

# Sparse Brain Network Recovery under Compressed Sensing

Hyekyoung Lee, Dong Soo Lee, Hyejin Kang, Boong-Nyun Kim, and Moo K. Chung

**Abstract**—Partial correlation is a useful connectivity measure for brain networks, especially, when it is needed to remove the confounding effects in highly correlated networks. Since it is difficult to estimate the exact partial correlation under the small- $n$  large- $p$  situation, a sparseness constraint is generally introduced. In this paper, we consider the sparse linear regression model with a  $l_1$ -norm penalty, also known as the least absolute shrinkage and selection operator (LASSO), for estimating sparse brain connectivity. LASSO is a well-known decoding algorithm in the compressed sensing (CS). The CS theory states that LASSO can reconstruct the exact sparse signal even from a small set of noisy measurements. We briefly show that the penalized linear regression for partial correlation estimation is related to CS. It opens a new possibility that the proposed framework can be used for a sparse brain network recovery. As an illustration, we construct sparse brain networks of 97 regions of interest (ROIs) obtained from FDG-PET imaging data for the autism spectrum disorder (ASD) children and the pediatric control (PedCon) subjects. As validation, we check the network reproducibilities by leave-one-out cross validation and compare the clustered structures derived from the brain networks of ASD and PedCon.

**Index Terms**—Brain Connectivity, Compressed Sensing, Partial Correlation, LASSO

## I. INTRODUCTION

The functional and anatomical connectivity of human brain has known to exhibit large and complex network structures with nontrivial topological characteristics [1], [2], [3], [4], [5],

Copyright (c) 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

H. Lee is with Department of Nuclear Medicine, the Department of Brain and Cognitive Sciences, and the Institute of Radiation Medicine, Medical Research Center, Seoul National University, Seoul, Republic of Korea (e-mail: leehk@postech.ac.kr).

D. S. Lee is with the Department of Nuclear Medicine, WCU Department of Molecular Medicine and Biopharmaceutical Sciences, and the Institute of Radiation Medicine, Medical Research Center, Seoul National University, Seoul, Republic of Korea (e-mail: dsl@snu.ac.kr).

H. Kang is with the Department of Nuclear Medicine and the Institute of Radiation Medicine, Medical Research Center, Seoul National University, Seoul, Republic of Korea (e-mail: hkang211@snu.ac.kr).

B.N. Kim is with the Division of Child and Adolescent Psychiatry and the Department of Neuropsychiatry, Seoul National University, Seoul, Republic of Korea (e-mail:kbn1@snu.ac.kr).

M. K. Chung is with the Department of Biostatistics and Medical Informatics, and the Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin, Madison, WI 53705, USA and also with the Department of Brain and Cognitive Sciences Seoul National University, Seoul, Republic of Korea (e-mail: mkchung@wisc.edu).

This work was supported by the WCU Grant from the government of Korea to M.K.C., by grant No. R31-2008-000-10103-0 from the WCU project of the MEST and the NRF, and by a grant (M103KV010016-08K2201-01610) from Brain Research Center of the 21st Century Frontier Research Program funded by the Ministry of Science and Technology.

[6], [7], [8], [9], [10], [11], [12]. By incorporating the graph theoretical approaches into connectivity analysis, we can gain a new understanding of the characteristics of human brain, from a microscale connectivity between single neurons to a macroscale connectivity between regions of interest (ROIs) in brain images. The brain connectivity has been usually categorized into well-known complex networks such as small-world [1], [3], [4], [5], [7], [11], scale-free [2], [5] or modular networks [8], [9], [10]. The human brain networks is formed from connectivity matrices defined between neuronal elements (single neurons for microscale and ROIs for macroscale network modeling). They are also known as ‘human connectome’ [13].

The majority of previous brain network studies have been based on thresholding correlation in localizing the focal regions of high connectivity [14], [15], [2]. The correlation is used as a similarity measure of network connectivity between two regions. However, the main limitation of correlation-based connectivity analyses is that they fail to explicitly factor out the confounding effect of other regions. To remedy this shortcoming, partial correlation has been naturally introduced in factoring out the dependency of other regions [16], [7], [6] or eliminating the effect of experimental designs [17]. Unfortunately, this type of problem usually belongs to the small- $n$  large- $p$  setting, where the number of regions  $p$  are substantially larger than the number of samples  $n$ , so it is not feasible to estimate the partial correlation accurately [18], [19]. So far the majority of literature have used the penalized likelihood method in imposing the sparseness on the partial correlation estimation [20], [21], [22], [23], [24], [25]. Moreover, since the brain networks are known to be sparse and highly clustered [26], [7], it is reasonable to incorporate the sparsity of network in estimating partial correlation. In this paper, we introduce a different approach based on the penalized linear regression for estimating sparse partial correlation [27], [28]. The penalized linear regression with  $l_1$ -norm, which is also known as the basis pursuit denoising in signal processing and least absolute shrinkage and selection operator (LASSO) in statistics, is usually formulated as the convex optimization to find the sparsest solution of the under-determined linear regression problem [29], [30].

LASSO is one of preferred decoding algorithms in the compressed sensing (CS) theory [31], [32], [33], [34]. The CS-theory states that if LASSO satisfies sparsity and incoherence, the exact recovery is guaranteed with the overwhelmingly high probability, even though the measurement data is not sufficient and contaminated with noise [35], [36]. Note that, incoherence is a stronger condition than a uniform uncertainty

principle (UUP) [37]. If our penalized linear regression for the partial correlation estimation satisfies sparseness and UUP, CS may provide a natural framework for modeling sparse brain networks, which has not been attempted before. We show that the proposed framework satisfies UUP based on the fact that the brain imaging data satisfies UUP for Gaussian ensemble.

In this paper, we focus on sparse model building of the macro-scale connectivity of human brain under CS. The proposed model is applied to the 97 ROIs extracted from FDG-PET data for autism spectrum disorder (ASD) children and pediatric control (PedCon) subjects. It is generally known that ASD has the global underconnectivity and the local overconnectivity in the key brain regions [38], [39]. The differences between ASD and PedCon are mostly found in connectivities between lobes, especially, connection between secondary association cortices such as frontal and parietal regions [40], [41], [5]. Dense internal and sparse external linkages are properties of a module. In particular, some studies suggest that the small-world network, which is one of famous characteristics of brain connectivity, induces a modular architecture [8], [9], [10]. Therefore, in this study, after estimating the partial correlation by the penalized linear regression, we seek the possible modular structures of ASD and PedCon brain network and observe their differences based on the lobe structures.

The main contributions of this paper are:

- to formulate the sparse brain connectivity based on correlation and partial correlation in the penalized linear regression framework,
- to show that the penalized linear regression for partial correlation estimation can near-optimally recover the sparse brain connectivity by showing our study satisfies UUP of the Gaussian ensemble,
- to show the reproducibilities of the estimated networks using the leave-one-out cross-validation,
- to show that controlling sparsity is related with determining the threshold of partial correlation matrix,
- and to suggest a new graph metric, the number of connected components, for thresholding. The proposed metric reflects the modular structures of brain network.

The organization of the paper is as follows. In Section II, we provide notations that will be used through the paper and present the standard methods for calculating correlation and partial correlation. We formulate the problem of estimating correlation and the partial correlation under a sparsity constraint as the sparse linear regression in Section III. Section IV deal with the implementation for estimating the sparse partial correlation. In Section V, after briefly introducing CS and prove that our LASSO-based connectivity method satisfies UUP. Numerical experiments are given in Section VI, where we use the 97 ROIs extracted from FDG-PET data for 26 autistic and 11 pediatric control subjects. We show that the proposed method consistently finds the brain networks which characterize the two groups and have significant group differences in network connectivity.

## II. NETWORK CONSTRUCTION

### A. Connectivity Matrix

Suppose that  $\{\mathbf{f}_1, \dots, \mathbf{f}_p\}$  is the  $n$ -dimensional data vector measured at the  $p$  selected ROIs on the FDG-PET images of  $n$  subjects. The observed data vector  $\mathbf{f}_i$  is the realization of random variable  $f_i$  at the  $i$ -th ROI. The collection of measurements  $f_i$  are assumed to be normally distributed with mean  $\mathbf{0} \in \mathbb{R}^p$  and covariance  $\Sigma = [\sigma_{ij}] \in \mathbb{R}^{p \times p}$ . We will further assume that  $\mathbf{f}_i$  is centered and normalized, i.e.  $\mathbf{f}_i^\top \mathbf{f}_i = 1$ . The covariance  $\sigma_{ij}$  is then estimated as  $\sigma_{ij} = \mathbf{f}_i^\top \mathbf{f}_j$ . If there is no ambiguity, we will interchangeably use  $\sigma_{ij}$  and its estimation  $\widehat{\sigma}_{ij}$ . Given the inverse covariance matrix  $\Sigma^{-1} = (\pi_{ij})$ , which is also known as the precision matrix or concentration matrix, the correlation coefficient  $\rho_{ij}$  and the partial correlation  $\theta_{ij}$  are given by

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \text{ and } \theta_{ij} = -\frac{\pi_{ij}}{\sqrt{\pi_{ii}\pi_{jj}}}.$$

The connectivity matrix of network is usually constructed as a function of correlation or partial correlation. The partial correlation is better in finding the true relationship between two nodes than correlation coefficient due to the ability of factoring out the influence of other regions [42].

