Remodeling Pearson’s Correlation for Functional Brain Network Estimation and Autism Spectrum Disorder Identification

Weikai Li¹,², Zhengxia Wang¹*, Limei Zhang², Lishan Qiao² and Dinggang Shen³,⁴*

¹ College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing, China, ² School of Mathematics, Liaocheng University, Liaocheng, China, ³ Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, ⁴ Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea

Functional brain network (FBN) has been becoming an increasingly important way to model the statistical dependence among neural time courses of brain, and provides effective imaging biomarkers for diagnosis of some neurological or psychological disorders. Currently, Pearson’s Correlation (PC) is the simplest and most widely-used method in constructing FBNs. Despite its advantages in statistical meaning and calculated performance, the PC tends to result in a FBN with dense connections. Therefore, in practice, the PC-based FBN needs to be sparsified by removing weak (potential noisy) connections. However, such a scheme depends on a hard-threshold without enough flexibility. Different from this traditional strategy, in this paper, we propose a new approach for estimating FBNs by remodeling PC as an optimization problem, which provides a way to incorporate biological/physical priors into the FBNs. In particular, we introduce an $L_1$-norm regularizer into the optimization model for obtaining a sparse solution. Compared with the hard-threshold scheme, the proposed framework gives an elegant mathematical formulation for sparsifying PC-based networks. More importantly, it provides a platform to encode other biological/physical priors into the PC-based FBNs.

To further illustrate the flexibility of the proposed method, we extend the model to a weighted counterpart for learning both sparse and scale-free networks, and then conduct experiments to identify autism spectrum disorders (ASD) from normal controls (NC) based on the constructed FBNs. Consequently, we achieved an 81.52% classification accuracy which outperforms the baseline and state-of-the-art methods.

Keywords: functional brain network, functional magnetic resonance imaging, Pearson’s correlation, sparse representation, scale-free, autism spectrum disorder

INTRODUCTION

Autism spectrum disorder (ASD) is a neural developmental syndrome defined by the defect in social reciprocity, restricted communication, and repetitive behaviors (Lord et al., 2000; Frith and Happé, 2005; Baio, 2014; Wee et al., 2016). The prevalence rate of ASD is fast growing in the worldwide. According to the report supported by the USA Centers for Disease Control and Prevention (Baio, 2014), 1.47% of American children was marred by some forms of ASD with...
a nearly 30% increasing rate in the last 2 years. However, the standard ASD diagnosis methods (e.g., parent interview and participant interview) are highly based on behaviors, and symptoms of the disease (Gillberg, 1993; Lord and Jones, 2012; Segal, 2013), resulting in missing the best cure opportunity. At the same time, the measurement at the gene level (Wang et al., 2009; O’Roak et al., 2012) can benefit an early diagnosis, but it is less popular due to high costs and complexity. Recent evidences show that the unusual brain activity (Brambilla et al., 2003; Ecker et al., 2010; Lo et al., 2011; Nielsen et al., 2013) and abnormal functional disruptions in some brain regions (Allen and Courchesne, 2003; Anderson et al., 2011; Delmonte et al., 2012) such as, hippocampus and frontal region have a high correlation with ASD. Thus, it is possible to identify informative biomarkers and help identify ASD by analyzing the activity data of brain.

Functional magnetic resonance imaging (fMRI) is currently a widely-used non-invasive technique for measuring brain activities (Brunetti et al., 2006; Kevin et al., 2008; Jin et al., 2010). However, it is hard to identify patients from normal controls (NC) by direct comparison of the fMRI data (i.e., time courses), since the spontaneous brain activities are random and asynchronous across subjects. In contrast, the functional brain network (FBN) constructed by, for example, the correlation of the time series can provide a more stable measurements for classifying different subjects (Smith et al., 2011; Sporns, 2011; Wee et al., 2012; Stam, 2014; Rosa et al., 2015). In fact, FBN identifies functional connections between brain regions, voxels, or ROIs (Horwitz, 2003), which has already been verified to be highly related to some neurological or psychological diseases such as, ASD (Theije et al., 2011; Gotts et al., 2012), mild cognitive impairment (Fan and Browndyke, 2010; Wee et al., 2012, 2014; Yu et al., 2016), Alzheimer’s disease (Supekar et al., 2008; Huang et al., 2009; Liu et al., 2012) and so on.

