# Bayesian Inference to Identify Cognitive Impairment from Functional Connectivity

**Marium Yousuf**
Department of Computer Science
University of Arizona

## Abstract

Alzheimer's Disease is a neurodegenerative disorder that weakens brain connectivity causing poor inter-regional communication. Functional Connectivity refers to the ongoing signal exchange in the brain that is drastically affected due to Alzheimer's. This change can be reflected using Gaussian precision matrix as it allows extraction of sparsity that represents the state of communication between brain regions. A Bayesian model that learns sparsity can be used to identify the cognitive state and further help identify the chances and predict any brain impairments.

## 1 Introduction

Neurons are involved in a signal exchange between brain regions creating a network-like structure within and between those regions. There are two kinds of configurations of such brain structure: anatomical and physiological [1]. Both of these configurations, also known as connections, allow the understanding of the brain as a network, which can be studied using the recent neuroimaging techniques [1], such as positron emission tomography (PET) scans and functional magnetic resonance imaging or functional MRI (fMRI). In this paper, we focus on the physiological connectivity, or functional connectivity, of the brain that describes the mechanical observations and patterns of statistical dependence among brain regions [1]. That is it refers to the constant neuron communication between the regions: when one part of the brain receives signals for a task it shows activity and further sends signals to another regions as such to complete the task received.

Observing this brain connectivity allows great insight into the cognitive state of a subject. Subjects with weaker cognitive state are expected to have a dysfunctional communication, that is to say while some signals sent from one region to another could be slower because of weak adaptivity or reception of the signals. Viewing the brain as a network, this dysfunctional structure can be understood as sparse in comparison to normally functioning brain. Exploiting the insight that functional connectivity provides, here is proposed a Bayesian inference model that learns sparsity structure in the brain network. To narrow down the focus of the problem, we only focus on the of age-related cognitive impairment, namely Alzheimer's disease. This model allows identification of patients with Alzheimer's disease (AD), Mild Cognitive Impairment (MCI), and Normal Control (NC) subjects.

With a goal to learn sparsity and understand the strength of information exchange between brain regions, we consider $p$ number of brain regions as variables and observe how they relate to each other. Considering the complexity of our problem in terms high-dimensions of brain regions, we use probabilistic graphical modeling. This allows to create a Bayesian framework, that incorporates graph uncertainty, focused on conditional independence between variables extracted from the existence of the edges in the underlying graph [2]. More on graphs and related background is discussed in Section 2, together with G-Wishart sampling, which is our chosen approach to draw samples of inverse covariance matrix, or precision matrix, that describes the conditional independence between brain regions. This helps determine how the sparsity of the functional connectivity differs for each of the category of subjects (AD, MCI, and NC).

## 2 Background

### 2.1 Alzheimer's Disease

Alzheimer's Disease (AD) is the most common form of dementia, which is a neurodegenerating illness that results in severe lifestyle changes. Development of AD gradually affects the cerebral cortex, which is a functional part of the brain operating on the sense perception, muscle movements, and assessment skills like problem-solving and critical thinking [3]. Other than this, the hippocampus - the brain region most affected by the AD - is disrupted affecting memory [4; 5].

AD is a progressive disease as it gets worse with time and is often only diagnosed when the patient is already showing extreme symptoms. Due to this, intensive research is dedicated to investigating ways to delay, cure, or prevent the progression of AD after, sometimes early, diagnosis [6]. In the U.S. alone, the number of patients with AD is estimated at 5.5 million [7]. Typical life expectancy after an Alzheimer's diagnosis being four to eight years [7], on diagnosis, the annual cost incurred to patient care is estimated at over $100 billion per year [6].

### 2.2 Functional Connectivity

There are two types of brain connectivity: structural connectivity and functional connectivity [1]. Structural connectivity concerns the white matter that forms due to long term neuron communication between brain regions, while functional connectivity refers to the active communication that takes place between brain regions through neurons [4]. Both types of brain connectivity eventually change [8].