### B. Small- $n$ large- $p$ problem

Under high-dimension-small-sample-size setting, the covariance matrix is singular and it cannot be inverted to the precision matrix directly. The pseudo-inverse of covariance matrix can be used (denoted as PINV hereafter), but it has low statistical power and lack of consistency [19]. Imposing the sparseness to the precision matrix is the most natural way to find the precision matrix under small- $n$  large- $p$  situations. It reduces the number of significant elements by forcing all other elements to be zero. There are two most widely used methods for estimating sparse partial correlation: (1) the penalized maximum likelihood (referred as PML hereafter) [22], [23], [43] and (2) the penalized linear regression (referred as PLR hereafter) [27]. The penalized linear regression using  $l_1$ -norm penalty (LASSO) is known to have better performance for the model selection and hub identification at the lower computational cost. Peng, et al. proved the consistency of identifying the true network neighborhood for  $n, p \rightarrow \infty$  [27].

## III. SPARSE CONNECTIVITY ESTIMATION

In this section, we present the penalized linear regression framework for estimating sparse correlation and partial correlation.

### A. Linear Regression for Correlations

Both correlation and partial correlation can be obtained by the linear regression.

- Correlation:

$$\mathbf{f}_i = \alpha_{ij} \mathbf{f}_j + \epsilon_i \quad (i = 1, \dots, p). \quad (1)$$

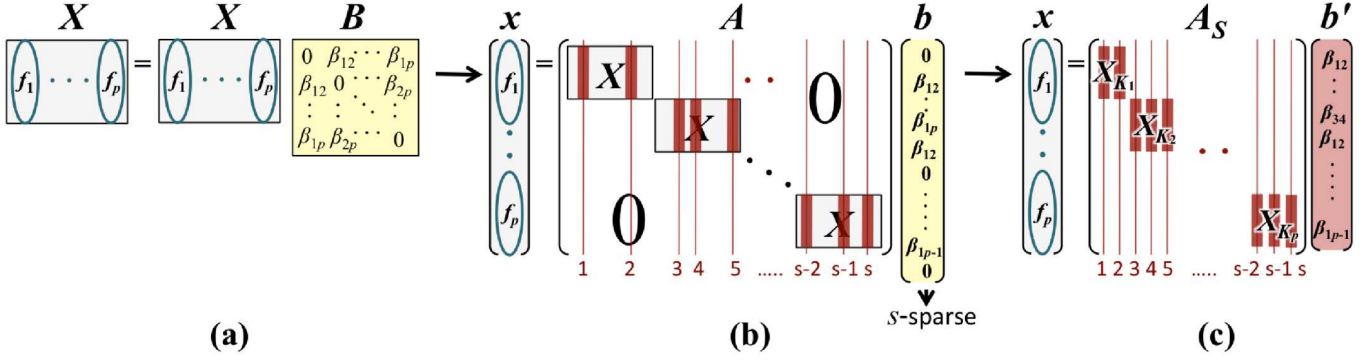


Fig. 1. Linear regression model for the partial correlation estimation. Linear regression model in (2) is represented as (a)  $\mathbf{X} = \mathbf{X}\mathbf{B}$ , where  $\mathbf{X} = [\mathbf{f}_1, \dots, \mathbf{f}_p] \in \mathbb{R}^{n \times p}$  and  $\mathbf{B} = [\beta_{ij}] \in \mathbb{R}^{p \times p}$ .  $\mathbf{B}$  is a symmetric matrix with zero diagonal terms. It can be written as (b)  $\mathbf{x} = \mathbf{A}\mathbf{b}$ , where  $\mathbf{x} = \text{vec}(\mathbf{X})$ ,  $\mathbf{A} = \mathbf{I} \otimes \mathbf{X} \in \mathbb{R}^{np \times p^2}$  and  $\mathbf{b} = \text{vec}(\mathbf{B}) \in \mathbb{R}^{p^2 \times 1}$ .  $\mathbf{I} \in \mathbb{R}^{p \times p}$  is a identity matrix. If  $\mathbf{b}$  is  $s$ -sparse, i.e. it has at most  $s$  number of nonzero elements.

The parameters  $\alpha_{ij}$  are estimated by minimizing the sum of squares

$$\begin{aligned} & \| \mathbf{f}_i - \alpha_{ij} \mathbf{f}_j \|^2 \\ &= (\mathbf{f}_i^\top \mathbf{f}_j) \alpha_{ij}^2 - 2(\mathbf{f}_i^\top \mathbf{f}_j) \alpha_{ij} + (\mathbf{f}_i^\top \mathbf{f}_i). \end{aligned}$$

The minimum is obtained when

$$\alpha_{ij} = \frac{\mathbf{f}_i^\top \mathbf{f}_j}{\mathbf{f}_j^\top \mathbf{f}_j} = \sqrt{\frac{\mathbf{f}_i^\top \mathbf{f}_i}{\mathbf{f}_j^\top \mathbf{f}_j}} \rho_{ij}.$$

For the normalization of the measurement  $\mathbf{f}_i^\top \mathbf{f}_i = 1$ , we have  $\rho_{ij} = \alpha_{ij}$ .

- Partial Correlation:

$$\mathbf{f}_i = \sum_{j \neq i} \beta_{ij} \mathbf{f}_j + \epsilon_i \quad (i = 1, \dots, p), \quad (2)$$

where  $\beta_{ij}$  is the measure of relationship between  $\mathbf{f}_i$  and  $\mathbf{f}_j$  given all other data vectors. When  $\text{var}(\epsilon_i) = (1/\pi_{ii})$  and  $\text{cov}(\epsilon_i, \epsilon_j) = \pi_{ij}/(\pi_{ii}\pi_{jj})$ , the partial correlation  $\theta_{ij}$  is given by [27]

$$\theta_{ij} = \beta_{ij} \sqrt{\frac{\pi_{ii}}{\pi_{jj}}}. \quad (3)$$

Now, we write the linear regression model in (2) to a matrix form (Fig. 1) in order to explicitly show the relationship between partial correlation, linear regression and CS. If we denote  $\mathbf{X} = [\mathbf{f}_1, \dots, \mathbf{f}_p] \in \mathbb{R}^{n \times p}$  and  $\mathbf{B} = [\beta_{ij}] \in \mathbb{R}^{p \times p}$ .  $\mathbf{B}$  is assumed to be symmetric with zero diagonal entries. Then (2) can be rewritten as

$$\mathbf{X} = \mathbf{X}\mathbf{B}. \quad (4)$$

Vectorize the both sides in (4) as

$$\text{vec}(\mathbf{X}) = \text{vec}(\mathbf{X}\mathbf{B}), \quad (5)$$

where  $\text{vec}(\mathbf{X}) = [\mathbf{f}_1^\top \dots \mathbf{f}_p^\top]^\top$  is the vectorization operator. Since  $\text{vec}(\mathbf{X}\mathbf{Y}\mathbf{Z}) = (\mathbf{Z}^\top \otimes \mathbf{X})\text{vec}(\mathbf{Y})$  with the Kronecker product  $\otimes$  [44], we have

$$\text{vec}(\mathbf{X}\mathbf{B}) = (\mathbf{I} \otimes \mathbf{X})\text{vec}(\mathbf{B}),$$

where  $\mathbf{I} \in \mathbb{R}^{p \times p}$  is a identity matrix. Then, (5) can be written in a matrix form

$$\mathbf{x} = \mathbf{A}\mathbf{b}, \quad (6)$$

where  $\mathbf{x} = \text{vec}(\mathbf{X})$ ,  $\mathbf{A} = (\mathbf{I} \otimes \mathbf{X}) \in \mathbb{R}^{np \times p^2}$  and  $\mathbf{b} = \text{vec}(\mathbf{B}) \in \mathbb{R}^{p^2 \times 1}$ .  $\mathbf{A}$  is a block diagonal matrix, but not a square matrix, of which main diagonal blocks consist of  $\mathbf{X}$ s as shown in Fig. 1(b).

Let  $N = np$  and  $P = p^2$ . For  $n \ll p$ , we have  $N \ll P$ . Subsequently, the problem of estimating the partial correlation in the linear model (6) fall under a high-dimension-small-sample-size situation. Thus, we need to incorporate the shrinkage method in regularizing the model parameters by adding the  $l_1$ -norm penalty to the model parameters.

### B. Adding Sparseness Constraint

The solution of linear data model in (6) is usually obtained by the least squares minimization:

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} \| \mathbf{x} - \mathbf{A}\mathbf{b} \|_2^2, \quad (7)$$

where  $\|\cdot\|_2$  is a  $l_2$ -norm. When the linear regression is under a small- $n$  large- $p$  problem, there exist infinitely many solutions. To obtain a unique solution, we need to add the sparseness constraint. The sparsest solution is obtained by the  $l_0$ -norm penalty, which measures the number of non-zero elements, as

$$\min_{\mathbf{b}} \| \mathbf{b} \|_0 \quad \text{subject to } \mathbf{x} = \mathbf{A}\mathbf{b}. \quad (8)$$

Since it is a combinatorial problem with NP-hard complexity, instead of using the  $l_0$ -norm, we employ the  $l_1$ -norm penalty (sum of absolute values of elements):

$$\min_{\mathbf{b}} \| \mathbf{b} \|_1 \quad \text{subject to } \mathbf{x} = \mathbf{A}\mathbf{b}, \quad (9)$$

which is related with the linear programming and the basis pursuit denoising problem [29]. The discussion about the  $l_0$  and  $l_1$  equivalence can be found in [45]. For a noisy case, we can transform (9) to the quadratic programming with a linear constraint

$$\min_{\mathbf{b}} \| \mathbf{x} - \mathbf{A}\mathbf{b} \|_2^2 \quad \text{subject to } \| \mathbf{b} \|_1 < \epsilon. \quad (10)$$

The formulation (10) is known as LASSO [30], [46]. By introducing the sparseness control parameter  $\lambda$ , the formulation (10) is rewritten by

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} \|\mathbf{x} - \mathbf{A}\mathbf{b}\|_2^2 + \lambda \|\mathbf{b}\|_1. \quad (11)$$

The optimization for solving the sparse partial correlation is done by the coordinate descent learning and the active-set algorithm [27], [47]. We can also apply the LASSO framework to (1) in estimating sparse correlation. The algorithms for the sparse correlation and partial correlation are outlined in the next section.

