Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced 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 or dissertation in whole or in part in all forms of media, now or hereafter known, 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 or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Yikai Wang Date

Statistical Methods for characterization and classification of brain functional networks : with application to Philadelphia Neurodevelopmental Cohort study

By

Yikai Wang

MPH

Emory University

Rollins School of Public Health

Department of Biostatistics

_____ [Chair's signature]

Ying Guo

_____ [Member's signature]

Suprateek Kundu

Statistical Methods for characterization and classification of brain functional networks : with application to Philadelphia Neurodevelopmental Cohort study

By

Yikai Wang

B.S., South China University of Technology, 2013

MPH, Emory University

Rollins School of Public Health

2015

Thesis Committee Chair: Ying Guo, PhD

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Public Health

in Biostatistics

2015

Abstract

Statistical Methods for characterization and classification of brain functional networks : with application to Philadelphia Neurodevelopmental Cohort study

By Yikai Wang

Functional magnetic resonance imaging(fMRI)-based Brain Network analysis has stimulated an enormous amount of interest in recent years. Network-oriented research on brain plays the key role in understanding the complex brain architecture and functional organization. Given its importance, mapping of brain functional connectivity is a highly challenging task due to the enormous number of connections in the brain and the many confounding factors that lead to difficulty in estimating the true functional connection. In our thesis, I develop two statistical methods for better characterization and classification of brain functional networks. In the first part, I present a statistical strategy for estimating partial correlation matrix for characterizing brain network. Compared to the commonly used correlation, partial correlation can provide more accurate assessment of direct connection by controlling the confounding effects from other brain regions. The proposed approach can overcome the major technical difficulties that have prohibited reliable estimation of the partial correlations in neuroimaging data. We applied the proposed partial correlation approach on a sample of 505 subjects from Philadelphia Neurodevelopmental Cohort(PNC) study to investigate the sex-related difference in brain connectivity. I also estimated the network using the standard correlation. The results showed that the partial correlation based network discovered more relatively significant differential modules compared with the correlation based network and tended to find more within-module differences while the correlation network discovered more between-module differences. Moreover, partial correlation network achieved a higher sex classification accuracy than correlation network based on cross validation. In the second part, I propose to classify brain networks in PNC study using symmetric-positive-definite(SPD) kernel based PCA method. This method provides a compact representation of the high-dimensional brain network connectivity. I applied SPD-PCA method on PNC study to classify male and female subjects based on their brain functional connectivity patterns. Our method achieved a high accuracy rate(73.27%) in the classification using only 60 features. In comparison, the standard approach based on the precision matrix had an accuracy rate of 69.70% using as many as 34980 features. In summary, these findings highlight the advantages of more efficient and advanced statistical tools in characterization and classification of brain functional networks.

Statistical Methods for characterization and classification of brain functional networks : with application to Philadelphia Neurodevelopmental Cohort study

By

Yikai Wang

B.S., South China University of Technology, 2013
MPH, Emory University
Rollins School of Public Health
2011

Thesis Committee Chair: Ying Guo, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Biostatistics
2015

Acknowledgements

I am really thankful for the faculty, advisors and staff of the Biostatistics Department at Rollins School of Public Health for these two years of learning and training. I learnt a lot of things here and this thesis is only a sample of the training and knowledge I had during these two years. I would especially say thank you to Dr. Ying Guo for her help and guidance for this thesis and thank you for Dr. Suprateek Kundu for his time to read this thesis. Moreover, I need to thank Phebe Kemmer for your help in my thesis. In the end, I want to say thank you to my parents for their support. You are the only reason I can have the opportunity studying here. Thank you very much!

Table of Contents

Chapter 1

Introduction

With the help of the advances in neuroimaging and statistic methods, the whole-brain functional magnetic resonance imaging (fMRI) connectivity analysis study have stimulated an enormous amount of interest in these years, [1][2][15][16][17]. For example, the Philadelphia Neurodevelopmental Cohort (PNC) is being funded by National Institute of Mental Health (NIMH) in year 2009 and simultaneously, the National Institutes of Health (NIH) has launched the Human Connectome Project (HCP) to study the brain connectivity and the brain development. Both projects are aimed at facilitating the study of Brain Connectivity, also known as Brain Networking Study. A commonly used tool to study the brain is resting-state fMRI (rs-fMRI), which measures spontaneous low-frequency fluctuations in blood oxygen level dependent (BOLD) signal in subjects at rest, and has the ability to measure the correlation between different brain regions, [18][19]. These correlations are also of fundamental interest to neuroscientists because they have the potential to reflect the functional connectivity between different brain regions and also can be used to explore the overall network structure of the whole brain, [20] [21].

However, in practice, although correlation analysis of fMRI successfully captures the pair-wise information between brain regions, partial correlation, also known as the conditional correlation, is recommended as a better approach for assessing true functional connection[4]. The reason is that the partial correlation measures the direct connectivity between two nodes in the brain by regressing out potential global or third-party confounding effects from other nodes. In particular, a zero partial correlation implies an absence of association given other nodes. Usually, the inverse of a covariance matrix, known as precision matrix, can be modeled in a sparse manner to

obtain a graph or the functional brain network, because the off-diagonals of a precision matrix have a one to one correspondence with partial correlations, [5]. However, estimating the precision matrix is not as straightforward as the correlation matrix which can be estimated directly from the Pearson correlation coefficients. The estimation of precision matrix involves inversion of a large covariance matrix. With neuroimaging data, this becomes challenging given the high dimension of the data where the number of features is usually larger than the sample size [2]. Consequently, estimating the precision matrix requires a huge computational cost and is often not stable. In addition, the precision matrix needs to satisfy the positive definite condition which further increases difficulty in estimation. In recent years, sparse inverse covariance estimation (SICE) has become a popular approach to estimate the precision matrix in high dimensional settings, because it offers a stable estimation under the sparse regularization, [8]. SICE has be used in neuroimaging area to model the functional brain connectivity. Some of the popular SICE methods are CLIME method, [7], Gaussian graphical models, [8]. One major challenge with SICE is that it often requires choosing a tuning parameter that controls for the sparseness of the estimated precision matrix. In this thesis, I propose a statistical procedure for estimating the partial correlation matrix under SICE framework. This procedure consists of three steps: (a) calculate sample covariance matrix; (b) estimate the precision matrix based on the covariance matrix using the CLIME; (c) derive the partial correlation matrix based on the estimated precision matrix.

Once the conditional graph of brain functional network has been estimated using the precision matrix, the next step is often to perform classification or prediction based on the estimated subject-specific precision matrices. For example, it is of interest to investigate how accurately one can classify the gender groups based on the brain network connectivity captured via the

precision matrix. However, the number of features in the precision matrix, which equals the total number of pair-wise connections, is much larger than the sample size. This curse of dimensionality usually restricts the classification performance of the approach directly using precision matrix as a predictor, [2]. One possible solution to this problem is to map the high dimensional precision matrix into a lower dimensional feature space to extract the most relevant information imbedded among all the pair-wise connections. Principle component analysis (PCA) is the most common choice for dimensional reduction. However, standard PCA is not designed for matrix outcomes which makes it unsuited for summarizing precision matrices. Moreover, precision matrix lies in the positive definite and symmetric space (SPD) which is called the Riemannian manifold, [13][22]. This restriction requires a more sophisticated method to perform dimension reduction in precision matrix.

In this thesis, I propose to obtain a compact representation of precision matrix with a SPD kernel-based PCA. Recently, advances in successfully measuring the similarity of the SPD matrix made the SPD kernel and the Kernel-based PCA a promising tool for summarizing useful information in precision matrices [11]. SPD kernel functions, such as the Stein kernel [12], Log-Euclidean kernel [14] and Cholesky kernel [13], take into account the underlying Riemannian manifold the precision matrix reside in[13][22]. The kernel PCA allows us to extract useful information from the high dimensional feature space of the precision matrix to perform classification and prediction based on neuroimaging data.

I applied the proposed methods in the thesis to the PNC study. Specifically, I applied the first method to estimate the partial correlation matrix for constructing the conditional brain networks based on resting fMRI data in the PNC study. I conducted extensive experimental study to compare the difference between the brain networks represented in full correlation matrix and in

the partial correlation matrix. I also studied the brain network difference between male and female subjects with the PNC dataset based on the full correlation matrix and the partial correlation matrix. I then applied the compact representation approach based on SPD kernel based PCA to classify the gender groups in the PNC based on the fMRI precision matrix. I further compared the classification performance of using the proposed SPD kernel based PCA versus classification using the linear PCA and using the original vectorized precision matrix. I applied support vector machine and the leave-one-out cross validation in all classifications. This thesis is organized as follows: Chapter 2 details the materials and the methods used in this thesis; Chapter 3 presents the experimental results on using PNC dataset; Chapter 4 discusses and concludes this thesis. The figures and tables are shown in the Appendix.

Chapter 2

2 Materials and Methods

2.1 Philadelphia Neurodevelopmental Cohort (PNC) Study and Description

The Philadelphia Neurodevelopmental Cohort (PNC) is a collaborative research between the Brain Behavior Laboratory at the University of Pennsylvania and the Children's Hospital of Philadelphia (CHOP), which is funded by NIMH through ARRA of 2009 [1]. The study targeted at over 9500 individuals from ages 8 to 21 years old in greater Philadelphia area, who went to the Children's Hospital of Philadelphia for a pediatric visit. The participants were the volunteers of the genomic studies of complex pediatric disorders. A subsample of 997 participants of PNC study was acquired via dbGaP at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi ?study_id=phs000607.v1.p1 in September 2014. Among these participants, 881 had resting state fMRI data. Additionally, 515 participants met inclusion criteria evaluating the quality of resting-state image quality. The inclusion criteria is that there could not be more than 20 volumes with relative displacement > 0.25 mm [1]. Of these, 290 are female, 215 are male and other 10 are unknown.