Neuroimaging techniques reveal that AD is closely related to alteration in the functional connection of the brain and compared to NC, patients with AD show a decrease in the amount of functional connectivity, especially around the hippocampus area [5]. In contrast, there is an increase in the amount of connectivity within the frontal lobe that causes high sensitivity to sense of touch and a sense of visual information [5], revealing the strong connection within the individual regions but weaker communication transfer to another brain regions.

### 2.3 Preliminaries and Notation

#### 2.3.1 Graph Theory

Let $G = (V, E)$ be a graph with $V$, a set a vertices, and $E$, a set of undirected edges. In this paper, we are concerned with two kinds of graphs: *decomposable* vs. *non-decomposable*. A graph is called decomposable if all the cycles ≥ 4 have a chord/edge between two vertices that are not in the vertex set of that cycle. Therefore, any graph that does not meet the said property of decomposable graph is called a non-decomposable graph.

Another important concept for this paper is of *prime components*. Prime components is related to the concept of graph decomposition (irrelevant to the decomposable/non-decomposable graph discussed above). Lauritzen (1996) [9] defines graph decomposition as a partitioning of $V$ into a three subsets: $A, B, S$ such that $A \perp\!\!\!\perp B|S$, and S is complete, called a separator set. A graph is called a prime component if it cannot be properly decomposed.

Built on top of the concept of graph decomposition, we used a *junction tree*, which is a tree representation of the prime components [10]. Each vertex of a junction tree represents a prime component and each edge represents a separator set, $S$, such that $v_i \in S$ are the common vertices from the two prime components that $S$ connects.

The purpose of using a junction tree is to decompose the graphs representing the sparsity structure induced from *Gaussian precision matrix*, which is the inverse covariance matrix from a Gaussian distribution that help capture the sparsity. The joint probability distribution of multivariate normal

$$f(x) \propto \exp\{-1/2(x-\mu)^T \Sigma^{-1}(x-\mu)\},$$

where $\Sigma^{-1}$ is the precision matrix and it is equivalent to adjacency matrix (plus identity matrix) of any undirected graph.

### 2.3.2 G-Wishart Distribution

Given a $G = (V, E)$ and a Gaussian precision matrix, $\theta$, when $\theta$ represents conditional independencies from $G$, we say that the *G-Wishart distribution* is a conjugate prior for $\theta$ [11]. That is, G-Wishart is a distribution over positive definite matrices and it represents the conditional independencies of a Gaussian graphical model. This property of applying $G$ to $\theta$ is characterised by the parameter set, $M^+(G)$, known as the cone of positive definite matrices [12; 11]. All the non-zero elements of these matrices correspond to the edges of $G$. Another significant concept related to precision matrix is of Cholesky decomposition, which is a decomposition of a positive definite matrix into the product of a lower (or upper) triangular matrix and its transpose [13]. For instance, given a matrix $D \in M^+(G)$, then $D$ can be written as a Cholesky decomposition $D = T \cdot T^T$, where $T$ is a lower (or upper) triangular matrix.

### 2.3.3 Rejection Sampling

Rejection sampling involves approximating a target distribution, $p$, using a proposal distribution, $q$. Samples are drawn from $q$ where direct sampling from $p$ is intractable. There are two main properties of rejection sampling:
1) $q$ is scaled by some factor, say $k$, such that $kq > p$, and
2) $\exists u$ that is uniformly sampled between $[0, 1]$.
If $u \leq \frac{p}{kq}$, then the sample is accepted, otherwise rejected.

## 3 Problem Statement

The functional connectivity of the brain reveals the state of cognitive impairment. Exploiting this knowledge, we would like to observe if given a sparse connectivity state - learned from the precision matrix using stimulated data - the cognitive state of a patient can be revealed. Diagnosing cognitive impairment itself is challenging as it requires intrusive and extensive techniques and so predicting impairments introduces further difficulties. Since functional connectivity can capture the deviation in expected brain communication and compromised brain communication, learning it can prove a great facility, helping categorizing subjects into AD, MCI, and NC groups, each of which have a certain sparsity structure to their brain connectivity.