#### IV. ESTIMATING SPARSE (PARTIAL) CORRELATIONS

In this section, we derive an algorithm for estimating the sparse (partial) correlations using the coordinate descent learning and the active-set algorithm [27], [47]. We also propose the method tuning the amount of sparsity  $\lambda$  in (11) based on the topological structure of the networks.

##### A. Algorithm for Sparse (Partial) Correlations

It requires huge computer memory to directly solve the optimization problem for large number of nodes (11). So we consider the LASSO frameworks of the element-wise formulas (1) and (2) in estimating sparse correlation and partial correlation [27]. The sparse correlation is obtained by

$$\hat{\alpha}_{ij} = \arg \min_{\alpha_{ij}} \sum_{i=1}^p \sum_{j \neq i} \|\mathbf{f}_i - \alpha_{ij} \mathbf{f}_j\|_2^2 + \lambda \sum_{i,j} |\alpha_{ij}| \quad (12)$$

while the sparse partial correlation is estimated by

$$\hat{\beta}_{ij} = \arg \min_{\beta_{ij}} \sum_{i=1}^p \|\mathbf{f}_i - \sum_{j \neq i} \beta_{ij} \mathbf{f}_j\|_2^2 + \lambda \sum_{i,j} |\beta_{ij}|. \quad (13)$$

The coordinate descent learning and the active-set algorithm are used to solve the optimization problems (12) and (13) [27], [47].

The objective function for estimating correlation without  $l_1$ -norm-penalty is given by

$$\mathcal{F} = \sum_{i=1}^p \sum_{j \neq i} \|\mathbf{f}_i - \alpha_{ij} \mathbf{f}_j\|_2^2.$$

Since  $\mathcal{F}$  is a convex function, the minimum is achieved at  $\hat{\alpha}_{ij}$  when

$$\frac{\partial \mathcal{F}}{\partial \alpha_{ij}} = -\mathbf{f}_i^\top \mathbf{f}_j + \alpha_{ij} = 0 \quad \Rightarrow \quad \hat{\alpha}_{ij} = \mathbf{f}_i^\top \mathbf{f}_j.$$

If we add the  $l_1$ -norm penalty, we have a new objective function

$$\mathcal{F}_1 = \sum_{i=1}^p \sum_{j \neq i} \|\mathbf{f}_i - \alpha_{ij} \mathbf{f}_j\|_2^2 + \lambda \sum_{i,j} |\alpha_{ij}|.$$

The derivatives are given by

$$\frac{\partial \mathcal{F}_1}{\partial \alpha_{ij}} = \begin{cases} -\mathbf{f}_i^\top \mathbf{f}_j + \alpha_{ij} + \lambda, & \text{for } \alpha_{ij} > 0 \\ -\mathbf{f}_i^\top \mathbf{f}_j + \alpha_{ij} - \lambda, & \text{for } \alpha_{ij} < 0 \end{cases}.$$

Hence, the minimum is obtained when

$$\hat{\alpha}_{ij} = \left[ \mathbf{f}_i^\top \mathbf{f}_j, \lambda \right]_+, \quad (14)$$

where

$$[a, b]_+ = \begin{cases} a - b & \text{if } a > 0 \text{ and } |a| > b \\ a + b & \text{if } a < 0 \text{ and } |a| > b \\ 0 & \text{if } |a| \leq b \end{cases}.$$

The partial correlation is also estimated in the same way. The objective function for sparse partial correlation is given by

$$\mathcal{G}_1 = \sum_{i=1}^p \|\mathbf{f}_i - \sum_{j \neq i} \theta_{ij} \sqrt{\frac{\pi_{jj}}{\pi_{ii}}} \mathbf{f}_j\|_2^2 + \lambda \sum_{i,j} |\theta_{ij}| \sqrt{\frac{\pi_{jj}}{\pi_{ii}}}.$$

Note that  $\pi_{ii}$  can be estimated directly from the sample variance the measurements. So assuming  $\pi_{ii}$  is given, the minimum is obtained when

$$\frac{\partial \mathcal{G}_1}{\partial \theta_{ij}} = \begin{cases} A\theta_{ij} - \mathcal{B}(\theta_{ik, k \neq i, j}) + \lambda, & \text{for } \theta_{ij} > 0 \\ A\theta_{ij} - \mathcal{B}(\theta_{ik, k \neq i, j}) - \lambda, & \text{for } \theta_{ij} < 0 \end{cases}$$

where

$$A = \frac{\pi_{jj}}{n\pi_{ii}} \mathbf{f}_j^\top \mathbf{f}_j$$

$$\mathcal{B}(\theta_{ik, k \neq i, j}) = \frac{1}{n} \left( \mathbf{f}_i^\top \mathbf{f}_j - \sum_{k \neq i, j} \theta_{ik} \frac{\sqrt{\pi_{kk}\pi_{jj}}}{\pi_{ii}} \mathbf{f}_k^\top \mathbf{f}_j \right).$$

$\mathcal{B}$  is a function of  $\theta_{ik}$  for all  $k = \{1, \dots, p\} \setminus \{i, j\}$ . Hence the partial correlation is estimated as

$$\hat{\theta}_{ij} = \frac{[\mathcal{B}(\theta_{ik, k \neq i, j}), \lambda]_+}{A}. \quad (15)$$

While the correlation estimation (14) is in a closed-form, the partial correlation estimation (15) is a function of all other partial correlation coefficients. In the coordinate descent optimization [48], we can obtain the global minimum by sequentially minimizing with respect to  $\theta_{ij}$  while fixing all other partial correlations. The algorithm for the coordinate descent optimization is given below:

<b>Input :</b>	$\mathbf{X} = [\mathbf{f}_1, \dots, \mathbf{f}_p] \in \mathbb{R}^{n \times p}, \lambda, \delta$
<b>Output :</b>	$\boldsymbol{\theta} \in \mathbb{R}^{p \times p}, \boldsymbol{\pi} \in \mathbb{R}^p$
1	Normalize $\mathbf{X}$ such that $\mathbf{f}_i^\top \mathbf{1} = 0$ and $\mathbf{f}_i^\top \mathbf{f}_i = 1$ for all $i$ ( $\mathbf{1} = [1, \dots, 1]^\top \in \mathbb{R}^n$ ).
2	Initialize $[\theta_{ij}]_{i,j=1, \dots, p(i < j)}$ and $[\pi_{ii}]_{i=1, \dots, p}$ .
3	While $l$
4	While $m$
5	For $i = 1, \dots, p$ and $j = i + 1, \dots, p$ ,
	$\theta_{ij}^{(m)} \leftarrow \frac{[\mathcal{B}(\theta_{ik, k \neq i, j})^{(m-1)}, \lambda]_+}{A}$ .
6	Repeat 4 until $ \theta_{ij}^{(m)} - \theta_{ij}^{(m-1)}  < \delta$ .
7	$\beta_{ij}^{(l)} \leftarrow \theta_{ij}^{(l)} \sqrt{\frac{\pi_{jj}^{(l-1)}}{\pi_{ii}^{(l-1)}}}$ for all $i, j$ .
8	For $i = 1, \dots, p$ ,
	$\pi_{ii}^{(l)} \leftarrow 1/\text{var}(\mathbf{f}_i - \sum_{j \neq i} \beta_{ij}^{(l)} \mathbf{f}_j)$ .
9	Repeat 3 until $ \pi_{ii}^{(l)} - \pi_{ii}^{(l-1)}  < \delta$ .

For the convergence of the algorithm, we used  $\delta = 10^{-3}$  and initialized  $[\theta_{ij}]_{i,j=1, \dots, p(i < j)}$  and  $[\pi_{ii}]_{i=1, \dots, p}$  using uniformly generated values.

Since we assume that most of the partial correlations  $\theta_{ij}$  are zero, if we update only nonzero partial correlation, we can reduce the computational complexity. Using this idea, Friedman and Peng exploited the active set algorithm [27], [47] which defines the nonzero partial correlation as the active set and updates only  $\theta_{ij}$  belonging to the active set. In the active set algorithm, the steps from 4 to 6 are then changed as follows:

4\* While  $m$   
 Construct the current active set  
 $\Lambda = \{(i, j) \mid \text{current } \theta_{ij}^{(m-1)} \neq 0\}$ .

5\* While  $m'$   
 For all elements on the active set,  
 Update  $\theta_{ij}^{(m')}$  ( $(i, j) \in \Lambda$  in (15)).  
 Repeat 5\* until convergence on the active set.

5\*\* For  $i = 1, \dots, p$  and  $j = i + 1, \dots, p$ ,  
 Update  $\theta_{ij}^{(m)}$  ( $(i, j) \in \Lambda$  in (15)).

6\* Repeat Step 4\* until convergence.

### B. Controlling Sparsity

For practical purposes, it is necessary to determine the amount of sparsity  $\lambda$ .

1) *Thresholding and Sparseness*: Estimating the partial correlation by fixing  $\lambda$  is equivalent to thresholding the partial correlations. This is evident from (15), where the partial correlation is given as a function of  $\lambda$ . Thus, we can apply the traditional thresholding methods for finding the suitable sparse network instead of controlling the sparsity. However, if the sparsity is too small in the small- $n$  large- $p$  situation, it is impossible to estimate the exact partial correlations. In that case, we cannot directly compare the sparse partial correlations obtained from CS and the partial correlations obtained from the usual thresholding.