The commonly used scheme to estimate FBNs is based on the second-order statistics that tend to work better than the high-order counterparts (Smith et al., 2011). The typical second-order estimation methods include Pearson’s correlation (PC), and sparse representation (SR), etc. PC estimates FBNs by measuring the full correlation between different brain regions (ROIs1). The full correlation is simple, computationally efficient and statistically robust, but tends to include confounding effects from other brain regions. In contrast, the partial correlation can alleviate this problem by regressing out the potential confounding influence. However, calculating the partial correlation involves an inverse operation on the covariance matrix, which is generally ill-posed, especially when the number of time points is fewer than the number of brain regions. Therefore, regularization techniques such as, SR (with a L1-norm regularizer) are generally used to achieve a stable solution (Lee et al., 2011).

In this paper, we mainly focus on the PC-based methods, because we empirically found that, in our experiments, the PC-based (full correlation) methods work better than the SR-based (partial correlation) counterpart. However, the original PC scheme always results in FBNs with a dense topological structure (Fornito et al., 2016), since the BOLD signals commonly contain noises, micro head-motion (Power et al., 2013; Yan et al., 2013) and/or mind wandering (Mason et al., 2007). In practice, a threshold is commonly used to sparsify the PC-based FBNs by filtering out the noisy or weak connections. Although it is simple and effective, the threshold scheme is hard without enough flexibility. To address this problem, in this paper, we reformulate the estimation of PC network as an optimization problem, and motivated by the SR model (see Section Related Methods), we introduce an L1-norm regularizer for achieving a sparse solution. Different from the traditional hard-threshold scheme, the proposed method is more flexible, and can in principle incorporate any informative prior into the PC-based FBN construction. Specifically, the main contributions of this paper can be summarized as follows.

1. We propose a novel strategy to estimate PC by remodeling it in an optimization learning framework. Consequently, biological/physical priors can be incorporated more easily and naturally for constructing better PC-based FBNs.

2. We introduce an L1-norm regularizer into the proposed framework for estimating sparse FBNs, and further extend it to a weighted version for constructing both sparse and scale-free FBNs. These two instantiations illustrate that the proposed method is more flexible than the traditional hard-threshold scheme.

3. We use the PC-based FBNs constructed by our framework to distinguish the ASDs from NCs, and achieve 81.52% classification accuracy, which outperforms the baseline and state-of-the-art methods.

The remainder of this paper is organized as follows. In Section Materials and Methods, we introduce the material and methods. In particular, we first introduce the participants and review two related methods, i.e., PC and SR. Then, we reformulate PC into an optimization model and propose two specific PC-based FBN estimation methods, including the motivations, models, and algorithms for these two methods. In Section Results, we evaluate the two proposed methods with experiments on identifying ASD. In Section Discussion, we discuss our findings and prospects of our work. In Section Conclusion, we conclude the entire paper briefly.

**MATERIALS AND METHODS**

**Data Acquisition**

In this paper, we have the same data set as the one in a recent study (Wee et al., 2016). Specifically, the data set includes resting-state fMRI (R-fMRI) data of 45 ASD subjects and 47 NC subjects (with ages between 7 and 15 years old). All these data are publicly available in the ABIDE database (Di et al., 2014). The demographic information of these subjects is summarized in Table 1. The ASD diagnostic was based on the autism criteria part in Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR). The psychopathology for differential diagnosis and

---

1In this paper, we will interchangeably use regions of interest (ROIs) and brain regions to denote network/graph nodes for the convenience of presentation.
comorbidities with Axis-I disorders was assessed by parent interview or participant interview. In particular, the parent interview was based on the Structured Clinical Interview for DSM-IV-TR Axis-I Disorders, Non-patient Edition (SCID-I/NP) and the Adult ADHD Clinical Diagnostic Scale (ACDS) for adults (>18.0 years old). Exclusion of the comorbid ADHD needs to meet all criteria for ADHD (except for criterion E) in the DSM-IV-TR. Inclusion as a NC needs to exclude the entire current Axis-I disorders by KSADS-PL, SCID-I/NP, and ACDS interviews.