All data were acquired on Siemens Tim Trio 3 Tesla, Erlangen, Germany using the same imaging sequences. Resting-state fMRI was acquired with 124 volumes, TR 3000ms, TE 32 ms, flip angle 90°, FOV 192*192 mm, matrix 64*64 and effective voxel resolution 3.0*3.0*3.0 mm. The more detailed experiment settings and image acquisition can be found in [1].

2.2 Functional Magnetic Resonance Imaging (fMRI) Data Preprocessing

A validated confound regression procedure is performed on time series data for each subject to reduce the influence of subject motion. The first 4 volumes of the functional time series is removed to stabilize the signal, leaving 120 volumes for subsequence preprocessing. Then,

functional time series were band-pass filtered to retain frequencies between 0.01 and 0.1 Hz. The confound regression contained 9 standard confounding signals (6 motion parameters + global / WM / CSF) as well as the temporal derivative, quadratic term and temporal derivative of the quadratic of each. Furthermore, motion-related spike regressors were included to bound the observed displacement. All the preprocessing work is based on FMRIB Software Library (FSL) in Rollins Cluster. The source scripts can be found online from 1000 Functional Connectomes Project (https://www.nitrc.org/projects/fcon_1000).

2.3 Brain Network Construction

I adopt a 264-node networking system to construct the brain network, which is proposed by Power et al. [3]. As the representative of the whole voxel-wise connectivity, these 264 nodes provide a good coverage of the entire brain connectivity. Additionally, compared with the original number of voxels, this network system affords a large amount of dimensionality reduction and computational efficiency, which is particularly helpful for multivariate methods such as support vector machine. Moreover, this 264-node system provides better spatial resolution than standard Anatomical Automatic Labeling (AAL) brain region system which partitions the whole brain only into 116 relatively larger regions. Then, for each subject, a voxel-averaged time series was calculated for each nodes.

2.4 Functional Brain Network Comparison : Correlation vs. Partial Correlation

2.4.1 Graph Construction: Correlation vs. Partial Correlation

We adopted two ways to construct the undirected graph among the 264-node system: correlation matrix and partial-correlation matrix. In the graph model G = (V, E), after nodes, V, are settled down, I need to estimate the edges among each pair of nodes, E. One common and native estimator of E is the sample correlation matrix for all nodes, which is the Pearson Correlation

among all nodes [1][7]. But, essentially, Pearson correlation cannot be used as an indicator of whether these two nodes share direct functional connection. More likely, it results from other nodes, which act as confounders in graph model. Thus, in order to explore a more biologically meaningful network, the conditional correlation is suggested [4]. Conditional correlation, also known as partial correlation, regresses out all the effects from other nodes to correctly estimate the direct functional connection between each pair of nodes. Thus, partial correlation matrix could be a better and more accurate way to estimate the underlying biological brain network. The procedure to estimate the partial correlation matrix is proposed in the following section.

2.4.2 The Proposed Procedure for Estimating Partial Correlation Matrix for Neuroimaging Data

Here, I proposed a procedure to estimate the partial correlation matrix for constructing the conditional brain network. In order to estimate the partial correlation matrix, I firstly need precision matrix S. Precision matrix is the inverse of covariance matrix $S = \Sigma^{-1}$. Thus, after the sample covariance matrix $\hat{\Sigma}$ is calculated from the averaged time series, I need to find a way to estimate the inverse of it. But this step is not straightforward. Due to the high dimension, estimating the inverse matrix is very challenging. There are many requirements an estimator must meet. The first of all is that an estimated precision matrix must be positive definite. Currently, there are many established methods to calculate to precision, such as graphical lasso, CLIME, fastclime and transelliptical graphical method [7][8][9]. Based on the experiments, CLIME method provides the most robust and stable estimator of them. Thus, in this procedure, I will adopt the CLIME algorithm to estimate the precision matrix.

2.4.2.1 A Constrained L1 Approach (CLIME) to Sparse Precision Matrix

The CLIME algorithm is a constrained L1 minimization method for estimating a sparse inverse

covariance matrix. The CLIME estimator Ω of the precision matrix S is defined as:

$$\min \left\| \tilde{\Omega} \right\|_1 \text{ subject to}$$

$$\left| \hat{\Sigma}\tilde{\Omega} - \mathrm{I} \right|_\infty \le \lambda, \qquad \Omega \in \mathbb{R}^{p*p}$$

where λ is the tuning parameter controlling the sparsity of the estimated precision matrix, which ranges from 0 to 1. Larger the λ, more sparse the precision matrix. Here, $\tilde{\Omega}$ is not constrained to be a symmetric matrix. The final CLIME estimator Ω is defined as :

$$\Omega = (\omega_{ij})_{p*p} \quad , \quad \tilde{\Omega} = (\tilde{\omega}_{ij})_{p*p}$$

$$\tilde{\omega}_{ij} = \omega_{ij} I(|\omega_{ij}| \le |\omega_{ji}|) + \omega_{ji} I(|\omega_{ij}| \ge |\omega_{ji}|).$$

Then, the final CLIME estimator Ω is symmetric and [7] also showed that it is positive definite with high probability.

2.4.2.2 Partial Correlation Matrix

After I have the estimated precision matrix $S = \Sigma^{-1} = (\sigma_{ij})_{\mathrm{p*p}}$, I can calculate the partial correlation matrix $Pcorr = (\rho_{ij})_{\mathrm{p*p}}$, where p is number of nodes in the brain network. In this thesis, there are 264 distinct nodes (p = 264). Actually, the off-diagnal part of the precision matrix is proportional to the partial correlation matrix $\rho_{ij} = -\sigma_{ij} / \sqrt{\sigma_{ii}\sigma_{jj}}$. Based on the following equation, the partial correlation matrix can be easily calculated:

$$\hat{P}corr = -diag(\Omega)^{-1/2} \bullet \Omega \bullet diag(\Omega)^{-1/2} + 2\mathrm{I}_p \qquad (1)$$

Then, I have the λ based estimated partial correlation matrix.

One challenging issue in the CLIME approach is the choice of the tuning parameter. In the following, I propose an approach for choosing this tuning parameter. We note that the larger λ is, the more sparse precision matrix becomes, which means faster computation but less

information is retained. When we decrease λ, the precision matrix will retain more information, however, the precision matrix becomes less sparse which means longer computation time. To balance the contradicting goals of retaining more information and having faster computation, I propose the following procedure. Start with a relatively large λ, estimate the precision matrix and summarize the amount of information by the absolute sum of the elements in the precision matrix. Then gradually decrease λ and re-calculate the amount of information retained in the estimated precision matrix. We continues decreasing λ towards zero until the information no longer increase with further decrease of λ. We then pick the largest λ value from which the curve of information reaches the plateau. This procedure aims to balance the goals of retaining more information and keeping the precision matrix sparse. The cross-validation approach is used in this selection procedure.

The overall procedure for estimating the partial correlation matrix is summarized in the following table:

Estimating the Partial Correlation Matrix:
Step1: calculate the sample covariance matrix $\hat{\Sigma}$;
Step2: apply the proposed procedure to choose a tuning parameter λ and calculate the CLIME estimator Ω for the precision based on the chosen λ;
Step3: derive $\hat{P}corr$ based on Equation (1).

2.4.3 Gender Classification based on Functional Brain Network: Correlation vs. Partial Correlation

In this thesis, in order to compare the difference between the correlation-based graph and the partial-correlation-based graph, I studied the brain network difference between male and female. First, I will study the edgewise difference between male and female. Second, I will perform the

multivariate pattern analysis on these networks. Moreover, I adopted the module definition proposed in [23] to study the module difference between male and female. In [23], the whole brain region is classified into 10 functional modules, which are medial visual (med vis), occipital pole visual (op vis),lateral visual (lat vis), default mode network (DMN), cerebellum (CB), sensorimotor (SM), auditory (Aud), executive control (EC), frontoparietal left (FPL), frontoparietal right (FPR). These 10 modules cover 232 out of 264 nodes providing a 88% coverage rate and their locations were visualized using BrainNet Viewer [25], Fig 6.0.

2.4.3.1 Edgewise Comparison

As a first step, I need to look at the edgewise difference between male and female within the correlation network and the partial correlation network to fully understand the difference between these two methods and the male and female. Thus, accordingly, I firstly performed a two-sample t test between male and female at each of the 34716 unique connections in the estimated correlation based network and partial correlation based network. Given the large number of tests in the networking data, it is necessary to perform the correction for multiple testing. In the following study, I set a significant threshold by controlling the false discovery rate ($Q< 0.2$) for both network systems. As discussed above, in order to understand the biological meaning behind the data, I will map the networks into 10 modules defined in [23]. The proportional of the number of significant differential edges is used to describe the connectivity of each module pair. All results are visualized in HEATMAP.

2.4.3.2 Multivariate Pattern Analysis

Edge-wise study alone cannot deal with the complex correlation structure existed in the biological system. In contrast to the edgewise comparison above, multivariate pattern analysis can jointly analyze the complex pattern in the brain networking system. In order to perform

multivariate pattern analysis, I adopt the linear support vector machine (SVM) implemented in LIBSVM [24]. SVM is a very powerful tool for classification and has the potential to discover the important features which are helpful to distinguish different groups. In this study, I utilized the linear SVM to classify between male and female based on their estimated functional brain network using correlation and partial correlation. In the linear kernel, I calculated the weight of each edges directly, of which the absolute value can be the representative of its power to differentiate different genders. Then, similarly, for each module pair, we calculate the mean of the absolute value of the edge weight, which can be used as the indicator of the significance between male and female. Again results are visualized in HEATMAP. At last, we examined the potential of using different functional brain network to classify the male and female using leave-one-out cross validation and linear support vector machine.