For selecting the sparsity parameter  $\lambda$ , we investigate the topological structures of the networks and check its stability during the cross validation as  $\lambda$  increases. We estimate the stability of the topological structures in two different ways. First, the sample variance of the total number of edges in a network is considered during the cross validation. The sparsity which have the small sample variance is chosen. Second, we estimate the number of edges which are included in the previous network with smaller sparsity among all edges in the network, i.e.,  $|\mathcal{E}_{\lambda_t} \cap \mathcal{E}_{\lambda_{t-1}}|$  for  $\lambda_{t-1} < \lambda_t$ , where  $|\mathcal{E}_{\lambda}|$  is the number of elements in a set of edges  $\mathcal{E}_{\lambda}$  in the partial correlation network with the sparsity  $\lambda$ . If  $|\mathcal{E}_{\lambda_t} \cap \mathcal{E}_{\lambda_{t-1}}| = |\mathcal{E}_{\lambda_t}|$ , i.e.,  $\mathcal{E}_{\lambda_t} \subset \mathcal{E}_{\lambda_{t-1}}$ , then, the sparsity  $\lambda_t$  is chosen. Because the thresholded network has the property that a set of edges in the network is a subset of a set of that with smaller threshold. If the obtained sparse network satisfies the same property of thresholded network, we can consider that it is the optimal solution of partial correlation.

2) *Clustered Structure via Thresholding*: The network obtained from the partial correlation with the chosen  $\lambda$  in the previous section still have many edges. Thus, we threshold it once again to find more suitable sparse network reflecting the characteristics of our dataset, ASD and PedCon.

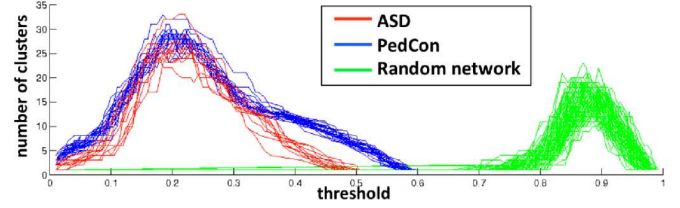


Fig. 2. The plots of the number of clusters for changing threshold on correlations. The cluster is defined as the connected components with more than two nodes.

The well-known characteristics of ASD is local overconnectivity and global underconnectivity. The overconnectivity is characterized by more number of edges within the lobes while the underconnectivity is characterized by less number of edges between the lobes. The local overconnectivity and global underconnectivity are related to the modular network which has many number of edges within a module and small number of edges between modules. In addition, the autistic and normal control subjects show significant group differences in connectivity between lobes and within lobes [40], [38], [39], [49]. Hence, we hypothesized that (1) finding the modular structure in the brain network can be used in differentiating ASD and PedCon and (2) the modular structure of ASD network shows the abnormal connectivity patterns within and between lobes.

A cluster is defined to be a connected component that has more than two nodes connected with edges. We consider that the number of clusters is a graph metric reflecting the modular structure of brain network. The modular structure generally allows the edges between disjoint modules, while the clustered structure does not allow such edges. In this sense, modular and clustered structures are different. However, in this paper, we will simply treat them equivalent to simply the problem and consider that a cluster reflects the modular structure of brain network. When the threshold is sufficiently small, all nodes are connected and the number of clusters becomes one. The number of clusters increases when the threshold increases but at a certain threshold, it obtains the maximum (Fig. 2). When the threshold is large, all nodes are disconnected and the number of connected components becomes the number of nodes, but the number of clusters becomes zero. We choose the threshold corresponding to the maximum as the representation of the network in subsequent analysis.

## V. RELATIONSHIP TO COMPRESSED SENSING

In this section, we show that the sparse linear model for estimating partial correlation is related to UUP. It opens a possibility that the near-optimal recovery of sparse brain network can be done by the proposed method.

### A. Uniform Uncertainty Principle

The coherence of  $\mathbf{A}$ ,  $\mu(\mathbf{A})$ , is defined as the maximum correlation coefficient among all correlations between two different column vectors in (6). If all basis vectors are orthogonal, the coherence is minimized. A signal is  $s$ -sparse if there

exists at most  $s$  number of nonzero elements in the signal. The exact sparse signal recovery is guaranteed for  $s$ -sparse signal, if the coherence is bounded by  $\mu(\mathbf{A}) \leq \mathcal{O}(1/s)$  [50]. However, since  $N \ll P$ , this condition is difficult to satisfy. A looser condition, which guarantees the near-optimal sparse data recovery is a UUP [35], [51].

*Definition 1:* A measurement matrix  $\mathbf{A}$  satisfies the *uniform uncertainty principle* (UUP) with the oversampling factor  $\lambda$  if, for every sufficiently small  $\gamma > 0$  and any  $s$ -sparse vector  $\mathbf{b}$  such that

$$s \leq \gamma \cdot N/\lambda,$$

$\mathbf{A}$  holds inequalities

$$\frac{1}{2} \cdot \frac{N}{P} \cdot \|\mathbf{b}\|_2^2 \leq \|\mathbf{A}\mathbf{b}\|_2^2 \leq \frac{3}{2} \cdot \frac{N}{P} \cdot \|\mathbf{b}\|_2^2, \quad (16)$$

with probability at least  $1 - \mathcal{O}(p^{-\rho/\gamma})$  for some fixed constant  $\rho > 0$ .

Let  $S$  be the index set consisting of  $s$  numbers from between 1 and  $P$ . Let  $\mathbf{A}_S \in \mathbb{R}^{N \times s}$  be a submatrix consisting of  $s$  column vectors of  $\mathbf{A}$  (Fig. 1(c)). The columns are obtained from the index set  $S$ . Then, it can be shown that the condition (16) is equivalent to

$$\frac{1}{2} \cdot \frac{N}{P} \leq \lambda_{\min}(\mathbf{A}_S^\top \mathbf{A}_S) \leq \lambda_{\max}(\mathbf{A}_S^\top \mathbf{A}_S) \leq \frac{3}{2} \cdot \frac{N}{P}, \quad (17)$$

for all subsets  $S$ .  $\lambda_{\max}(\cdot)$  and  $\lambda_{\min}(\cdot)$  denote the largest and smallest eigenvalues.

*Lemma 1:* The Gaussian ensemble  $\mathbf{X} \in \mathbb{R}^{n \times p}$ , which are i.i.d.  $\mathcal{N}(0, 1/p)$ , holds the UUP with the oversampling factor  $\lambda = \log p$ .

Lemma 1 implies that, if we pick  $\mathbf{A}$  in the sparse linear model  $\mathbf{x} = \mathbf{A}\mathbf{b}$  as a Gaussian ensemble  $\mathbf{X}$  with  $n \geq \gamma \cdot s \log p$ , sparse recovery of  $\mathbf{b}$  can be done with overwhelmingly large probability [35].

Now, we will show that the brain connectivity obtained by the penalized linear regression in (6) can be recovered under UUP for the first time. It is sufficient to show that the Gaussianness of data matrix  $\mathbf{X}$  in (6).

### B. Gaussianness of Data Matrix

After centering and normalizing the column vectors, we check the Gaussianness of the data matrix  $\mathbf{X} = [\mathbf{f}_1, \dots, \mathbf{f}_p] \in \mathbb{R}^{n \times p}$  using both the Lilliefors test and quantile-quantile plots (QQ-plots) [52]. Because it is not possible to visualize the QQ-plots for all ROIs here, we measured the correlation coefficients  $r$  of scatter points in the QQ-plots for quantification. If the empirical distribution follows Gaussian,  $r$  should asymptotically converge to 1. For the random numbers from  $\mathcal{N}(0, 1/p)$  whose dimension is identical to the data matrix of ASD and PedCon,  $r = 0.98 \pm 0.01$  and  $r = 0.97 \pm 0.03$ , respectively. For the actual measurements from ASD and PedCon,  $r = 0.97 \pm 0.02$  and  $r = 0.96 \pm 0.03$  (Fig. 3). The high correlation in the QQ-plot guarantees that our data matrix follows a normal distribution. Using Lilliefors statistic, we also tested Gaussianness. Since the Lilliefors statistics of data matrix are mostly smaller than the cutoff values of 0.19 (ASD) and 0.29 (PedCon) at 1% level, there is no

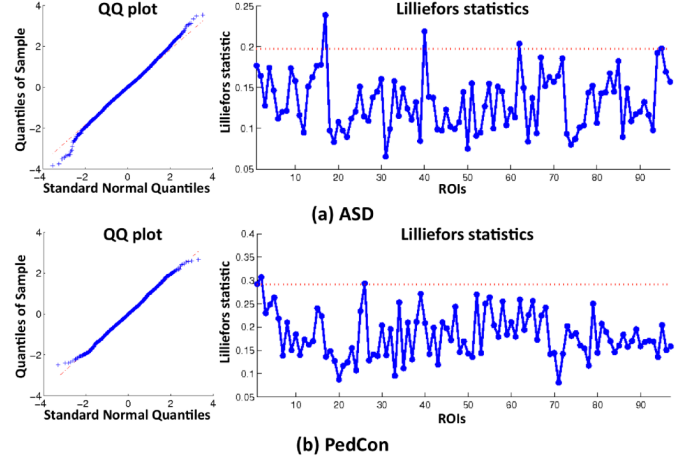


Fig. 3. Checking Gaussianness of data matrix for (a) ASD and (b) PedCon. In the QQ-plots, the horizontal axes are the quantiles of a normal distribution while the vertical axes are the quantiles of data. If the data are Gaussian, the QQ-plot should be close to the straight red line. In the right panels, the horizontal axes display the index of ROIs while the vertical axes displays Lilliefors statistic which measures the maximum difference between empirical and theoretical Gaussian distributions. Most ROIs (blue solid lines) are below the cutoff values 0.19 for ASD (a) and 0.29 for PedCon (b) at the  $\alpha$  level of 1%. The two test procedures confirm the Gaussianness of our data.

reason not to assume normality for the given data matrix, i.e.,  $X_{ij} \sim \mathcal{N}(0, 1/p)$  (Fig. 3). Since FDG-PET measurements are obtained in 97 nodes that are not close to each other, the elements of  $\mathbf{X}$  are likely to be i.i.d. so we did not check the i.i.d. assumption here.