## 4 Related Work

There is plenty of literature available that discusses the identification of AD from functional connectivity. Some papers mentioned here provide models on inferring functional connectivity, with a couple particularly focused on AD. Huang et al. [5] provides a Graphical LASSO inspired model called Sparse Invariance Covariance Estimation (SICE) that identifies functional brain connectivity networks from PET data. Taking $p$ number of brain regions, they assume their data to follow a Gaussian distribution and use SICE to find an estimate for the precision matrix of the $p$ brain regions using the $l_1$ regularization with the following optimization.

$$\hat{\theta} = \text{argmax}_{\theta>0} \log(det(\theta)) - tr(S\theta) - \lambda||\theta||_1$$

$\hat{\theta}$ is the estimate of precision matrix $\theta$, $S$ is the sample covariance matrix, and det(·) and tr(·) are determinant and trace respectively. Their model assumes the precision matrix as a non-random quantity and they use $l_1$ regularization from Graphical LASSO to induce sparsity.

Similar (SICE) model is provided in Sun et al. [6] but with a focus on anatomical changes (structural connectivity) due to long-term impaired functional connectivity. They look at the connectivity patterns that provide image-based ways to distinguish between NC, MCI, and AD subjects. They propose an algorithm based on "block coordinate descent approach" to estimate inverse covariance matrix that has a "user feedback" feature instilled in the estimation process. They apply this algorithm, and provide supportive experimental results, to PET images of 232 NC, MCI, and AD subjects that helps discover the connectivity patterns and differences between the three categories.

Deligianni et al. [14] also focuses on structural connectivity and observes it across subjects to predict functional connectivity. They present a probabilistic framework using cross-validation to learn the

covariance structure of the brain. They consider the correlation of the parameters and introduce a model, based on conditional independence structure of structural connectivity, with a loss function independent to the parameters so to maintain the property of positive definite matrices that match the functional connectivity. Their results emphasize that using statistical learning, functional connectivity can be extracted from the anatomy of the brain.

Hinne et al. [8] propose a Bayesian framework for the estimation of functional connectivity in the presence of uncertainty. They also use the precision matrices to understand the correlations between brain regions and depict the ongoing communication between them. They propose a Bayesian model that estimates a posterior density over precision matrices to analyse functional connectivity, work very close to ours but without the consideration of cognitive impairments. They use diffusion imaging data and test their model using simulated data as well as resting-state fMRI data, eventually comparing their approach to the graphical lasso.

Apart from the neuro-scientific aspect, sparsity matrix is also used to investigate other biological datasets. For instance, Knowles et al. [15] takes a Bayesian approach treating sparsity connectivity matrix as a random variable to model gene expression data of increasing complexity. In their paper, they show that inducing sparsity from datasets for E. Coli and breast cancer data improves predictive performance together with providing an easy interpretation.

# 5 Approach

## 5.1 Model

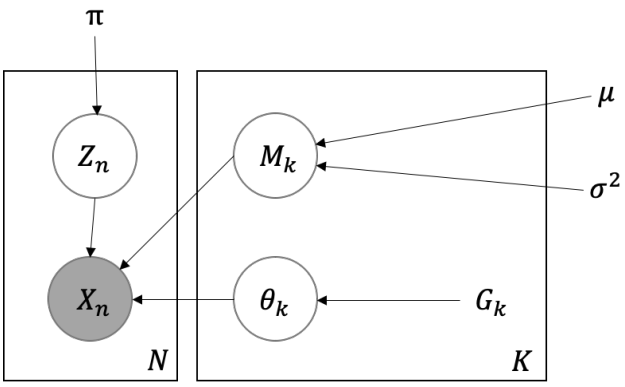


Figure 1: Mixture Graphical model representing the Bayesian network for the proposed where $K$ = 3 representing categories: AD, MCI, and NC and N number of subjects. $G_k$ is G-Wishart prior on Gaussian precision matrices, $\theta_k$, $M_k$ is the Gaussian mean with hyperparameters $\mu$ and $\sigma^2$. $Z_n$ defines the cognitive state for each subject, while $X_n$ is the observed data defining the sparsity for each subject.

$$
\begin{gathered}
Z_i \sim \text{CAT}(\pi) \\
M_k \sim N(\mu, \sigma^2) \\
\theta_i = LL^T \text{ such that } L_{i>j} \sim \text{Laplace}(\lambda) \\
X_i | Z_i, M_k, \theta_k \sim N(M_k, \theta_z^{-1})
\end{gathered}
$$