2.5 A Compact Represent Method in Functional Brain Network Classification

In the above sections, we talked about brain network and the method to construct the brain network. However, the constructed brain network exists many potential problems. Although the conditional brain network has regressed out the confounding effect from other nodes, the high dimensionality and the scarcity of training subjects of the brain network usually lead to unsatisfying discrimination results. Additionally, the brain network is usually represented as a symmetric matrix. Thus, the traditional vector-based methods such as linear PCA, SVM, are not able to capture all the information a brain network contains. Moreover, as the representative of the brain network, the precision matrix is symmetric positive definite (SPD). This inherent property restricts it into a lower dimensional space rather than the full p*p Euclidean space [2]. All these limitations and ideas drive us to explore a compact represent method to map the brain

network into a compact feature space. Before I come into that, I will firstly introduce the most common used compact mapping method - principle component analysis.

2.5.1 Principle Component Analysis (PCA)

Principle component analysis (PCA) is the commonly used unsupervised dimensionality reduction method in analyzing the high dimensional biological data [10]. As the background of kernel based PCA, I will introduce it under the concept of singular value decomposition. Assume that we have N subjects with p features, $X_1,...,X_N \in \mathbb{R}^p$, $X = (X_1,...,X_N)^T \in \mathbb{R}^{N*p}$ and each X_i has the same covariance matrix $\Sigma \in \mathbb{R}^{p*p}$. Based on the eigen-value decomposition theorem, the covariance matrix Σ can be decomposed as $\tilde{E}\Lambda\tilde{E}^T$, where the ith column of $\tilde{E}$ is the ith eigen-vector of Σ and Λ is a p*p diagonal matrix with the ith diagonal element is the ith eigen-value of Σ. Based on [10], the principle component of X_i is given by $\tilde{E}^T X_i$, which means the principle component matrix is given by

$$Prin = X\tilde{E} \in \mathbb{R}^{N*p} \tag{2}$$

, where each column is the principle component vector of each subject. Numerically, we do not have Σ but an unbiased estimator of it:

$$\hat{\Sigma} = \frac{1}{n-1} X^T X \tag{3}$$

where each each X_i is assumed to have the mean of zero and are mutually independent. Then, based on singular value decomposition, we will have:

$$X = UVE^T \tag{4}$$

,where $U \in \mathbb{R}^{N*N}$, $E \in \mathbb{R}^{p*p}$ and also the unitary matrix and V is a N * p rectangular diagonal matrix with non-negative real numbers on the diagonal. Thus, put the (4) into (3), we have $\hat{\Sigma} = E\tilde{\Lambda}E^T$ and $\tilde{\Lambda} = V^T V/(n-1)$ which is a diagonal matrix. Then, the equation (2) turns into:

$$\hat{P}rin = XE = UVE^T E = UV \tag{5}$$

Furthermore, U is the eigen-vector matrix of XX^T which is exactly the kernel matrix, also known as the similarity matrix, with a linear kernel. One more notice is that although the estimated principle component matrix is N*p, in brain imaging data the number of features is much larger than the sample size leading to the last (p - N) columns being zero. Thus, the estimated $\hat{P}rin$ is actually a N*N matrix which successfully reduce the dimension of the data from p to N and the transposed pc's have the most variance and are mutually independent to capture the original information as much as possible.

2.5.2 Symmetric Positive Definite (SPD) Kernel based PCA method

Although promising, linear PCA cannot be used for the brain network data which is usually a matrix. The precision matrix, estimator of the brain network, is defined to be symmetric positive definite (SPD). With the help of the SPD kernels, the similarity among different precision matrix can be studied. Thus, SPD-kernel-based PCA method can be used for analyzing the brain network data [11].

First, define the SPD space for p*p matrix: $Sym_p^+ = \{\mathrm{A} \mid \mathrm{A} = \mathrm{A}^T, \forall x \in \mathbb{R}^p / \{0\}, \mathrm{x}^T \mathrm{Ax} > 0\}$ and denote F to be the kernel-induced feature space. We assume that there is a kernel mapping $\Phi : Sym_p^+ \mapsto F$ which can capture all the useful information a SPD matrix contains to a feature space. The kernel mapping cannot be explicitly solved but implicitly induced by SPD kernel [2]. Assume there are N precision matrixes $\{\Omega_1, ..., \Omega_N\} \subset Sym_p^+$ and we map them to the feature space $\{\Phi(\Omega_1), ..., \Phi(\Omega_N)\} \subset F$. Without loss of generality, $\Phi(\Omega_1), ..., \Phi(\Omega_N)$ are assumed to be centered [2][11]. Then, we can calculate the kernel matrix $K = (\mathrm{K}_{ij})_{N*N}$ with $\mathrm{K}_{ij} = \left\langle \Phi(\Omega_i), \Phi(\Omega_j) \right\rangle = \mathrm{k}(\Omega_i, \Omega_j)$, where k is the SPD kernel function. The kernel matrix here is

comparable to the XX^T in section 2.6.1. Similarly, we perform the eigen-decomposition on kernel matrix $K = U\Upsilon U^T$, where $\Upsilon = diag(\lambda_1, ..., \lambda_N)$. Here, I assume the number of feature in F is larger than the sample size. Then, the last p-N columns in equation 5 is zero and the principle component matrix will have the format:

$$\hat{P}rin = U\Upsilon^{1/2} \tag{6}$$

, where $\Upsilon^{1/2} = diag(\lambda_1^{1/2}, ..., \lambda_N^{1/2})$. Thus, in SPD-kernel PCA, we do not need to solve the explicit form of the kernel mapping Φ. We only need to select a SPD kernel function k and perform the eigen-decomposition of kernel matrix to have the principle component matrix. Based on the performance test in [2], I will use the Stein kernel (SK) [12] of which the distance function is named root Stein divergence:

$$\mathrm{k}(\Omega_i, \Omega_j) = \exp(-\theta \bullet \mathrm{d}^2(\Omega_i, \Omega_j))$$

$$d(\Omega_i, \Omega_j) = \left[\log(\det(\frac{\Omega_i + \Omega_j}{2})) - \frac{1}{2}\log(\det(\Omega_i \Omega_j)) \right]^{1/2}.$$

2.5.3 Gender Classification based on Functional Brain Network using Compact Representation Methods

Although the promising properties in theory, we still need experiments to study its performances. In the thesis, we utilized the PNC study to evaluate the efficacy of the SPD-kernel PCA method in enhancing the information hidden in the precision matrix. In order to study its efficacy, we performed linear support vector machine and leave-one-out cross validation method. Here, we compared the classification performance of the SPD - kernel based principle component analysis with linear principle component analysis directly performed on the vectorized precision matrix. Moreover, we directly used the vectorized precision matrix as the input features of each subject to study the potential of using the original functional brain networks to distinguish between male

and female. The number of principle components of SPD-kernel PCA and linear PCA is selected by using cross validation procedure.

Chapter 3

3 Experiments and Results

This chapter is divided into two main parts. In the first part, I will investigate on the difference between full correlation matrix and partial correlation matrix based brain network on various aspects. In the second part, the performance of SPD-kernel based PCA will be compared with linear PCA and the original precision matrix using support vector machine and leave-one-out cross validation. All the analysis before is based on the procedure in section 2.4 to estimate the precision matrix and partial correlation matrix. Thus, as a basis, I will firstly talks about the strategy I used in choosing tuning parameter for CLIME algorithm.

3.0 Choosing Tuning Parameter in CLIME

In order to choose tuning parameter, we randomly selected 20 subjects and calculated the precision matrix based on the λ at 0.4, 0.3, 0.2, 0.1, 0.01, 1e-3, 1e-4, 1e-6, 1e-8. For example, the results of one subject is showed in Fig 1.1, 1.2, 2.1, 2.2. In the beginning, we started from a relative larger value 0.4 and gradually decreased it. With the λ decreasing, the shape of the histogram of the precision matrix and the information it contains is becoming stable. The changing point of λ is 0.001, Fig 1.1, 1.2. When λ gets smaller than 0.001, the shapes of histogram for precision matrix are very similar and the same goes for the information. It means the information a precision can capture is becoming stable and saturated. Moreover, heatmap of the estimated precision matrix at .001 shows that it has already contained the main patterns in covariance matrix and is very similar to the one at 1e-8, Fig 2.1, 2.2; also, the pattern in precision matrix at 0.1 and 0.01 are much weaker than 0.001. This kind of trend has also been discovered in most of other subjects. Thus, I will choose $\lambda = 0.001$ for all subjects in the following study.

3.1 Results of Comparing Full-correlated Graph vs. Partial-correlated Graph

We have constructed the brain network based on full-correlation matrix and partial-correlation matrix for 503 subjects, where 289 are females and 124 are males (2 subjects were removed there because they do not have the full-rank covariance matrix). In order to study the difference between these two graphs, we compare the brain network of female versus male. The averaged full and partial correlation matrix for male and female is visualized in Heatmap, Fig 3, which is based on the same color scale. The two networks share the similar main pattern in the diagonal part but the partial correlation matrix is much more concise than full correlation matrix. Moreover, in the full covariance matrix, the block-wise strong signal in diagonal seems like they were jointly generating some complex signal pattern in their overlapped area, which could be the confounder.

3.1.1 Results of Edgewise Comparison

To fully understand the differences in brain network, we compared the edgewise connectivity at each of the 34716 unique edges. In Fig 4, p-value list from correlation based brain network is more skewed towards zero than partial correlation matrix. After multi-testing correction based on false discovery rate, Fig 5, correlation based brain network tends to find more significant edges than partial correlation based brain network ($Q < 0.2$). To understand the spatial pattern of the significant edges, significant differential edges are visualized within the 10 functional brain modules and the proportional of the significant differential edges are used to evaluated the module pair difference between male and female. In Fig 6.1, the significant differential edges of correlation network exist a more strong pattern than partial correlation network. However, when we consider the relative significance of each module, the pattern of partial correlation network

becomes clear, Fig 6.2. Based on the relative significance of the module pair, there are two major difference between correlation network and partial correlation network.

First, the significant differential module pairs of correlation network are more concentrated in the bottom right corner and on the top left corner it becomes much weaker. In contract, the significant differential module pairs in partial correlation network are more evenly distributed across all module pairs where the partial correlation network not only discover the significant differential module pair in the bottom right corner, it also discover some module pairs in top left corner which is not discovered by correlation network. One potential reason is that in correlation matrix, the locally strong information (bottom right corner) would produce confounding effects in its neighbors preventing other differential module pairs from being discovered.