### C. Sparse Brain Connectivity Recovery under CS

If the measurement matrix  $\mathbf{X}$  follows i.i.d. Gaussian, from Lemma 1,  $\mathbf{X}$  satisfies UUP. In other words, for any submatrix  $\mathbf{X}_{K'}$  of  $\mathbf{X}$  with  $K' \subset K$ , it satisfies the condition (17) like

$$\begin{aligned} \frac{1}{2} \cdot \frac{n}{p} &\leq \lambda_{\min}(\mathbf{X}_{K'}^\top \mathbf{X}_{K'}) \leq \lambda_{\min}(\mathbf{X}_K^\top \mathbf{X}_{K'}) \\ &\leq \lambda_{\min}(\mathbf{X}_{K'}^\top \mathbf{X}_{K'}) \leq \lambda_{\max}(\mathbf{X}_K^\top \mathbf{X}_K) \leq \frac{3}{2} \cdot \frac{n}{p}. \end{aligned}$$

Let  $\mathbf{A}_S \in \mathbb{R}^{N \times s}$  be a submatrix consisting of  $s$  column vectors of  $\mathbf{A}$  (Fig. 1(c)). To check if a measurement matrix  $\mathbf{A}$  in (11) holds UUP, we should show that all submatrices  $\mathbf{A}_{S'}$  ( $S' \subset S$ ) satisfy (17). Because  $\mathbf{A}_S$  is a block diagonal matrix (not a square matrix) of which block matrices are  $\mathbf{X}_{K_1}, \dots, \mathbf{X}_{K_p}$  with the number of column vectors  $k_1, \dots, k_p$  (Fig. 1(c)), the smallest and largest eigenvalues of  $\mathbf{A}_S^\top \mathbf{A}_S$  are

$$\begin{aligned} \lambda_{\min}(\mathbf{A}_S^\top \mathbf{A}_S) &= \min \left\{ \lambda_{\min}(\mathbf{X}_{K_1}^\top \mathbf{X}_{K_1}), \dots, \lambda_{\min}(\mathbf{X}_{K_p}^\top \mathbf{X}_{K_p}) \right\}, \\ \lambda_{\max}(\mathbf{A}_S^\top \mathbf{A}_S) &= \max \left\{ \lambda_{\max}(\mathbf{X}_{K_1}^\top \mathbf{X}_{K_1}), \dots, \lambda_{\max}(\mathbf{X}_{K_p}^\top \mathbf{X}_{K_p}) \right\}. \end{aligned}$$

If we write the index set  $K = K_1 \cup \dots \cup K_p$ , then,

$$\begin{aligned} \lambda_{\min}(\mathbf{X}_K^\top \mathbf{X}_K) &\leq \lambda_{\min}(\mathbf{A}_S^\top \mathbf{A}_S) \\ &\leq \lambda_{\max}(\mathbf{A}_S^\top \mathbf{A}_S) \leq \lambda_{\max}(\mathbf{X}_K^\top \mathbf{X}_K). \end{aligned}$$

For all subsets  $S' \subset S$ , the block matrices of  $\mathbf{A}_{S'}$  come from  $\mathbf{X}_{K'}$  ( $K' \subset K$ ). Therefore,

$$\begin{aligned} \frac{1}{2} \cdot \frac{n}{p} &\leq \lambda_{\min}(\mathbf{A}_S^\top \mathbf{A}_S) \leq \lambda_{\min}(\mathbf{A}_{S'}^\top \mathbf{A}_{S'}) \\ &\leq \lambda_{\min}(\mathbf{A}_{S'}^\top \mathbf{A}_{S'}) \leq \lambda_{\max}(\mathbf{A}_S^\top \mathbf{A}_S) \leq \frac{3}{2} \cdot \frac{n}{p}. \end{aligned}$$

The larger measurement matrix  $\mathbf{A}$  will satisfy the condition (17), showing that  $\mathbf{A}$  also satisfies UUP with the same probability as  $\mathbf{X}$ . For this reason, the sparse partial correlation obtained by LASSO is a near-optimal under CS and the brain network recovery based on the partial correlation can be theoretically guaranteed.

## VI. NUMERICAL EXPERIMENTS

### A. Imaging Data

1) *Subjects*: There are twenty six children with ASD (24 boys, mean age:  $6.0 \pm 1.8$  years) and eleven children with PedCon (8 boys, mean age:  $9.73 \pm 2.55$  years). The ASD group, who was diagnosed by the Korean version of Autism Diagnostic Interview-Revised (K-ADI-R) [53], was recruited from Child and Adolescent Psychiatric Outpatient Clinic of Seoul National University Hospital, South Korea. The pediatric controls comprised as children who failed to meet the criteria of any psychiatric disorder and visited the clinic for IQ evaluation.

2) *Image Acquisition*: All PET scans were obtained from ECAT EXACT 47 (Siemens-CTI, Knoxville, USA) PET scanner with an intrinsic resolution of 5.2 mm FWHM. PET images were 47 contiguous planes with a thickness of 3.4 mm. After transmission scan measured by  $^{68}\text{Ge}$  rod sources for attenuation correction, emission scan was administered. All participants were scanned under the normal environmental noise of the scanner room. Image reconstruction was performed using a filtered back-projection algorithm (Shepp-Logan filter at a cutoff frequency of 0.3 cycles/pixel as  $128 \times 128 \times 47$  matrices of size  $2.1 \times 2.1 \times 3.4$  mm).

3) *Preprocessing*: All PET data were preprocessed using Statistical Parametric Mapping (SPM 2, University College of London, UK), implemented in the Matlab 6.5 (Mathworks Inc., USA) environment. The PET data is spatially normalized to Korean standard template space developed by 78 Korean normal right-handed volunteers (Male/Female=49/29) based on MR and PET images. The mean FDG uptake within ROIs was extracted using Statistical Probabilistic Anatomical Map-Korean version (SPAM-K) [54]. The values of FDG uptake were globally normalized to the individual's total gray matter mean count.

### B. Controlling the sparsity

We selected the optimal sparsity  $\lambda$  using the leave-one-out cross-validation in Section IV-B. Since the number of ASD and PedCon data was 26 and 11 respectively, the leave-one-out scheme produced 26 and 11 partial correlation maps of ASD and PedCon for  $\lambda = 0.01, 0.1, 0.2, \dots, 2$ . The blue plots in Fig. 4 (f) and (g) show the number of edges in the network as we increases  $\lambda$  values. If the sample variance of the number

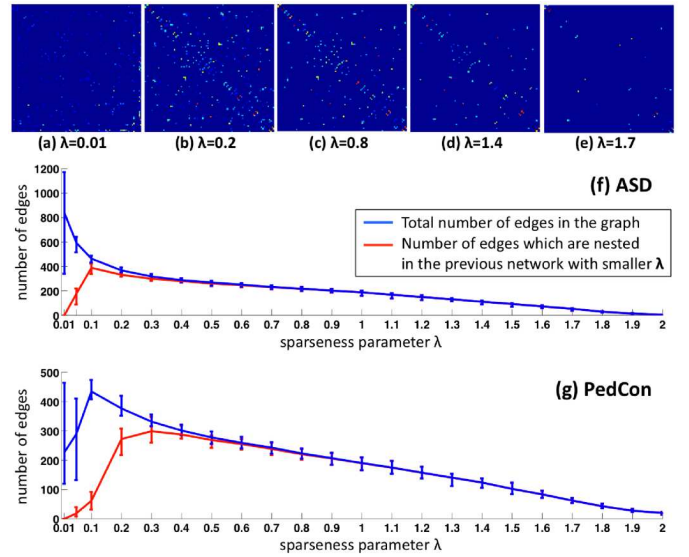


Fig. 4. (a)-(e) show the partial correlation maps for different  $\lambda = 0.01, 0.2, 0.8, 1.4$  and  $1.7$ . The number of edges in the network changes as  $\lambda$  value changes for ASD (f) and PedCon (g). The blue plots are the total numbers of edges in the network. The red plots are the numbers of edges which are included in the previous network with smaller  $\lambda$  among all edges in a graph.

of edges is large during the cross validation, it implies that the obtained partial correlation is not stable. The variance is relatively large between  $0 \leq \lambda \leq 0.3$ . But for  $\lambda \geq 0.4$ , the variance gets relatively smaller.

For the sparse network with  $\lambda \geq 0.4$ , most of edges in a network start to belong to previous networks which have smaller sparsity (red line in Fig. 4). If the sparseness is not too small in the small- $n$  large- $p$  situation, finding the sparseness parameter is related to fixing the threshold. Thus, we chose the sparseness parameter  $\lambda = 0.4$  for estimating the optimal partial correlation.

### C. Reproducibility of Brain Network

To validate our method, we checked the reproducibility (consistency or stability) of networks using the cross-validation. We computed the mean and standard deviation of partial correlations. If the standard deviation is less than 0.1, we consider the network to be reproducible [55]. If the mean is less than 0.1, we categorized it into zero. In this way, we categorized the partial correlation into 4 classes: stable zero, stable nonzero, unstable zero and unstable nonzero.