Consider the mixture model from Figure 1. $N$ represents the number of subjects and $K$ represents the three categories: AD, MCI, and NC. For each subject, $\theta_i$ is learned that represents the sparsity in the functional connectivity from the precision matrix. The plan is to learn $Z_i$ that represents the functional connectivity state that defines if the subject has AD or not. The state of each subject is observed, and represented as $X_i$, which is conditionally dependent on $Z_i$, $M_k$, and $\theta_k$. The goal is to compute the posterior $p(Z_i|X_i)$ that reveals the cognitive state given the information associated with each subject.

### 5.2 G-Wishart Sampling

The algorithm below is adapted from the Atay-Kayis et al. [12], Wang et al. [16], and Roverato et al. [13].

Let $p$ be the number of brain regions considered and $G = (V, E)$ be an arbitrary graph that represents the brain connectivity of a subject. The nodes $V = v_1, v_2, \ldots, v_n$ denote the brain regions and $E = e_1, e_2, \ldots, e_k$ denote the edges linking the nodes representing the brain regions that are connected. Note that the total edges $k$ is arbitrary and depends on each subject.

Atay-Kayis et al [12] provides steps to compute samples of precision matrices, $\theta$, from G-Wishart sampling. This sampling distribution consists of a product of independent univariate standard normal distribution and chi-squared distribution that can be extracted from $G$. Following these samples, Wang et al. [16] provides a rejection sampler to either accept or reject $\theta$. Lastly, both these papers use a concept of triangular-completion, denoted $t_{<ij]}$ introduced in Roverato et al. [13]. Since, we are dealing with non-decomposable graphs, for any triangular matrix, we need a tool to in some sort "complete" an incomplete triangular matrix so that the relation $T \cdot T^T$ from Cholesky decomposition is properly defined [13]. Steps to sample $\theta$ are as follows:

1. Given a graph $G$, for each prime component of $G$, $G_{p_i}$, compute a matrix, $D \in M^+(G)$, with an upper triangular Cholesky factor, $T$ such that $D = T \cdot T^T$, where $T = (t_{ij})_{1 \le i \le j \le p}$.

2. For each $D = T \cdot T^T$, compute a relevant $t_{<ij]}$ such that $t_{<ij]} = \frac{t_{ij}}{t_{jj}}$.

3. Compute an upper triangular matrix, $A$, of size $p \times p$ from adjacency matrix of $G_{p_i}$ plus an identity matrix of size $p \times p$, where any zero element, $a_{ij} = 0$, is off-diagonal and represents missing edges in G.

4. Compute $\bar{V} = W_{\sim V}$, where $W$ is the arbitrary set of vertices if $G_{p_i}$ were a complete graph.

Now, for a sample size, $N$:

5. For $(i, j) \in V$ and $i = 1, ..., p$, sample $\theta_{ii} = \sqrt{U_i}$ where $U_i \sim \chi^2$, and for $i = 1, ..., (p-1)$ and $j = (i+1), ..., p$, if $a_{ij} = 1$, sample $\theta_{ij} \sim V_{ij}$, where $V_{ij} \sim N(0, 1)$.

6. For $(i, j) \in \bar{V}$ and $i = 1, ..., (p-1)$ and $j = (i+1), ..., p$, if $a_{ij} = 1$, compute $\theta_{ij}$:

If $i = 1$ and $a_{ij} = 0$:

$$\theta_{ij} = -\sum_{k=i}^{j-1} \theta_{ik} t_{<kj]}$$

Otherwise, if $i > 1$ and $a_{ij} = 0$:

$$\theta_{ij} = -\sum_{k=i}^{j-1} \theta_{ik} t_{<kj]} - \sum_{r=1}^{i-1} \left( \frac{\theta_{ri} + \sum_{l=r}^{i-1} \theta_{rl} t_{<li]}}{\theta_{ii}} \right) \left( \theta_{rj} + \sum_{l=r}^{j-1} \theta_{rl} t_{<lj]} \right)$$