The second major difference between the two networks is that partial correlation based network shows that the within module difference are relatively significant compared with the between module difference. However, in correlation based network, the within module difference are much weaker than between module difference. For example, relative within module difference of lat vis, CB, SM, Aud in partial correlation network are larger than the difference in correlation based network.

3.1.2 Results of Multivariate Pattern Analysis

In order to study the overall complex pattern in brain network, we perform multivariate pattern analysis using the complete multivariate structure of the data with the help of support vector machine. First, we vectorized the correlation matrix and partial correlation matrix and trained the linear SVM classifier using all subjects and features and calculated the absolute weight of each features (edges). The edge with larger weight means it is more helpful to distinguish between male and female. To remove the scale difference existed in two networks, we divided the weight

with their standard deviation and mapped them into the module based system, Fig 7.1 and Fig 7.2.

Based on the same scale, the results are comparable. From Fig 7.1, the important edges in partial correlation network are much sparse than correlation network. Similarly, within module difference are much significant than between module difference in partial correlation network and for correlation network it reversed, Fig 7.1 and Fig 7.2. In Fig 7.2, the distribution of the module-wise difference of the two network have similar pattern with the one in edge-wise analysis. Partial correlation network discovered some strong pattern in the top left corner in the module-wise comparison and correlation network did not. This means that both edgewise analysis and multivariate analysis give consistent results.

At last, in order to evaluate the quality of these two networks, we vectorized the network matrix and performed leave-one-out cross validation with SVM on male and female. As results, full correlation based SVM classification has an accuracy of 74.95% and partial correlation based SVM classification has an accuracy of 79.13%. This means partial correlation matrix contains more useful information than full correlation matrix and it can give more reliable founding than full correlation matrix.

3.2 Results of Using Compact Representation of the Conditional Graph

In this selection, we compared the classification performance of SPD kernel PCA method with linear PCA and the original vectorized precision matrix. The whole 505 subjects are used for classification which is also based on sex. θ in the SK kernel is chosen to be 1/10 which is small enough to capture the similarity between different subjects, where the Heatmap of the kernel matrix is in Fig 8. In order to ensure the number of principle components, we follow the method

in [2] which is tuned out by cross-validation. The classification accuracy plot is shown in Fig 9, which has a parabola shape and is optimal when the number of PC's reach 60. Same goes for the linear PCA method. The optimal classification accuracy for these methods is in Table 1. The SPD kernel based PCA reaches the best classification accuracy at 73.27% with the smallest number of features (60 features). The accuracy of linear PCA is 64.55% which means linear PCA failed to capture the important information in the brain network, and the accuracy for using the whole brain network (34980 features) is 69.70%. This means that compact representation method of the network can increase the information to distinguish between male and female. However, we need to pay attention to the method we use, because using the inappropriate representation method would lead to worse results.

Chapter 4

4 Discussion

Overall, the study included 505 subjects from the PNC study. First, we compared the difference between full correlation network and the partial correlation network with the help of discriminating between males and females with two-sample t test, FDR correction, module-based analysis and support vector machine based multivariate pattern analysis. Second, we examined the performance of the compact representation - SPD kernel based principle component analysis. We compared the classification accuracy between SPD kernel PCA and other methods like linear PCA based on SVM and leave-one-out cross validation. Moreover, we also visualized the constructed brain network in Heatmap for better understanding of these methods. In the following sections, I will summarize the result findings and discuss about the strength and limitations.

4.1 Conclusions and Recommendations

From the results of the comparison between full correlation matrix and partial correlation matrix in section 3.1, the averaged partial correlation matrix showed a similar pattern with the averaged correlation matrix which means under a broad view, these two methods are consistent. However, for the following comparison between male and female, the p-values and FDR from these two brain network exhibited relatively different results. Full correlation matrix tends to discover more significant edges than partial correlation matrix under the same significant level. One reason for this is that the partial correlation matrix already regressed out the confounding effects from all another nodes. Thus, there should be less significant edges than full correlation matrix. Based on the module analysis, correlation network tends to emphasize the most significant differential patterns and ignores other relatively weaker differential pattern which can be

discovered within the partial correlation network. Moreover, partial correlation based network discovered the relatively strong within module difference. However, in the correlation based network, within module difference is much weaker than between module difference. In the multivariate pattern analysis, the differential edges in partial correlation network are more sparse than correlation matrix and the module analysis gives the similar results with the edgewise analysis. At last, the results from the leave-one-out classification based on SVM showed that the partial correlation matrix is more helpful in discriminating between male and female. This could be another support for partial correlation matrix containing more concise and important information than full correlation matrix which has too much confounding effects disturbing the SVM to find the true signals. Thus, in summary, as an estimator of the brain network, the partial correlation matrix is not only more concise and sparse than the full correlation matrix, but it also contains more important and useful information.

From the results of the comparison between the classification accuracy between SPD-kernel PCA and the linear PCA and the original vectorized precision matrix, SPD-kernel based PCA received the highest classification accuracy at 73.27%, higher than linear PCA for over 8% and also higher than original precision matrix for over 3.5%. This means the compact representation method successfully enhanced the important information hidden in the high dimensional precision matrix. Moreover, the performance of linear PCA is even worse than directly using the whole precision matrix. It means that not all compact mapping method is helpful. It needs to consider the structure of the data.

Based on this study, we observed the difference between partial correlation based network and the correlation based network. Thus, in the future study, we suggested that we need to look into both network system and for classification purpose we suggest using partial correlation based

network. Moreover, SPD-kernel based PCA is suggested for dimension reduction for precision matrix. On the other side, when performing the compact representation, we need pay more attention to the method we choose which should be well designed for the data we are dealing with.

4.2 Strengths and Limitations

One strength of this study is that we have more than 500 subjects and the quality of this dataset is very high, which makes the results more reliable. However, this huge dataset makes it hard and time-consuming to analyze the data, which is one of the limitations. Moreover, with the help of the discrimination between male and female, we successfully studied the efficacy of the conditional network and the full correlated network which lead to an important founding. Another limitation is that because of the time limit, we only consider the SK kernel. There are still many SPD kernels like Cholesky kernel [13], Power Edclidean kernel [13] and Log-Euclidean kernel [14], which should be fully studied. One limitation of the SPD kernel PCA is that it can only deal with the precision matrix which is assumed to be positive definite and symmetric. It cannot deal with the partial correlation matrix which is not ensured to be SPD even if the precision matrix is.

4.3 Future Direction

In the future work, we will study the performance of other three interesting SPD kernels and study the potential of directly using partial correlation matrix as predictor. Moreover, rather than the unsupervised learning methods, we are planning to move on to the supervised learning methods, such as kernel linear discriminant analysis (KLDA), to explore discriminative representation.

Reference

[1] Satterthwaite T D, Wolf D H, Roalf D R, et al. Linked sex differences in cognition and functional connectivity in youth[J]. Cerebral cortex, 2014: bhu036.

[2] Zhang J, Zhou L, Wang L, et al. Functional Brain Network Classification With Compact Representation of SICE Matrices[J]. 2015.

[3] Power J D, Cohen A L, Nelson S M, et al. Functional network organization of the human brain[J]. Neuron, 2011, 72(4): 665-678.

[4] Smith, Stephen M. "The future of FMRI connectivity." *Neuroimage* 62.2 (2012): 1257-1266.

[5] Peng, Jie, et al. "Partial correlation estimation by joint sparse regression models." *Journal of the American Statistical Association* 104.486 (2009).

[6] Dempster, Arthur P. "Covariance selection." *Biometrics* (1972): 157-175.

[7] Cai, Tony, Weidong Liu, and Xi Luo. "A constrained ℓ 1 minimization approach to sparse precision matrix estimation." *Journal of the American Statistical Association* 106.494 (2011): 594-607.

[8] Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. "Sparse inverse covariance estimation with the graphical lasso." *Biostatistics* 9.3 (2008): 432-441.

[9] Liu, Han, Fang Han, and Cun-hui Zhang. "Transelliptical graphical models."*Advances in Neural Information Processing Systems*. 2012.

[10] Härdle W, Simar L. Applied multivariate statistical analysis[M]. Springer Science & Business Media, 2007.

[11] Schölkopf B, Smola A, Müller K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. Neural computation, 1998, 10(5): 1299-1319.

[12] Sra S. Positive definite matrices and the symmetric Stein divergence[R]. 2011.

[13] Jayasumana S, Hartley R, Salzmann M, et al. Kernel methods on the riemannian manifold of symmetric positive definite matrices[C]//Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013: 73-80.

[14] Arsigny V, Fillard P, Pennec X, et al. Log‐Euclidean metrics for fast and simple calculus on diffusion tensors[J]. Magnetic resonance in medicine, 2006, 56(2): 411-421.

[15] Biswal, Bharat, et al. "Functional connectivity in the motor cortex of resting human brain using echo‐planar mri." *Magnetic resonance in medicine* 34.4 (1995): 537-541.

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

[17] Deco, Gustavo, Viktor K. Jirsa, and Anthony R. McIntosh. "Emerging concepts for the dynamical organization of resting-state activity in the brain." *Nature Reviews Neuroscience* 12.1 (2011): 43-56.

[18] Dosenbach, Nico UF, et al. "Prediction of individual brain maturity using fMRI."*Science* 329.5997 (2010): 1358-1361.

[19] Ogawa, Seiji, et al. "Brain magnetic resonance imaging with contrast dependent on blood oxygenation." *Proceedings of the National Academy of Sciences*87.24 (1990): 9868-9872.

[20] Church, Jessica A., et al. "Control networks in paediatric Tourette syndrome show immature and anomalous patterns of functional connectivity." *Brain* 132.1 (2009): 225-238.

[21] Seeley, William W., et al. "Neurodegenerative diseases target large-scale human brain networks." *Neuron* 62.1 (2009): 42-52.