In Table I, we compared the reproducibility of 3 different methods, the pseudo-inverse (PINV), the penalized maximum likelihood (PML) and the penalized linear regression (PLR), by measuring the ratio of partial correlations belonging to 4 classes. The methods were briefly explained in Section II-B. The percentages of stable elements obtained by PINV, PML and PLR are 61.90 %, 96.39 % and 100 % for ASD and 11.69 %, 89.54 % and 96.91 % for PedCon. Among nonzero elements, PML has the stable and unstable elements in the ratios (1.74:0.92) and (1.29:2.60) for ASD and PedCon. On the other hand, PLR has the ratios (2.68:0) and (1.83:1.48) for

ASD and PedCon. Since the number of PedCon data is smaller than ASD, PedCon is less stable than ASD. For PedCon, PLR finds more stable elements than PML. This empirical evidence for the consistency and near-optimal-recovery of PLR-based partial correlation is also discussed in Peng, et al. [27].

#### D. Visualization of Modular Structure

After obtaining the optimal partial correlation with  $\lambda = 0.4$ , we thresholded the brain network when the number of clusters are maximized in Section IV-B. For simplicity, we visualized the estimated 3D brain networks in a 2D space using ISOMAP (Fig. 6). ISOMAP is an embedding technique that preserves the relative distance between neighboring nodes [56]. Fig. 6 (c) and (d) show the ASD and PedCon brain networks thresholded. The color of nodes corresponds to a lobe the nodes belong. The nodes and edges of the networks in (c) and (d) have different colors for different clusters. The color representing the cluster in (c) and (d) is selected by blending node colors in the cluster. If the cluster consists of nodes in the same lobe, the color of cluster is identical to the lobe color. We observe that the ASD network is more similar to the lobe coloring scheme compared to the PedCon network. Based on the Bonferroni correction, the corrected p-value of each thresholds obtained by the maximum number of clusters are  $p = 0.0189 \pm 0.0095$  for 26 ASD networks and  $p = 0.0253 \pm 0.0164$  for 11 PedCon networks [19].

#### E. Significance of Network Differences

We constructed 26 ASD and 11 PedCon networks by thresholding the partial correlation based on four different methods: PINV, PML and PLR and the correlation (CORR). The CORR method was introduced in Section IV-A. We also generated 100 weighted random networks using Erdős-Rényi (ER) model [57]. In the ER random network, the weight of edges is chosen randomly and uniformly in the interval  $[0, 1]$ . The generated random networks are thresholded in the same way.

1) *Global Inference:* From each network, we extracted 4 global features: the number of (1) edges, (2) clusters, (3) edges connected between two ROIs in different lobes and (4) edges connected between two ROIs within the same lobe. To quantify the network differences among ASD, PedCon and random networks (Rand), the two-sample Wilcoxon rank sum test was applied in a pairwise fashion. In Fig. 5, each panel shows the box plots of 4 global features for ASD (red), PedCon (blue) and Rand (green) using 4 different methods. The significant group difference at 0.01 level is marked with the asterisk (\*). All global features obtained by PLR are significantly different from random networks. Although the number of edges and clusters of ASD and PedCon were not different in PLR, the number of edges connected between lobes and the number of edges connected within a lobe were significantly different ( $p < 0.001$ ). The results show that the autistic brain network has local overconnectivity and long-range underconnectivity [58]. The network obtained by CORR also shows similar results.

2) *Local Inference:* We also extracted 28 local features from ASD, PedCon and random networks (RAND). We considered ROIs belong to 7 lobes, frontal (F), subcortical (S), limbic (L), temporal (T), parietal (P), occipital (O) and limbic (L) lobes and Cerebellum (C). 7 features are the number of edges connected within each lobe and 21 features are the number of edges connected between two lobes. We performed the Wilcoxon rank sum test on three pairs (ASD,RAND), (PedCon,RAND) and (ASD,PedCon) and used the Bonferroni procedure for the multiple comparison correction.

Fig. 7 shows the brain networks and the connectivity matrices of 7 lobes on two groups (column) and three methods (row). The ROI locations of the brain networks are shown in Fig. 6. We obtained the thresholded brain networks based on the number of connected components and counted the number of connection of each edge during the cross validation. Thicker edges represent more reproducible connections. The matrix entries are the mean and standard deviation of number of edges between the corresponding lobes during the cross validation. The gray colored entries are the connections that are statistically different from the random networks at 0.05 level. The red box indicates the significantly different connections in all pairwise comparisons. PML, PLR and CORR found 18, 26 and 27 gray colored entries in ASD connectivity matrices and 14, 18 and 26 one in PedCon. The number of red boxes for PML, PLR and CORR is 4, 9 and 5, respectively. The results show that PLR finds more significant network than PML and CORR.

From Fig. 7 (b), the local overconnectivity was found in the ASD network in the frontal, parietal, limbic and subcortical lobes ( $p < 0.001$ , corrected for comparison  $p < 0.01$ ). The long-range underconnectivity patterns between lobes were observed in the ASD network: frontal-parietal, frontal-temporal, frontal-limbic, parietal-temporal and occipital-temporal ( $p < 0.001$ , corrected for comparison  $p < 0.01$ ). Functional underconnectivity between frontal and parietal regions in ASD was quite consistent with other studies, because it is associated with deficits of planning and problem solving in ASD [41]. Frontal mirror neuron system was suggested to mediate understanding of other's emotional states in concert with limbic center, such as amygdala. This emotional dysfunction in ASD children might be explained through the abnormal connectivity between frontal and limbic system [59]. The occipital regions showed reduced functional connectivity with the temporal regions, which was associated with mentalizing impairment in ASD [60]. Abnormal behavioral phenotypes in ASD could be involved in these long-range dysconnectivities. Thus, we can say that PLR finds more representative and discriminative networks which fit the previous studies better.

## VII. CONCLUSIONS

In this paper, we showed that the problem of estimating correlation and partial correlation can be formulated in the sparse linear regression framework. The partial correlation is widely used in modeling highly correlated networks since it shows the actual dependency between two nodes by factoring out the redundant dependences of other nodes. However, the



TABLE I

AFTER LEAVE-ONE-OUT CROSS VALIDATION, THE ENTRIES OF THE PARTIAL CORRELATION MAP ARE CATEGORIZED INTO 4 CLASSES: STABLE ZERO, STABLE NONZERO, UNSTABLE ZERO AND UNSTABLE NONZERO. THE STABLE AND UNSTABLE CLASSES ARE DETERMINED BY WHETHER THE STANDARD DEVIATION IS LESS THAN 0.1 OR NOT. THE ZERO AND NONZERO CLASSES ARE DETERMINED BY WHETHER THE MEAN IS LESS THAN 0.1 OR NOT. WE COMPARED PINV, PML AND PLR METHODS ON ASD, PEDCON AND TWO RANDOM NETWORKS OBTAINED BY PERMUTING AND SELECTING THE GIVEN DATA RANDOMLY. THE RESULTS ARE GIVEN IN TERMS OF THE PERCENTAGE OF EDGES BELONGING TO EACH CLASS AMONG TOTAL 4656 EDGES.

Group	Class	PINV				PML				PLR			
		TRUE		Random		TRUE		Random		TRUE		Random	
ASD	stable	61.90		0.02		96.39		82.26		100		96.56	
	zero nonzero	39.78	22.12	0	0.02	94.65	1.74	82.24	0.02	97.32	2.68	95.98	0.58
	unstable	38.10		99.98		3.60		17.74		0		3.44	
	zero nonzero	25.73	12.37	70.79	29.19	2.68	0.92	12.93	4.81	0	0	1.31	2.13
PedCon	stable	11.69		0.04		89.54		66.32		96.91		92.33	
	zero nonzero	6.64	5.05	0	0.04	88.25	1.29	66.32	0	95.08	1.83	92.33	0
	unstable	88.32		99.96		10.46		33.68		3.09		7.67	
	zero nonzero	54.23	34.09	73.41	26.55	7.86	2.60	26.01	7.67	1.61	1.48	4.96	2.71

estimation of the partial correlation using the traditional least squares method is unreliable when the number of observations is smaller (small- $n$ ) than the large number of nodes (large- $p$ ) in complex networks. To remedy the small- $n$  large- $p$  problem, the  $l_1$ -penalty for the sparseness constraint is usually introduced to the regression. The penalized linear regression, as known as LASSO, naturally leads to sparse brain network modeling. Under the i.i.d. Gaussian assumption, the proposed brain network model can recover sparse underlying signal even from small number of noisy measurements.

The numerical experiments show that the sparse brain network can be estimated consistently. The proposed method was applied in characterizing the local overconnectivity and long-range underconnectivity in the autistic brain. Our result is consistent with previous autism-related clinical studies [38], [39], [49].

#### ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their useful suggestions. Hyekyoung Lee would like to thank Jiho Yoo for helpful discussions and comments.