Data Preprocessing

All R-fMRI images were acquired using a standard echo-planar imaging sequence on a clinical routine 3T Siemens Allegra scanner. During 6 min R-fMRI scanning procedure, the subjects were required to relax with their eyes focusing on a white fixation cross in the middle of the black background screen projected on a screen. The imaging parameters include the flip angle = 90°, 33 slices, TR/TE = 2000/35 ms with 180 volumes, and 4.0 mm voxel thickness. Data preprocessing was made by the standard software, statistical parametric mapping (SPM8 http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). Specifically, the first 10 R-fMRI images of each subject were discarded to avoid signal shaking. The remaining images were calibrated as follows: (1) normalization to MNI space with resolution 3 × 3 × 3 mm³; (2) regression of nuisance signals (ventricle, white matter, global signals, and head-motion) with Friston 24-parameter model (Friston et al., 1996); (3) band-pass filtering (0.01–0.08 Hz); (4) signal de-trending. After that, the pre-processed BOLD time series signals were partitioned into 116 ROIs, according to the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). At last, we put these time series into a data matrix $X \in R^{170 \times 116}$.

Functional Brain Network Estimation

After extracting the data matrix $X$ from the R-fMRI data, we construct FBNs for these subjects based on the methods that will be given in the following subsections.

### Related Methods

It is well known that PC is possibly the most popular method to estimate FBNs (Smith et al., 2013). The mathematical expression of PC is defined as follows:

$$W_{ij} = \frac{(x_i - x_j)^T(x_j - x_j)}{\sqrt{(x_i - x_j)^T(x_j - x_j)(x_j - x_j)^T(x_j - x_j)}},$$

(1)

where $x_i \in R^t$ is the observed time course associated with ith brain regions, $t$ is the number of time nodes, $x_i \in R^t$ has all entries being the mean of the elements in $x_i, i = 1, 2, \cdots, n$, and $n$ is the number of ROIs. Consequently, $x_i - x_j$ is a centralized counterpart of $x_i$.

As discussed previously, PC always generates dense FBNs. Thus, a threshold is often used to sparsify the PC-based FBNs (namely PC$_\text{threshold}$), which can be expressed as follows:

$$W'_{ij}(\text{new}) = \begin{cases} W_{ij}, & W_{ij} > \text{threshold} \\ 0, & \text{otherwise} \end{cases}$$

(2)

where $W'_{ij}(\text{new})$ denotes the connection value between nodes i and j after thresholding.

Different from PC that measures the full correlation, SR is one of the widely-used schemes for modeling the partial correlation (Lee et al., 2011). The model of SR is shown as follows:

$$\min_W \sum_{i=1}^{n} \left\| x_i - \sum_{j \neq i} W_{ij}x_j \right\|^2 + \lambda \sum_{j \neq i} |W_{ij}|,$$

(3)

or equivalently, its matrix form is

$$\min_W \| X - XW \|^2_F + \lambda \| W \|_1$$

(4)

s.t. $W_{ii} = 0, \forall i = 1, \cdots, n$, where $X = [x_1, x_2, \cdots, x_n] \in R^{t \times n}$ represents the fMRI data matrix associated with a certain subject. Each column of $X$ corresponds to the time course from a certain brain region. Note that the L1-norm regularizer in Equation (4) plays a key role in achieving a sparse and stable solution (Lee et al., 2011).