[22] Harandi, Mehrtash T., et al. "Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach." *Computer Vision–ECCV 2012*. Springer Berlin Heidelberg, 2012. 216-229.

[23] Smith, Stephen M., et al. "Correspondence of the brain's functional architecture during activation and rest." *Proceedings of the National Academy of Sciences*106.31 (2009): 13040-13045.

[24] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: a library for support vector machines." *ACM Transactions on Intelligent Systems and Technology (TIST)*2.3 (2011): 27.

[25] Xia, Mingrui, Jinhui Wang, and Yong He. "BrainNet Viewer: a network visualization tool for human brain connectomics." *PloS one* 8.7 (2013): e68910.

Appendix: Figure and Table

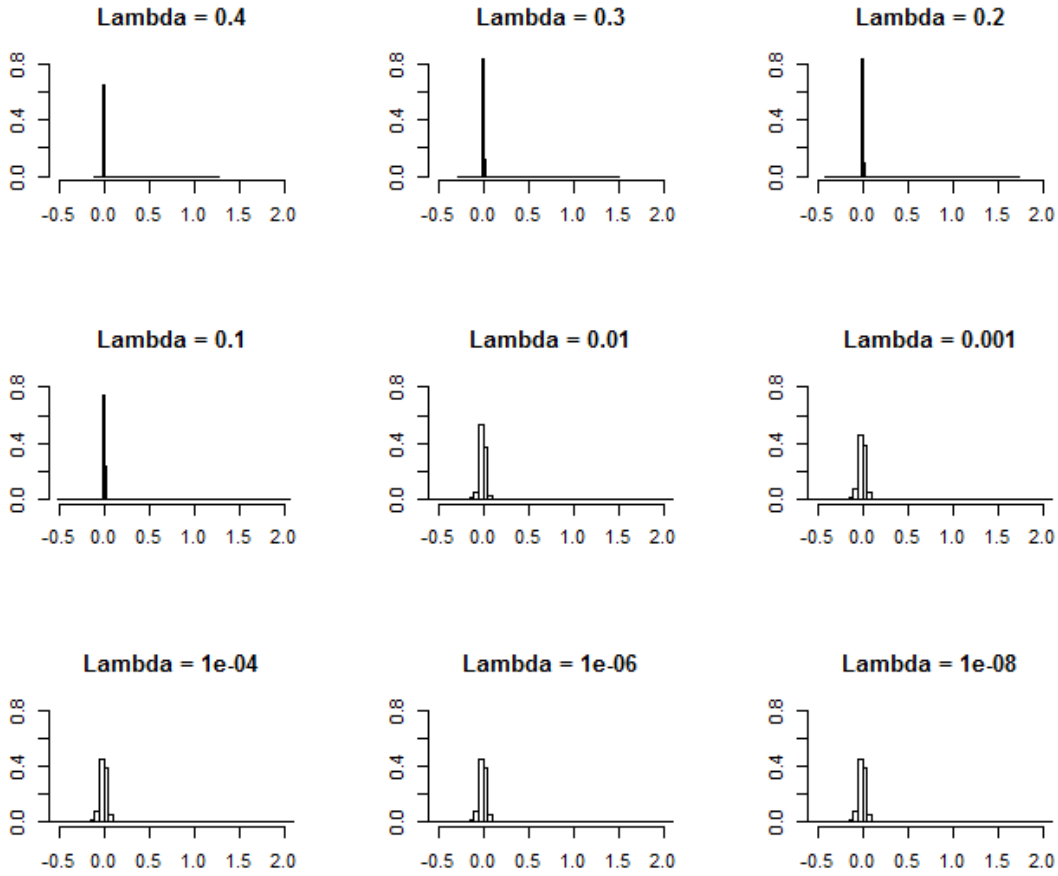

Fig 1.1: Histogram of Precision Matrix on different tuning parameters

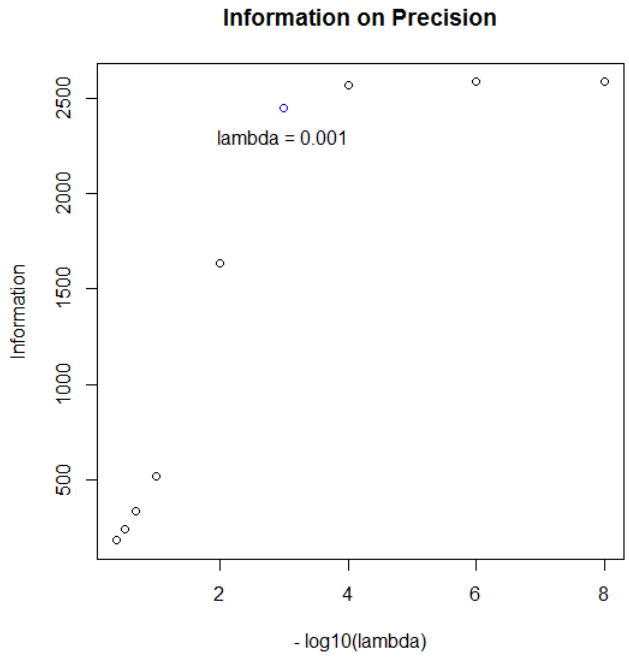

Fig 1.2 : Information in Precision Matrix on different tuning parameters

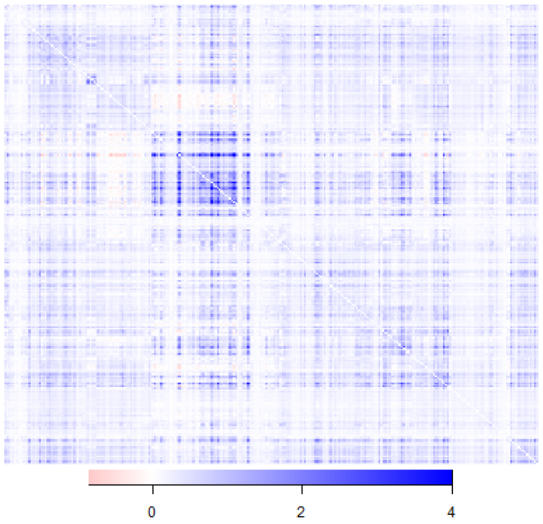

Fig 2.1: Covariance matrix

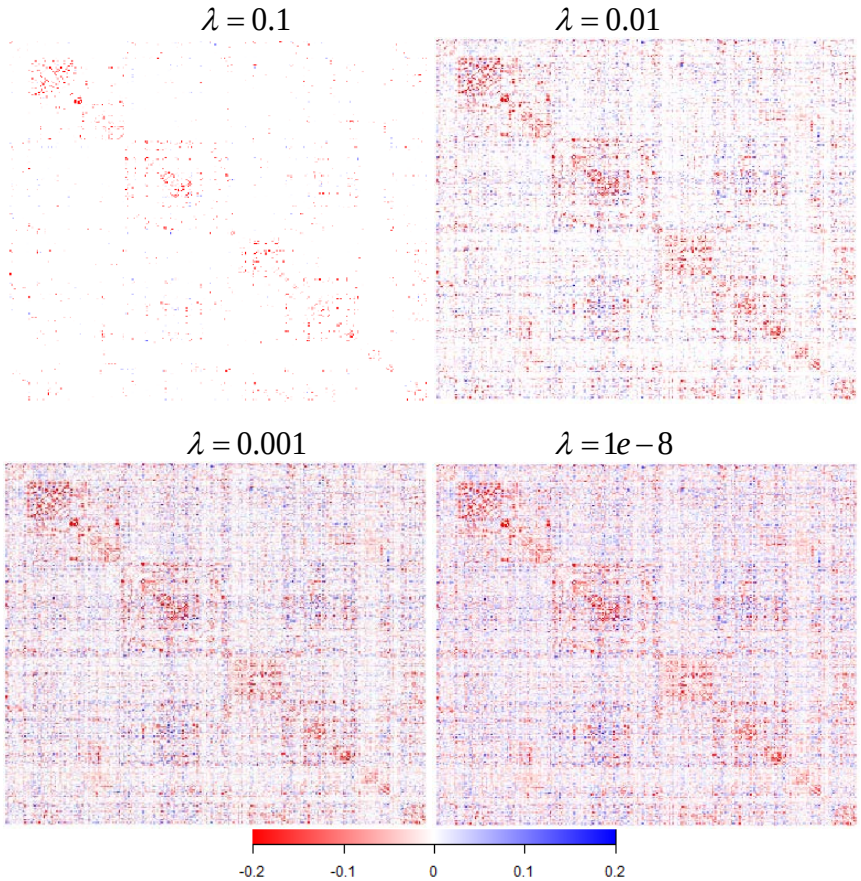

Fig 2.2: Precision Matrix on different tunning parameters (same subject with Fig 1 - Fig 3)