#### REFERENCES

- [1] O. Sporn and J. Zwi, "The small world of the cerebral cortex," *Neuroinformatics*, vol. 2, pp. 145–162, 2004.
- [2] V. Eguiluz, D. Chialvo, G. Cecchi, M. Baliki, and A. Apkarian, "Scale-free brain functional networks," *Physical Review Letters*, vol. 94, p. 018102, 2005.
- [3] C. Stam, "Functional connectivity patterns of human magnetoencephalographic recordings: a small-world network?" *Neuroscience Letters*, vol. 355, pp. 25–28, 2004.
- [4] C. Stam, B. Jones, G. Nolte, M. Breakspear, and P. Scheltens, "Small-world networks and functional connectivity in alzheimer's disease," *Cerebral Cortex*, vol. 17, pp. 92–99, 2007.
- [5] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore, "A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs," *Journal of Neuroscience*, vol. 26, pp. 63–72, 2006.
- [6] R. Salvador, J. Suckling, M. Coleman, J. Pickard, D. Menon, and E. Bullmore, "Neurophysiological architecture of functional magnetic resonance images of human brain," *Cerebral Cortex*, vol. 15, pp. 1332–1342, 2005.
- [7] Y. He, Z. Chen, and A. Evans, "Small-world anatomical networks in the human brain revealed by cortical thickness from MRI," *Cerebral Cortex*, vol. 17, pp. 2407–2419, 2007.
- [8] Z. Chen, Y. He, P. Rosa-Neto, J. Germann, and A. Evans, "Revealing modular architecture of human brain structural networks by using cortical thickness from MRI," *Cerebral Cortex*, pp. 2374–2381, 2008.
- [9] M. Valencia, M. Pastor, M. Fernandez-Seara, J. Artieda, J. Martinerie, and M. Chavez, "Complex modular structure of large-scale brain networks," *Chaos*, vol. 19, 2009.
- [10] P. Laurienti, C. Hugenschmidt, and S. Hayasaka, "Modularity maps reveal community structure in the resting human brain," *Nature Precedings*, 2009.
- [11] D. Bassett, "Small-world brain networks," *The Neuroscientist*, vol. 12, pp. 512–523, 2006.
- [12] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, pp. 1059–1069, 2010.
- [13] O. Sporns, G. Tononi, and R. Kotter, "The human connectome: a structural description of the human brain," *PLoS Computational Biology*, vol. 1, 2005.
- [14] J. Cao and K. Worsley, "The geometry of correlation fields with an application to functional connectivity of the brain," *Annals of Applied Probability*, vol. 9, pp. 1021–1057, 1999.
- [15] M. Koch, D. Norris, and M. Hund-Georgiadis, "An investigation of functional and anatomical connectivity using magnetic resonance imaging," *NeuroImage*, vol. 16, pp. 241–250, 2002.
- [16] G. Marrelec, A. Krainik, H. Duffau, M. Péligrini-Issac, S. Lehericy, J. Doyon, and H. Benali, "Partial correlation for functional brain interactivity investigation in functional MRI," *NeuroImage*, vol. 32, 2006.
- [17] A. McIntosh, F. Bookstein, J. Haxby, and C. Grady, "Spatial pattern analysis of functional brain images using partial least squares," *NeuroImage*, vol. 3, pp. 143–157, 1996.
- [18] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the lasso," *Annals of Statistics*, vol. 34, pp. 1436–1462, 2006.
- [19] J. Schäfer and K. Strimmer, "An empirical bayes approach to inferring large-scale gene association networks," *Bioinformatics*, vol. 21, pp. 754–764, 2005.
- [20] O. Banerjee, L. El Ghaoui, and A. d'Aspremont, "Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data," *Journal of Machine Learning Research*, vol. 9, pp. 485–516, 2008.
- [21] P. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *Annals of Statistics*, vol. 36, pp. 199–227, 2008.
- [22] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, pp. 432–441, 2008.
- [23] L. Sun, R. Patel, J. Liu, K. Chen, T. Wu, J. Li, E. Reiman, and Y. Ye, "Mining brain region connectivity for alzheimer's disease study via sparse inverse covariance estimation," in *Proceedings of the ACM*

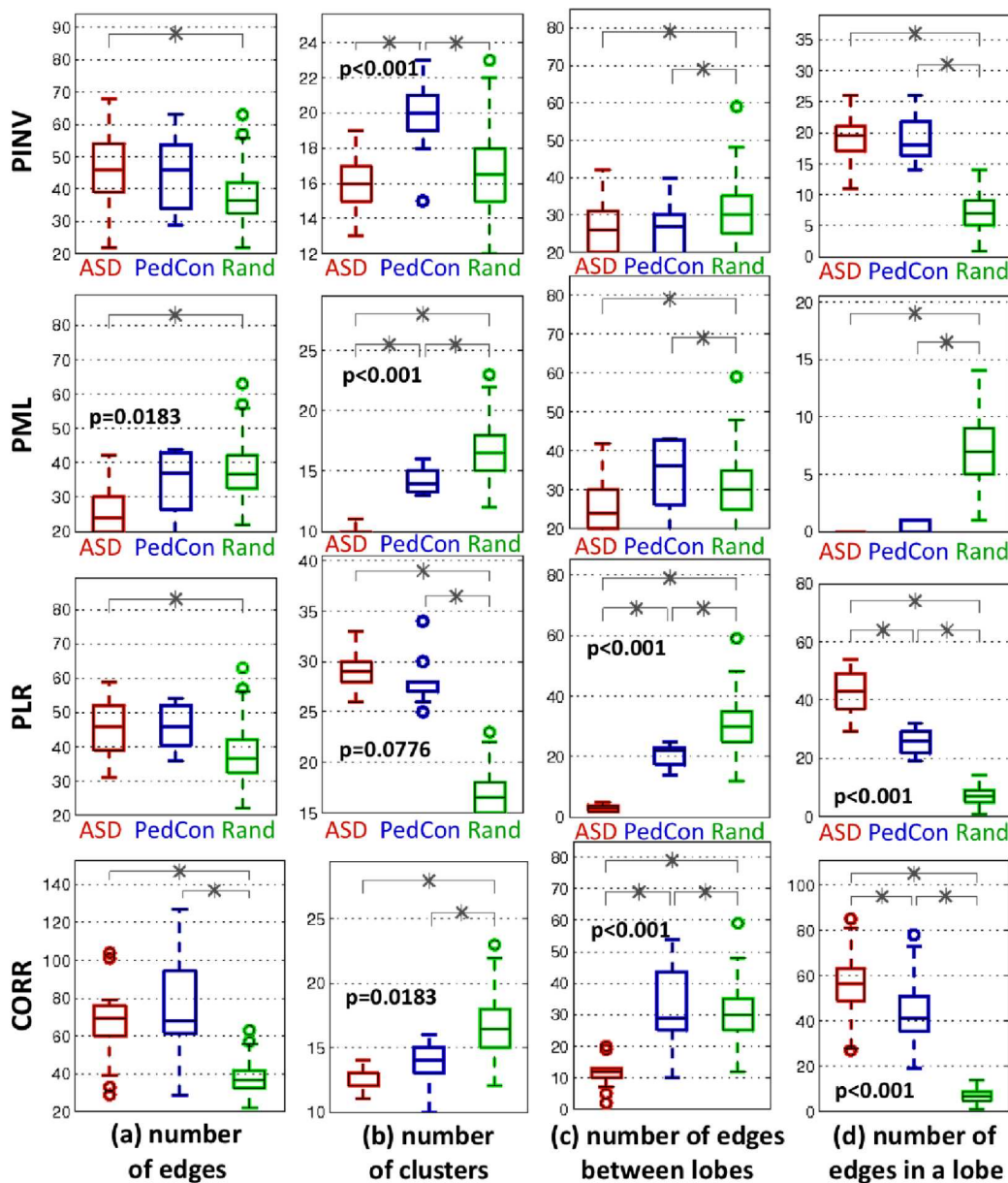


Fig. 5. The statistical significance of the pairwise group difference among ASD, PedCon and random networks on (a) the number of edges, (b) the number of clusters, (c) the number of edges connected between lobes and (d) the number of edges connected in a lobe using the four different methods (PINV, PML, PLR and CORR). The asterisk (\*) represents  $p < 0.01$  based on the Wilcoxon rank sum test. In PLR, all features are significantly different with  $p = 0.7898$ ,  $p = 0.0776$ ,  $p < 0.001$  and  $p < 0.001$  respectively.

- SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2009.
- [24] G. Cecchi, I. Rish, B. Thyreau, B. Thirion, M. Plaze, M. Paillere-Martinot, C. Martelli, J. Martinot, and J. Poline, "Discriminative network models of schizophrenia," in *Advances in Neural Information Processing Systems (NIPS)*, 2009.
- [25] S. Huang, J. Li, L. Sun, J. Liu, T. Wu, K. Chen, A. Fleisher, E. Reiman, and J. Ye, "Learning brain connectivity of alzheimer's disease from neuroimaging data," in *Advances in Neural Information Processing Systems (NIPS)*, 2009.
- [26] S. Achard and E. Bullmore, "Efficiency and cost of economical brain functional networks," *PLoS Computational Biology*, vol. 3, 2007.
- [27] J. Peng, P. Wang, N. Zhou, and J. Zhu, "Partial correlation estimation by joint sparse regression models," *Journal of the American Statistical Association*, vol. 104, pp. 735–746, 2009.
- [28] O. Banerjee, L. Ghaoui, A. d'Aspremont, and G. Natsoulis, "Convex optimization techniques for fitting sparse gaussian graphical models," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2006.
- [29] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1999.
- [30] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society B*, vol. 58, pp. 267–288, 1996.
- [31] E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, pp. 21–30, 2008.
- [32] D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, 2006.
- [33] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory*, pp. 489–509, 2006.
- [34] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, 2007.
- [35] E. Candès and T. Tao, "Near optimal signal recovery from random

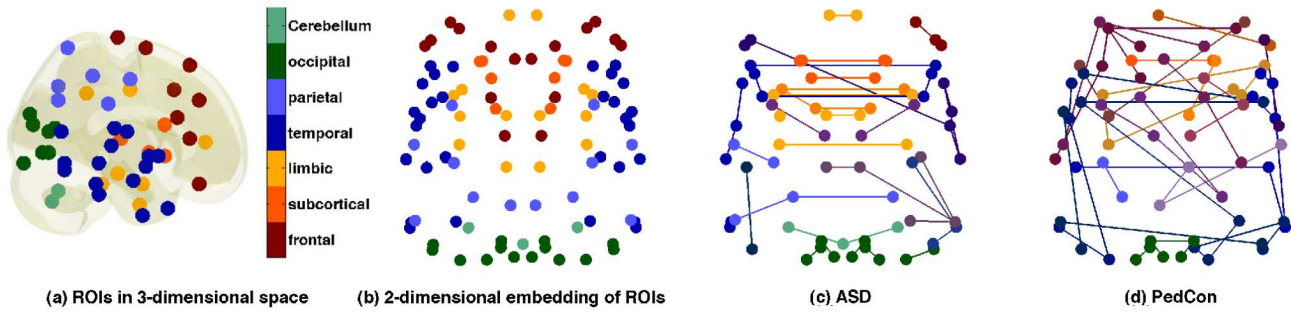


Fig. 6. Visualization of ROIs in (a) 3D and (b) 2D spaces. 3D ROIs are embedded into the 2D space by ISOMAP, which preserves the relative distance between nodes. Each lobe is represented by different color as shown in the colorbar (a). The clustered brain networks are given for (c) ASD and (d) PedCon. In (c) and (d), the color represents cluster.