### Our Methods

As two typical examples, PC and SR have been demonstrated to be more sensitive than some complex higher-order methods (Smith et al., 2011). Therefore, in this paper, we mainly focus on these two methods, and we empirically found that PC tends to work better than SR in our experiments. However, compared with SR that controls the sparsity in an elegant mathematical model, the PC sparsifies the networks using an empirical hard threshold. Thus, a natural goal is to develop a new FBN estimation method that can inherit the robustness of PC and meanwhile has a flexible sparsification strategy as in SR. To this end, we first formulate the PC scheme as an optimization model, and then introduce an L1-norm regularizer into the model for achieving a sparse solution.
Without loss of generality, we suppose that the observed fMRI time series $x_i$ of each node is centralized by $x_i - x_i$ and normalized by $\sqrt{\left( x_i - x_i \right)^T \left( x_i - x_i \right)}$. That is, we define the new time series $x_i = \frac{(x_i - x_i)}{\sqrt{\left( x_i - x_i \right)^T \left( x_i - x_i \right)}}$. As a result, we can simplify the PC as the formula $W_{ij} = x_i^T x_j$, which can be easily proved to be the optimal solution of the following optimization problem:

$$
\min_{W_{ij}} \sum_{i,j=1}^n \left\| x_i - W_{ij} x_j \right\|^2.
$$

(5)

In fact, we first expand the objective function in Equation (5) as follows:

$$
\sum_{i,j=1}^n \left\| x_i - W_{ij} x_j \right\|^2 = \sum_{i=1}^n \sum_{j=1}^n \left( x_i - W_{ij} x_j \right)^T \left( x_i - W_{ij} x_j \right)
= \sum_{i=1}^n \left( x_i x_i - 2 W_{ij} x_i x_j + W_{ij} x_j x_j \right)
= \sum_{i=1}^n \left( 1 - 2 W_{ij} x_i x_j + W_{ij} x_j x_j \right).
$$

(6)

Then, letting the derivative $\frac{d \sum_{i,j=1}^n \left\| x_i - W_{ij} x_j \right\|^2}{d W_{ij}}$ be 0, we have the following result:

$$
2 W_{ij} - 2 x_i x_j = 0 \rightarrow W_{ij} = x_i x_j.
$$

(7)

Based on Equation (7), Equation (5) can be further formulated to a matrix form as follows:

$$
\min_{W} \left\| W - X^T X \right\|^2_F.
$$

(8)

Below, we will note that such an optimization view of PC can help improve the traditional PC and further develop new flexible FBN estimation methods.

Motivated by the model of SR, we can naturally incorporate a regularized term into the objective function of Equation (8) for constructing a new platform to estimate FBNs. More specifically, the platform can be formulated using a matrix-regularized learning framework as follows:

$$
\min_{W} \left\| W - X^T X \right\|^2_F + \lambda R(W)
$$

(9)

where $R(W)$ is a regularized term, $\lambda$ is a trade-off parameter, and is a set of additional constraints on the constructed FBNs, such as, the positive definiteness and non-negativity, etc.

Here, we argue that the PC-based FBN learning framework shown in Equation (9) has two advantages: (1) it is statistically robust and scales well, without the ill-posed problem involved in the SR-based (partial correlation) method; (2) biological/physical priors (e.g., sparsity) can be naturally introduced into the model in the form of regularizer for constructing more meaningful FBNs. In order to illustrate the flexibility of the proposed framework, we develop two specific remodeling PC-based FBN estimation methods (I and II) that will be discussed below.