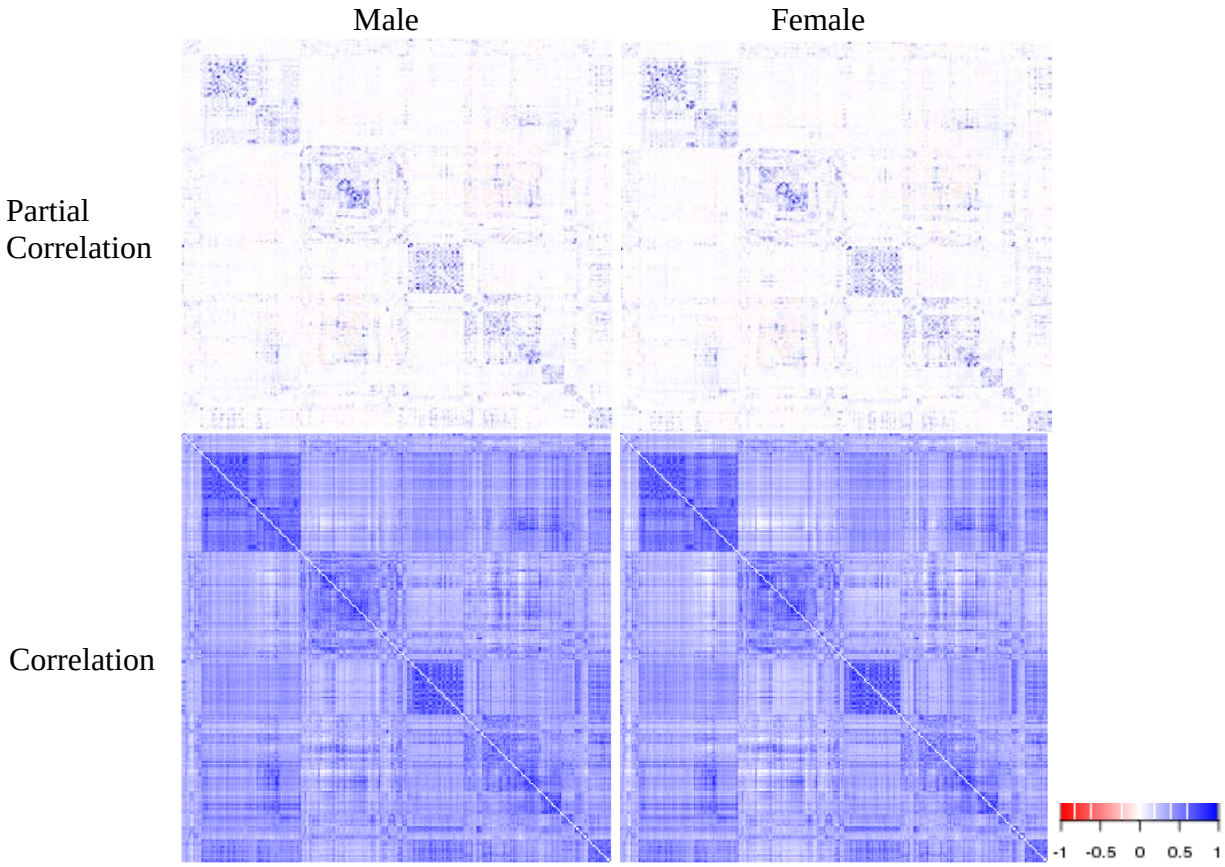

Fig 3: Averaged Full & Partial Correlation Matrix on Male and Female

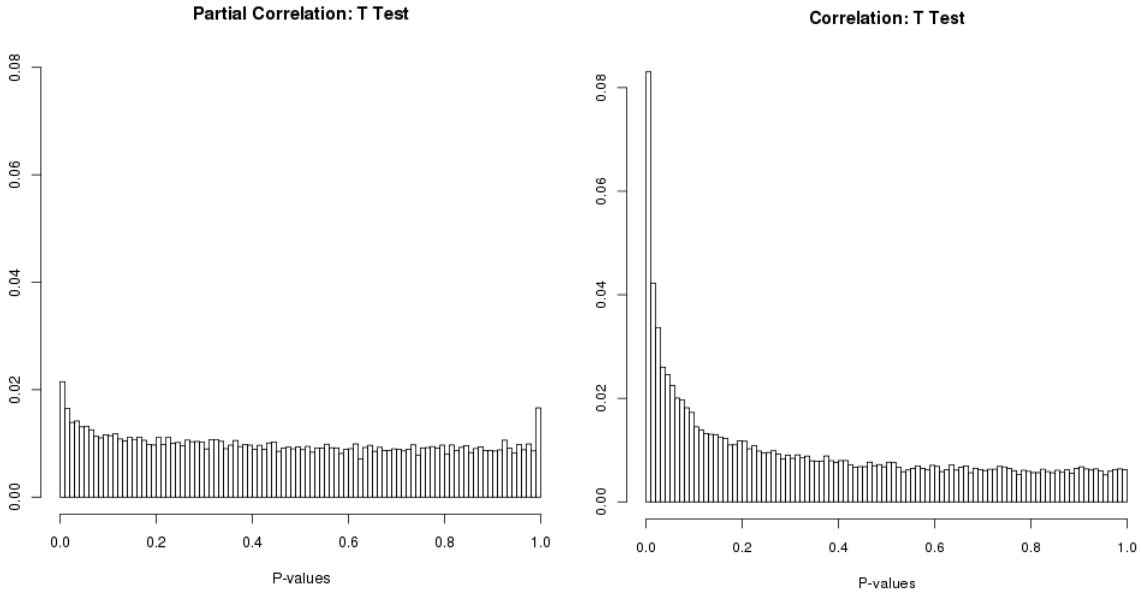

Fig 4: P-value List for Male vs. Female

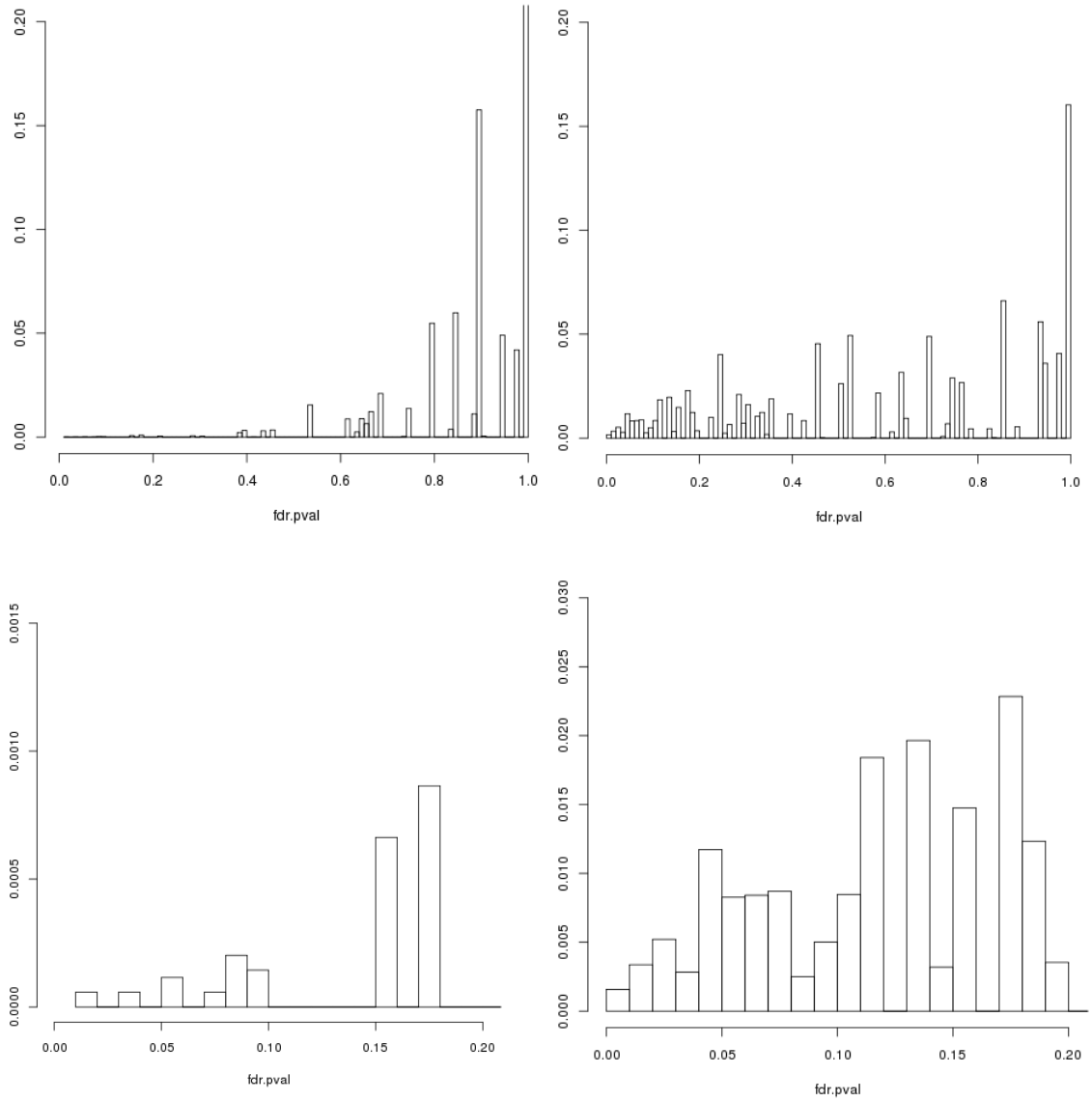

Fig 5: Edge-wise FDR for Female vs. Male (left is partial correlation, right is full correlation; first line is overall histogram of FDR, second row is the histogram of significant edge Q < 0.2)