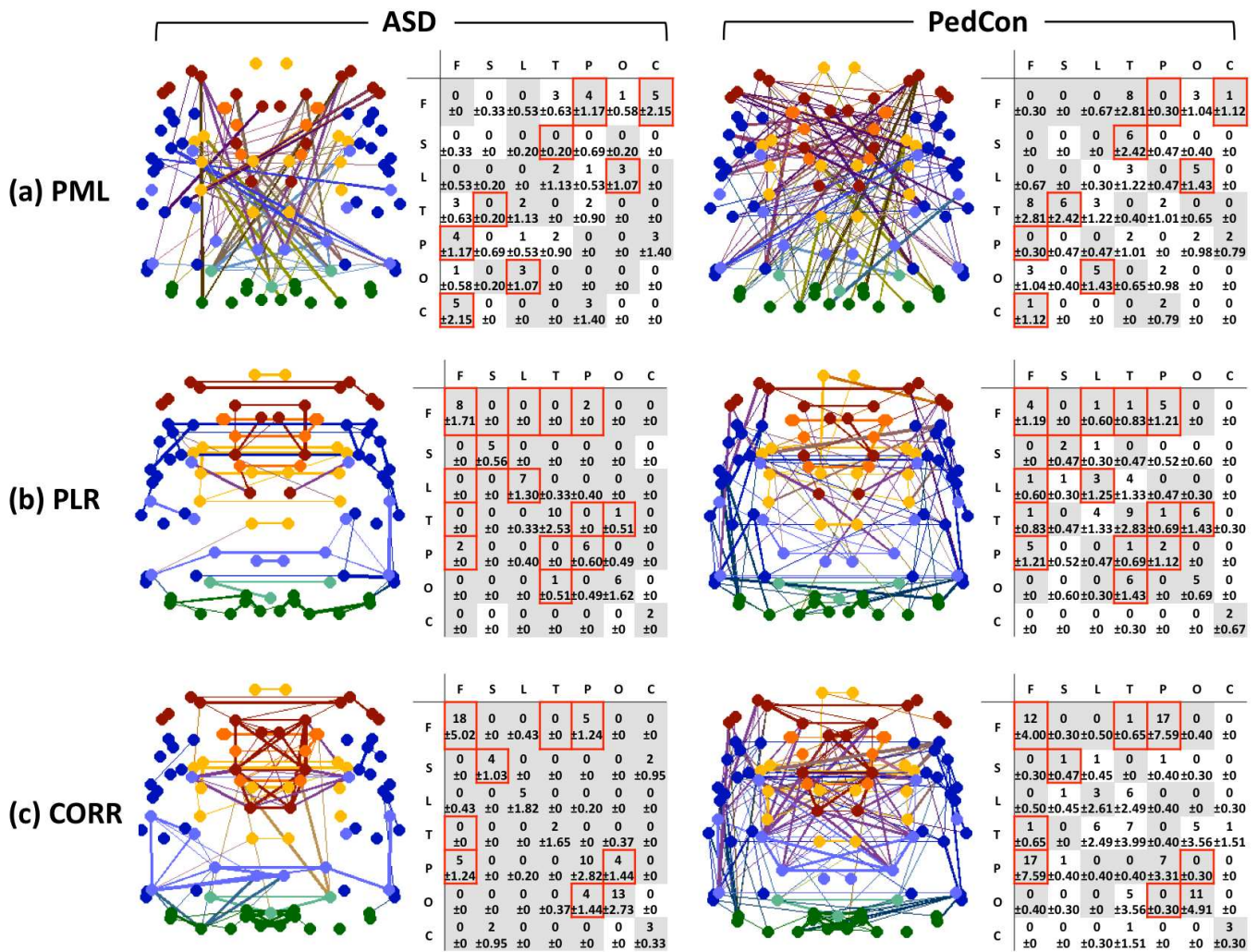


Fig. 7. The Brain networks and the connectivity matrices of 7 lobes for PML, PLR and CORR methods. The lobes are frontal (F), subcortical (S), limbic (L), temporal (T), parietal (P) and occipital (O) lobes and cerebellum (C). The nodes represent ROIs which are embedded into the 2D space by ISOMAP. The width of edge represents the number of connections during the cross validation. If the edge is connected more frequently, it is thicker. The elements of the connectivity matrices are the mean and standard deviation of the number of edges between 7 lobes during the cross validation. The gray colored entries are significant connection difference from the random networks at 0.05 level. The  $p$ -value is determined using the Wilcoxon rank sum test and the Bonferroni correction. The red boxes are the significant connection differences between ASD and PedCon at 0.05 level.

- projections: Universal encoding strategies?" *IEEE Trans. Information Theory*, pp. 5406–5425, 2006.
- [36] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, pp. 1207–1223, 2006.
- [37] N. Meinshausen and B. Yu, "Lasso-type recovery of sparse representations for high-dimensional data," Department of Statistics, UC Berkeley, Tech. Rep. 720, 2006.
- [38] M. Murias, S. Webb, J. Greenson, and D. G., "Resting state cortical connectivity reflected in EEG coherence in individuals with autism," *Biological Psychiatry*, vol. 62, pp. 270–273, 2007.
- [39] D. Williams, G. Goldstein, and N. Minshew, "Neuropsychologic functioning in children with autism: Further evidence for disordered complex information-processing," *Child Neuropsychol.*, pp. 279–298, 2006.
- [40] E. Courchesne, K. Pierce, C. Schumann, E. Redcay, J. Buckwalter, D. Kennedy, and J. Morgan, "Mapping early brain development in autism," *Neuron*, vol. 56, pp. 399–413, 2007.
- [41] M. Just, V. Cherkassky, T. Keller, and N. Minshew, "Cortical activation and synchronization during sentence comprehension in high-functioning autism: Evidence of underconnectivity," *Brain*, vol. 127, pp. 1811–1821, 2004.
- [42] D. Bassett and E. Bullmore, "Human brain networks in health and disease," *Current Opinion in Neurology*, vol. 22, pp. 340–347, 2009.
- [43] S. Huang, J. Li, L. Sun, J. Ye, A. Fleisher, T. Wu, K. Chen, and E. Reiman, "Learning brain connectivity of alzheimers disease by sparse inverse covariance estimation," *NeuroImage*, vol. 50, pp. 935–949, 2010.
- [44] A. Laub, *Matrix Analysis for Scientists and Engineers*. SIAM, 2004.
- [45] Y. Sharon, J. Wright, and Y. Ma, "Computation and relaxation of conditions for equivalence between  $l_1$  and  $l_0$  minimization," UIUC, Tech. Rep. UILU-ENG-07-2008, 2007.
- [46] E. Candès, M. Rudelson, T. Tao, and R. Vershynin, "Error correction via linear programming," 2005, pp. 295–308.
- [47] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," Stanford University, Tech. Rep., 2008.
- [48] D. G. Luenberger, *Linear and Nonlinear Programming*. Addison-Wesley, 1984.
- [49] N. Minshew and D. Williams, "The new neurobiology of autism: Cortex, connectivity, and neuronal organization," *Arch Neurol.*, pp. 945–950, 2007.
- [50] D. Hsu, S. Kakade, J. Langford, and T. Zhang, "Multi-label prediction via compressed sensing," in *Advances in Neural Information Processing Systems (NIPS)*, 2009.
- [51] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathématique*, pp. 589–592, 2008.
- [52] M. Chung, K. Worsley, S. Robbins, T. Paus, J. Taylor, J. Giedd, J. Rapoport, and A. Evans, "Deformation-based surface morphometry applied to gray matter deformation," *NeuroImage*, vol. 18, pp. 198–213, 2003.
- [53] H. Yoo, *Korean version of Autism Diagnostic Interview-Revised*. Seoul: Hakjisa, 2007.
- [54] J. Lee, D. Lee, Y. Kim, J. Kim, J. Lee, B. Koo, J. Kim, J. Kwon, T. Yoo, K. Chang, S. Kim, H. Kang, E. Kang, J. Chung, and M. Lee, "Quantification of brain images using korean standard templates and structural and cytoarchitectonic probabilistic maps," *Korean Journal of Nuclear Medicine*, vol. 38, pp. 241–252, 2004.
- [55] "IUPAC compendium of chemical terminology – the gold book," 2009. [Online]. Available: <http://www.iupac.org/goldbook>
- [56] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," pp. 2319–2323, 2000. [Online]. Available: <http://isomap.stanford.edu/>
- [57] P. Erdős and A. Rényi, "On the evolution of random graphs," in *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 1960, pp. 17–61.
- [58] M. Belmonte, G. Allen, A. Beckel-Mitchener, L. Boulanger, R. Carper, and S. Webb, "Autism and abnormal development of brain connectivity," *Journal of Neuroscience*, vol. 24, pp. 9228–9231, 2004.
- [59] M. Dapretto, M. Davies, J. Pfeifer, A. Scott, M. Sigman, S. Bookheimer, and M. Iacoboni, "Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders," *Nature Neuroscience*, pp. 28–30, 2005.
- [60] F. Castelli, C. Frith, F. Happé, and U. Frith, "Autism, asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes," *Brain*, pp. 1839–1849, 2002.