the Chow-Liu graph.

The error curves and the degree distributions of these graphs are plotted in Figures 17(a) and (b), respectively. Comparing the degree distributions of the three techniques indicates that restricting the number of edges to $n - 1 = 139$ results in many disconnected nodes in the thresholded sample correlation and the graphical lasso graphs. In order to obtain fully connected graphs from these techniques, the threshold level and the regularization parameter need to be set lower. This results in dense graphs with a large number of edges, many of which might not indeed reveal direct interactions among distinct brain regions.

**Brain connectivity based on the combined resting-state fMRI data sets:** The sample correlation matrix of the combined 20 fMRI data sets is given in Figure 18. The sample correlation matrix for this new data set is well-conditioned and invertible. As a result, the partial correlation matrix can be computed. This matrix is given in Figure 19.

The graphs for the combined fMRI data obtained from thresholding the sample correlation matrix, graphical lasso algorithm, Chow-Liu algorithm and thresholding the partial correlation matrix are plotted in Figure 20. Interestingly, the graphs derived from thresholding the sample correlation matrix and graphical lasso algorithm are dense again in the lower left corner, similar to the graphs associated with the single fMRI data set analyzed in the previous section. The similarity between the graphs obtained from thresholding the partial correlation matrix and the Chow-Liu algorithm is also noticeable. The error curves and degree distributions for the three graphs are plotted in Figures 21(a) and (b), respectively. The similarity between the graphs derived from
thresholding the correlation matrix and the graphical lasso as well as the analogy between the graphs obtained from Chow-Liu and thresholding the partial correlation matrix can also be observed in the degree distributions of these graphs.

Remark 2: In both fMRI cases studied here, the correlation matrices as well as the graphs obtained from thresholding and graphical lasso have structural properties similarly to those observed in the synthetic data with grounded nodes. This suggests that these techniques should be used with caution in the brain connectivity analysis to avoid any wrong or inaccurate conclusions.

IV. CONCLUSIONS

The objective of this paper is to study some of the commonly used techniques in brain functional connectivity studies. To this end, we apply four well-known techniques of thresholding the correlation matrix, graphical-lasso, Chow-Liu algorithm and partial correlation to some synthetic data derived from a circuit model with some structural properties. We have shown that thresholding the correlation and graphical lasso algorithm are more prone to errors (false positives and negatives). We also applied these techniques to some resting-state fMRI data and observed similar effects.

REFERENCES