7. The condition for rejection sampler is as follows:

$$u < \exp\left\{ -\frac{1}{2} \sum_{(i,j) \in \bar{V}} \theta_{ij}^2 \right\}$$

### 5.3 Potential Directions for Validation

#### 5.3.1 Cognitive State Prediction Accuracy

On given the data, $X_{new}$, for a new subject with $N$ brain regions, we would like to be able to classify the subject as either AD, MCI, or NC. Similarly to [5], this can be achieved by comparing the likelihoods of state of the subject in terms of the sparsity of the precision matrix given that the subject is AD, MCI, or NC:

$$f(X_{new}|\theta_{AD}) = \frac{|\theta_{AD}|^{1/2}}{(2\pi)^{K}/2} \exp\left( -\frac{1}{2} X_{new}^T \theta_{AD} X_{new} \right),$$

$$f(X_{new}|\theta_{MCI}) = \frac{|\theta_{MCI}|^{1/2}}{(2\pi)^{K}/2} \exp\left( -\frac{1}{2} X_{new}^T \theta_{MCI} X_{new} \right),$$

$$f(X_{new}|\theta_{NC}) = \frac{|\theta_{NC}|^{1/2}}{(2\pi)^{K}/2} \exp\left( -\frac{1}{2} X_{new}^T \theta_{NC} X_{new} \right),$$

where $\theta_{AD}$, $\theta_{MCI}$, and $\theta_{NC}$ are inverse covariance matrices, derived from a Cholesky Decomposition $L_K L_K^T$, where each $l_{ij}^K$ follows a Laplace distribution with some $\lambda$ and the mean vector considered zero. If

$$f(X_{new}|\theta_{AD}) > f(X_{new}|\theta_{MCI}), \text{ and}$$
$$f(X_{new}|\theta_{AD}) > f(X_{new}|\theta_{NC}),$$

then the new subject would be classified as AD. Similarly, for MCI and NC. Finally, compare with the ground truth by measuring the prediction accuracy.

#### 5.3.2 Comparison to gLASSO for Estimation Accuracy

Huang et. al [5] estimate the precision matrix, $\hat{\theta}$, with $l_1$ regularization to induce sparsity and treat $\hat{\theta}$ as a non-random variable, unlike the estimation proposed here. The SICE model is widely used for estimating brain connectivity [4, 6] and since it uses gLASSO, we would like to measure estimation accuracy of $\theta$ in the proposed model to the $\hat{\theta}$ of gLASSO. One way to achieve this is by minimizing matrix norm.

## 6 Experimental Results

### 6.1 Clique

To validate the performance of the G-Wishart sampler, we first tested it on two complete graphs. Since in this case $W = V$, hence, $\bar{V} = \varnothing$. Therefore, the acceptance rate was 100% because the exponential condition in the rejection sampler was always 1. Moreover, Step 6 was entirely skipped.

### 6.2 Non-Clique

We also tested the G-Wishart sampler on non-decomposable graphs. For simplicity, $p = 7$ was arbitrarily chosen. Figure 2. shows instances of the graphs used to test the sampler.

Consider the matrix, $A$, shown below. The upper triangular matrix from Step 1 computes this matrix for a graph with 7 vertices and two 3-cycles and two 4-cycles.

$$\text{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

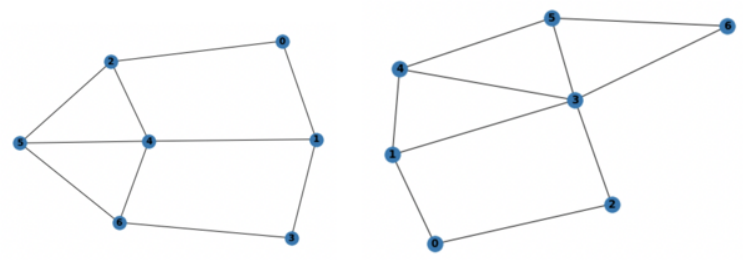

Figure 2: Two examples of non-decomposable graphs used to test the G-Wishart sampler on. Note that all the cycles $\geq 4$ have no chords passing through the vertices in the cycle. The matrix above corresponds to the graph on the left.