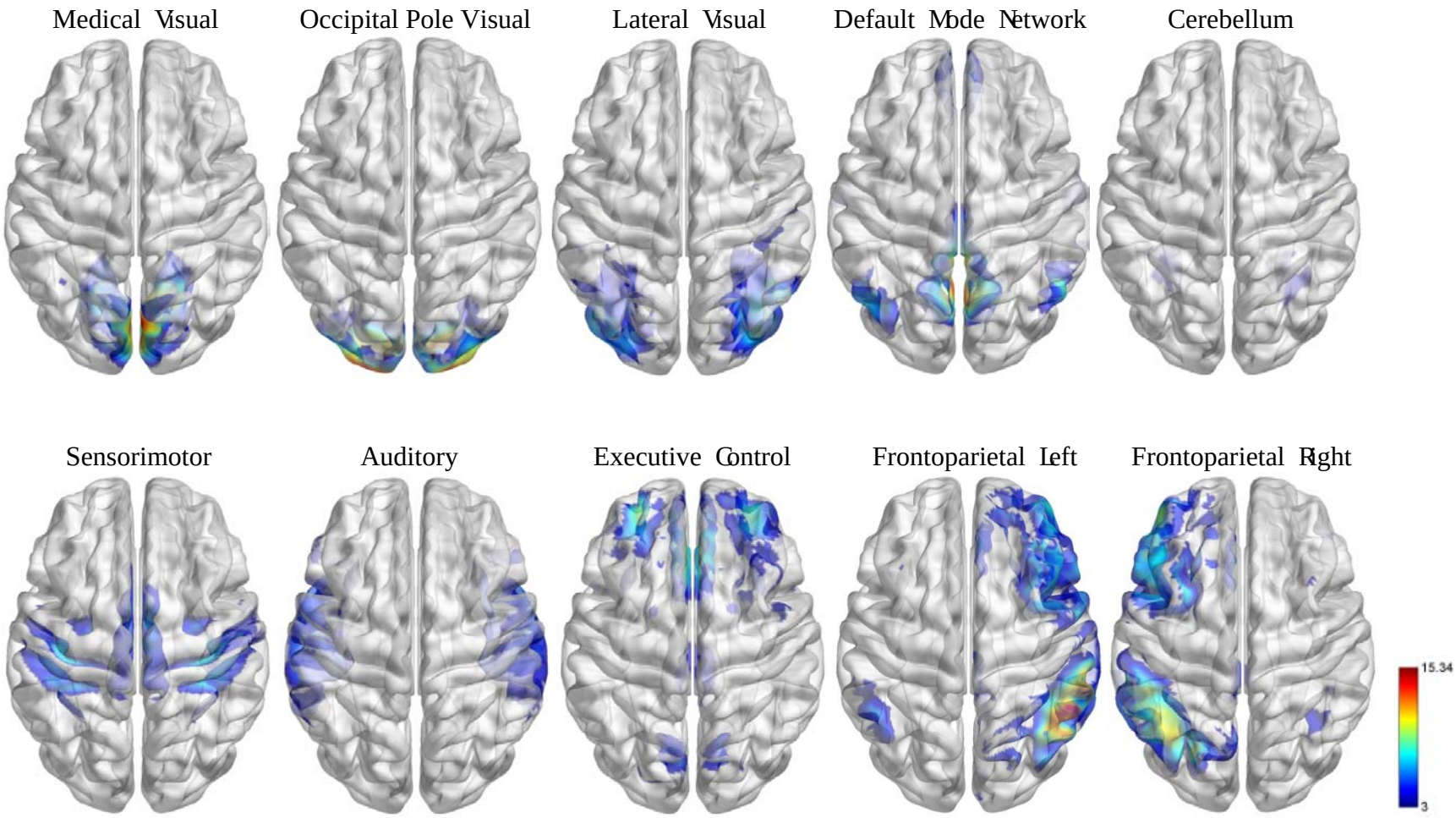

Fig 6.0: Spatial Location of 10 Functionl Modules

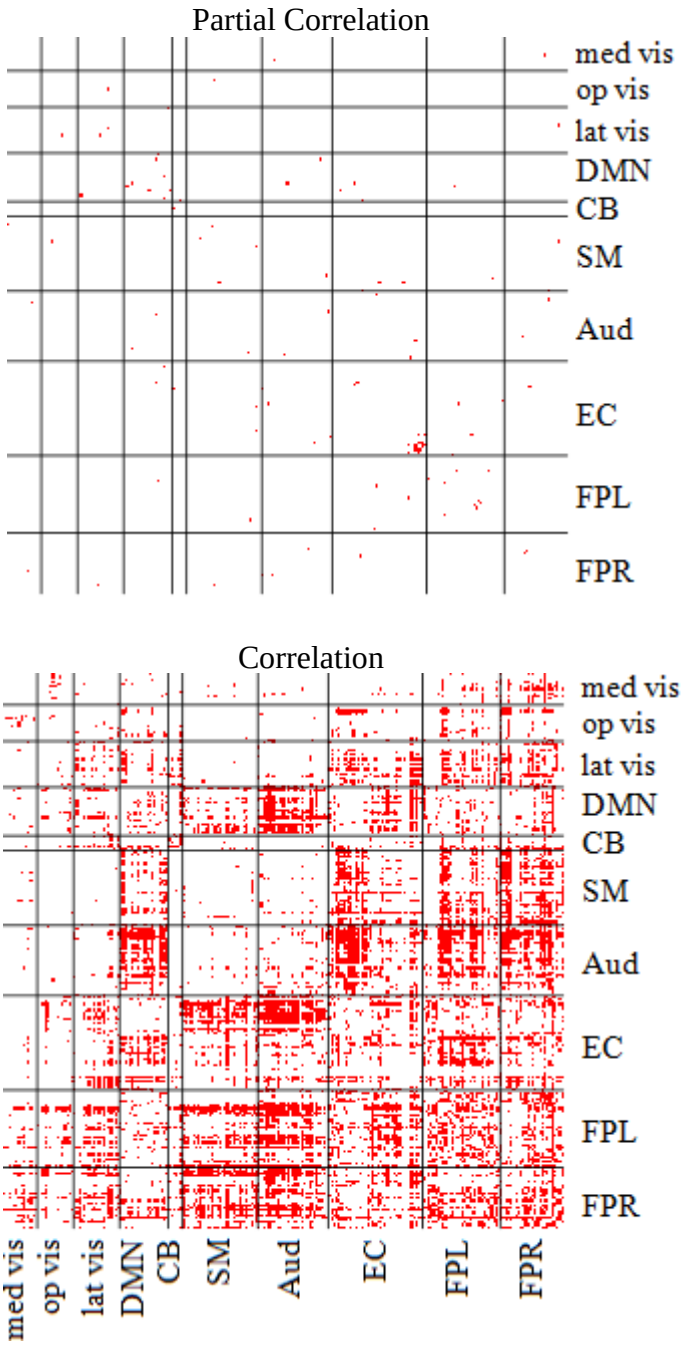

Fig 6.1: Significant Differential Edges between Male and Female (Q < 0.2)