### 6.3 Junction Tree Decomposition

Consider another instance of a non-decomposable graph in Figure 3. We computed a junction tree decomposition function, known as *treewidth_decomp()* in Python library NetworkX with a heuristic that counts the number of edges used while computing a smallest possible clique from a chosen vertex and its adjacent vertices and returns a vertex with smallest edge count. This results in the tree decomposition shown in Figure 4.

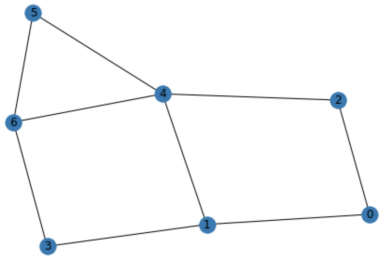

Figure 3: Non-Decomposable graph used for junction tree decomposition shown in Figure 4.

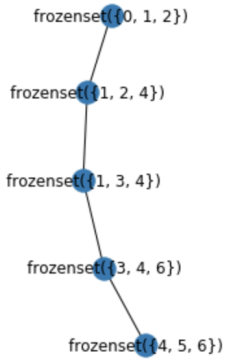


Figure 4: Junction tree decomposition for a non-decomposable graph shown in Figure 3.

As it can be observed, each vertex in a tree decomposition in Figure 4. corresponds to a prime component, while a seperator set can be computed taking an intersection of vertices from a chosen prime component and it's adjacent vertex.

## References


[1] Olaf Sporns. Structure and function of complex brain networks. *Dialogues in clinical neuroscience*, 15(3):247, 2013.

[2] Mohammadi and Abdolreza Ernst C. Wit. Bayesian structure learning in sparse gaussian graphical models. *Bayesian Analysis*, 10(1):109–138, 2015.

[3] Medical Art Library. Cerebral cortex - functional areas., 2017. URL `https://medicalartlibrary.com/cerebral-cortex/`.

[4] Delbeuck X. Alzheimer' disease as a disconnection syndrome? *Neuropsychology review*, 13(2): 79–92, 2003.

[5] Huang Shuai. Learning brain connectivity of alzheimer's disease by sparse inverse covariance estimation. *NeuroImage*, 50(3):935–949, 2010.

[6] Liang Sun. Mining brain region connectivity for alzheimer's disease study via sparse inverse covariance estimation. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM:1335–1344, 2009.

[7] Alzheimers News Today Erum Naqvi. Alzheimer's disease statistics, 2019. URL `https://alzheimersnewstoday.com/alzheimers-disease-statistics/`.

[8] Max Hinne. Structurally-informed bayesian functional connectivity analysis. *NeuroImage*, 86: 294–305, 2014.

[9] Steffen L. Lauritzen. *Graphical models.* Clarendon Press, 1996.

[10] Carlos M. Carvalho. Simulation of hyper-inverse wishart distributions in graphical models. *Biometrika*, 94(3):647–659, 2007.

[11] Wang and Hao Sophia Zhengzi Li. Efficient gaussian graphical model determination under g-wishart prior distributions. *Electronic Journal of Statistics*, 6:168–198, 2012.

[12] Aliye Atay-Kayis. A monte carlo method for computing the marginal likelihood in nondecomposable gaussian graphical models. *Biometrika*, 92(2):317–335, 2005.

[13] Alberto Roverato. Hyper inverse wishart distribution for non-decomposable graphs and its application to bayesian inference for gaussian graphical models. *Scandinavian Journal of Statistics*, 29(3):391–411, 2002.

[14] Fani Deligianni et al. A probabilistic framework to infer brain functional connectivity from anatomical connections. *Biennial International Conference on Information Processing in Medical Imaging.*, Springer(Berlin, Heidelberg), 2011.

[15] Knowles and David Zoubin Ghahramani. Nonparametric bayesian sparse factor models with application to gene expression modeling. *The Annals of Applied Statistics*, 5(2B):1534–1552, 2011.

[16] Wang and Hao Carlos M. Carvalho. Simulation of hyper-inverse wishart distributions for non-decomposable graphs. *Electronic Journal of Statistics*, 4:1470–1475, 2010.