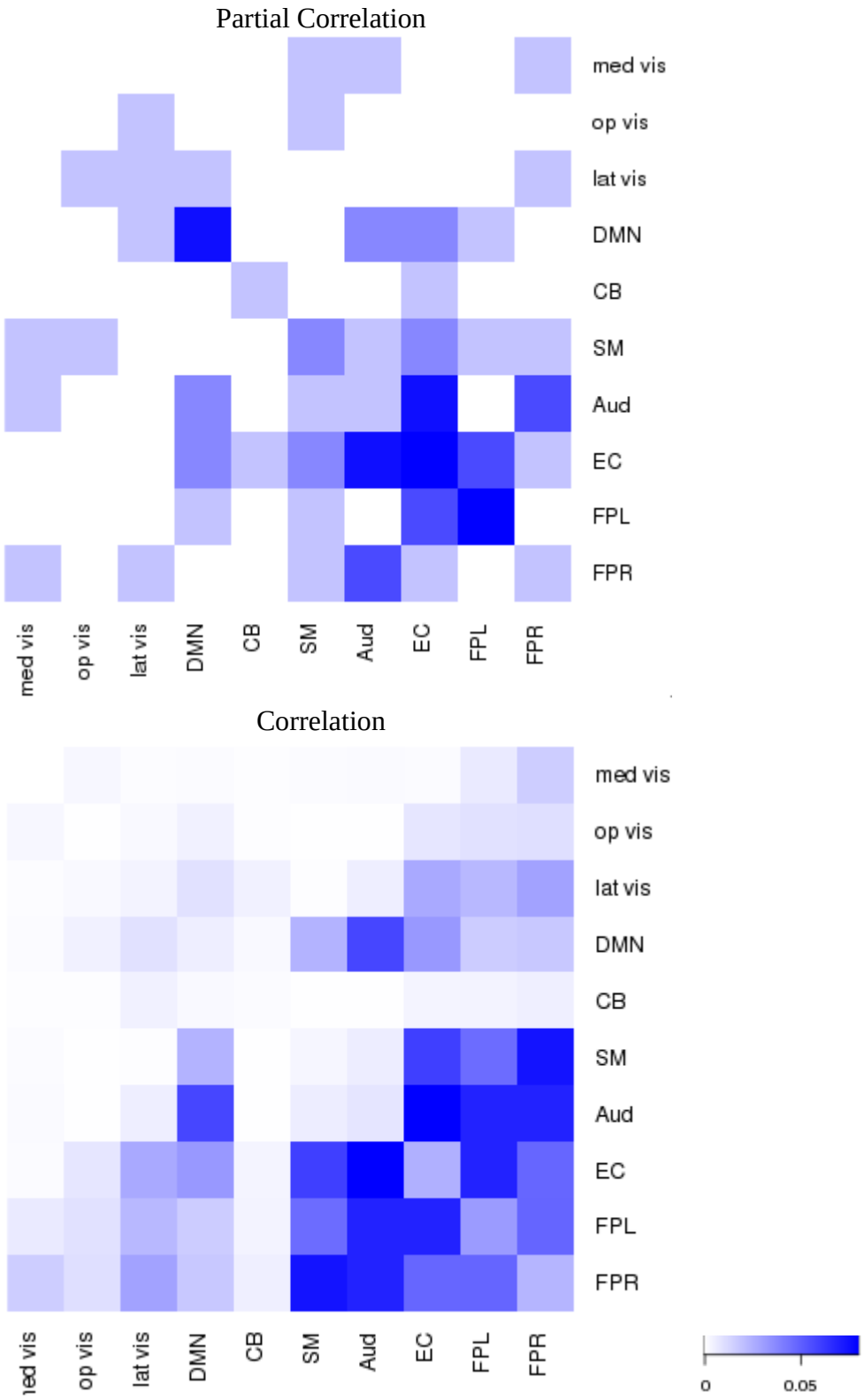

Fig 6.2 : Differential Modules between Male and Female

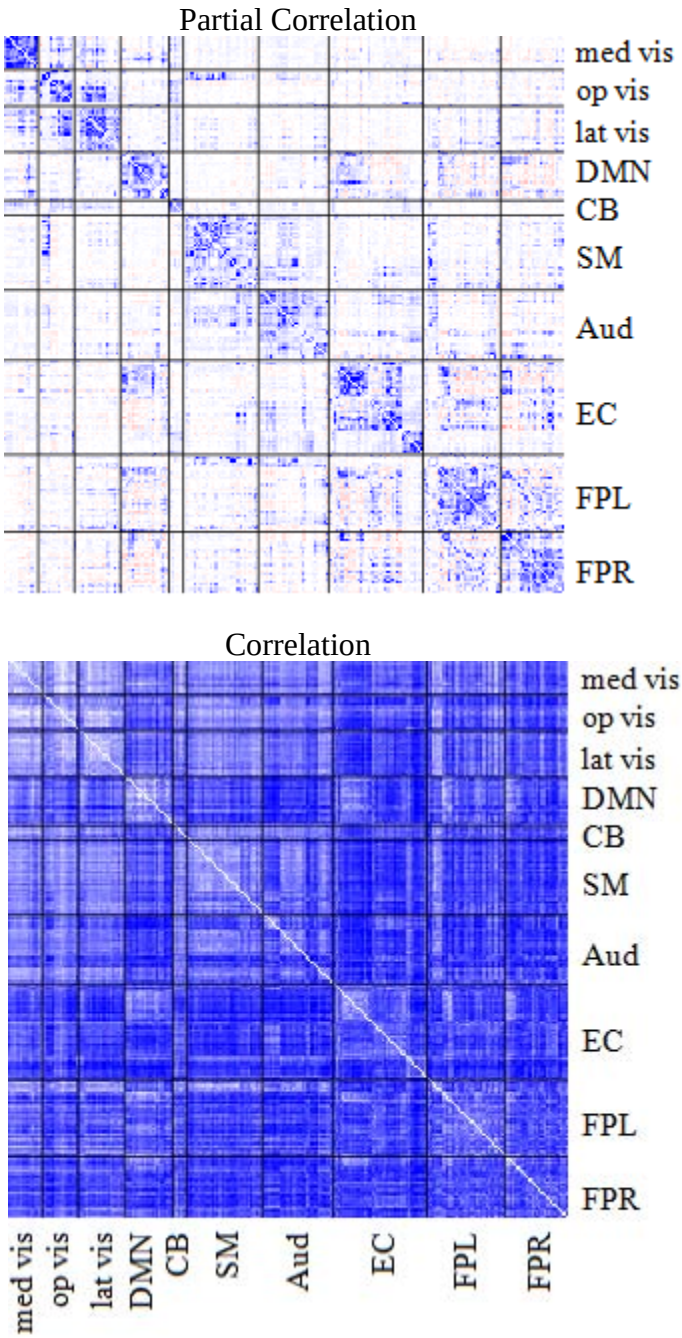

Fig 7.1: SVM Edges Weight between Male and Female

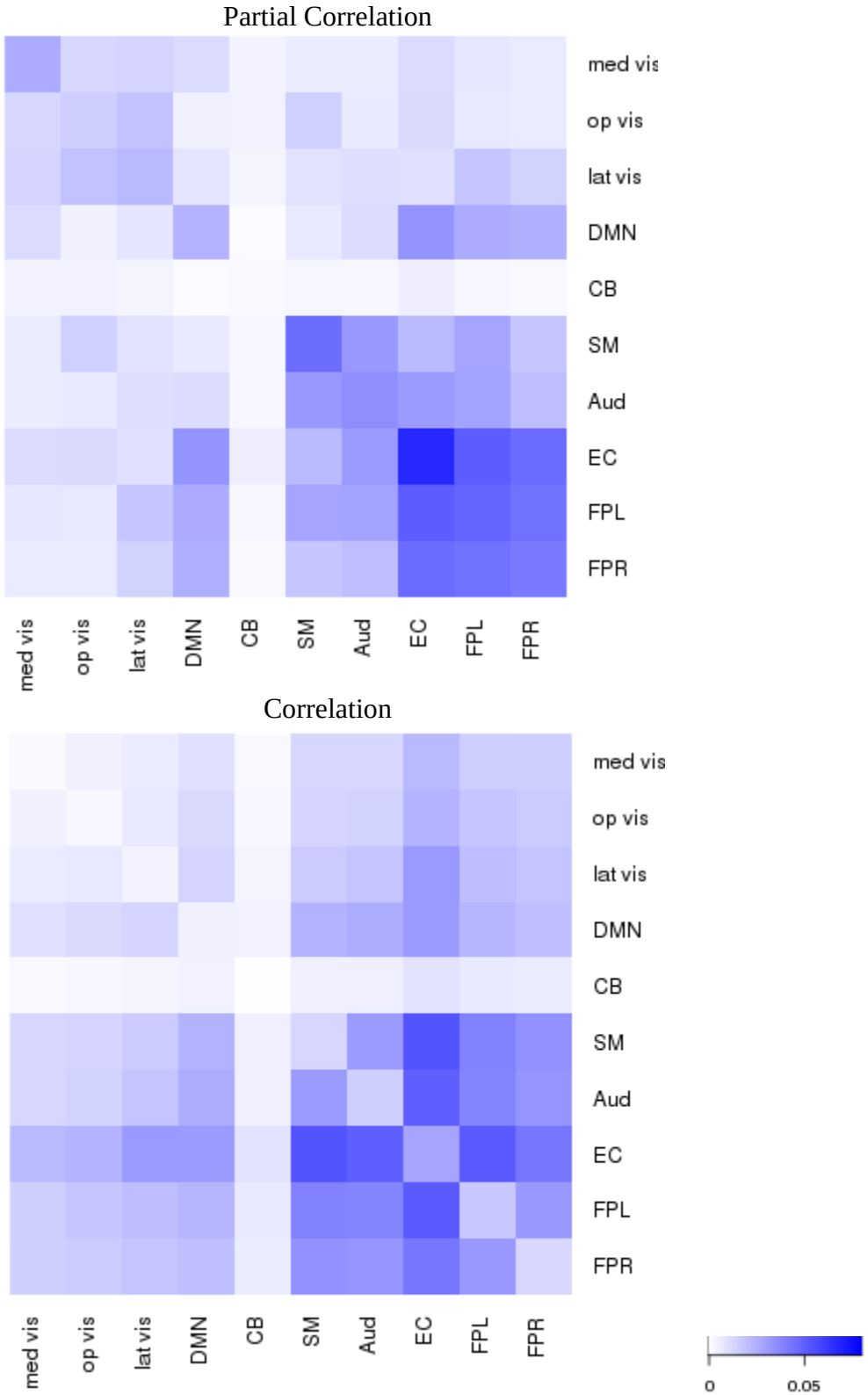

Fig 7.2: Connection between Modules based on SVM Edge Weight

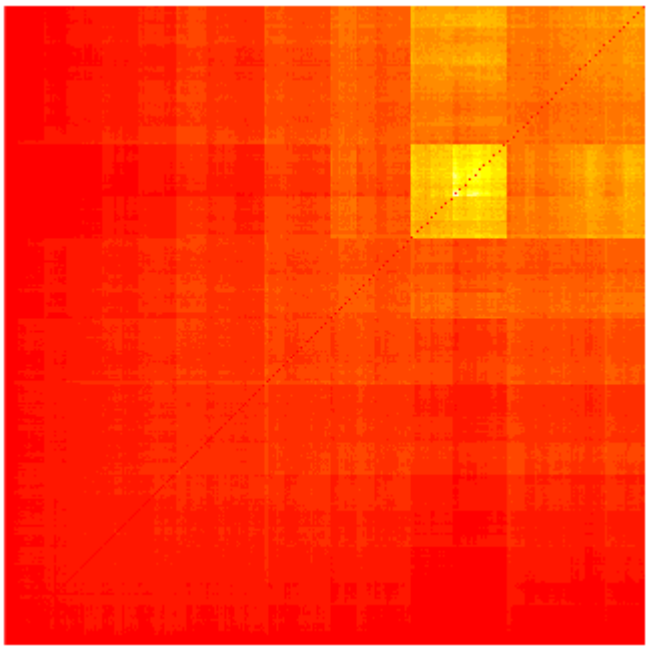

Fig 8: Similarity Matrix of SK kernel

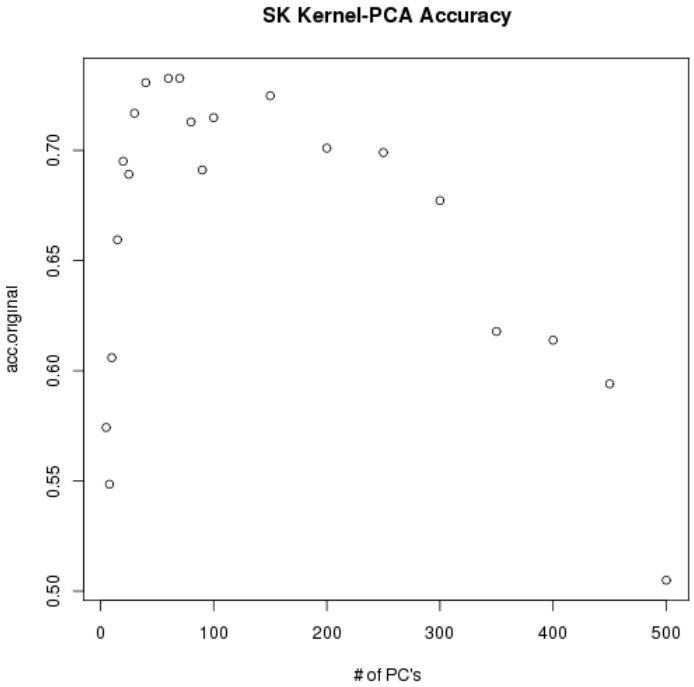

Fig 9: SK-PCA Accuracy (leave-one-out)

Table 1 Classification Accuracy Comparison Among Different Methods

Transformation Method	Accuracy	# of Predictors
Vectorized Precision Matrix	69.70%	34980
Linear PCA	64.55%	150
SPD-kernel PCA (SK)	73.27%	60