**Method I: Remodeling PC-Based FBN with a Sparsity Prior**

As pointed out previously, the hard-threshold scheme is an effective scheme to sparsify the FBNs, which can be regarded as a special format of the $L_1$-norm. However, generally, the threshold selection is empirical without an elegant mathematical representation. In addition, it is hard to incorporate other biological/physical priors into FBN construction task. In this paper, based on the proposed FBN learning framework, we first introduce the $L_1$-norm as an instantiation of the regularized term $R(W)$, resulting in a new remodeling PC-based FBN estimation model (namely PC$_{\text{sparsity}}$) as follows:

$$
\min_{W} \left\| W - X^T X \right\|^2_F + \lambda \left\| W \right\|_1,
$$

(10)

where $\lambda$ is a regularized parameter for controlling the sparsity of $W$. Obviously, the PC$_{\text{sparsity}}$ reduces to the original PC when $\lambda = 0$. Besides the $L_1$-norm, there are some alternative regularizers, such as, the log-sum strategy (Shen et al., 2013), can be introduced in the proposed framework to sparsify FBNs. Here, we select the $L_1$-norm since it is simple and popular.

The objective function of Equation (8) is convex but indiffferentiable due to the $L_1$-norm regularizer. A number of algorithms have been developed to address the indiffferentiable convex optimization problem in the past few years (Donoho and Elad, 2003; Meinshausen and Bühlmann, 2006; Tomioka and Sugiyama, 2009; Zhao, 2013). Here, we employ the proximal method (Combettes and Pesquet, 2011; Bertsekas, 2015) to solve Equation (10) for the main reason of its simplicity and efficiency. In particular, we first consider the fidelity term $f(X, W) = \left\| W - X^T X \right\|^2_F$ in Equation (10), which is differentiable, and its gradient is $\nabla f(X, W) = 2(W - X^T X)$. As a result, it is easy to get the following gradient descent step:

$$
W_k = W_{k-1} - \alpha_k \nabla f(X, W_{k-1}),
$$

(11)

where $\alpha_k$ denotes the step size of the gradient descent.

Then, according to Combettes and Pesquet (2011) and the definition Data Acquisition therein, the proximal operator of $L_1$-norm regularizer on $W$ can be given as the following soft-threshold operation:

$$
\text{proximal}_{\lambda, \text{||} \cdot \text{||}_1}(W) = [\text{sgn}(W_i) \times \max(\text{abs}(W_i) - \lambda), 0]_{p \times p}.
$$

(12)

Finally, we use the proximal operation $\text{proximal}_{\lambda, \text{||} \cdot \text{||}_1}$ in Equation (12) on $W$ to keep $W$ in the “feasible region” (regularized by the $L_1$-norm) after each gradient descent step. Consequently, we get a simple algorithm for solving Equation (10) as shown in Table 2.
Method II: Remodeling PC-Based FBN with a Scale-Free Prior

It is well known that a brain network has more topological structures than just sparsity (Sporns, 2011) such as, modularity (Qiao et al., 2016), hierarchy (Zhou et al., 2006), small-worldness (Watts and Strogatz, 1998; Achard et al., 2006), clustering (White et al., 1986), degeneracy (Tononi et al., 1999), and scale-free (Eguíluz et al., 2005; Li et al., 2005). In order to verify the flexibility of the proposed framework in Equation (9), we develop a new PC-based FBN estimation model by incorporating a scale-free prior. Consequently, we have the following optimization model (namely PC\textsubscript{scale-free}):

$$\min_W \|W - X^TX\|_F^2 + \lambda \sum_{ij = 1}^n \gamma_{ij} |W_{ij}|. \quad (13)$$

Similar to the PC\textsubscript{sparsity} in Equation (10), \( \lambda \) is the regularized parameter. In order to incorporate the node degree information, a weight \( \gamma_{ij} \) related to the node degree of each \( W_{ij} \) is introduced in the PC\textsubscript{scale-free} model, which essentially makes the PC\textsubscript{scale-free} be a weighted version of PC\textsubscript{sparsity}. We argue that such a weighted extension can achieve a scale-free network by assigning the weight \( \gamma_{ij} \) properly as discussed below.

Note that the fidelity term \( f(X, W) = \|W - X^TX\|_F^2 \) of Equation (13) is the same as the one in Equation (10). Thus, the two problems share the same gradient descent step as shown in Equation (11). Then, we consider the regularized parameter \( \lambda \sum_{ij = 1}^n \gamma_{ij} |W_{ij}| \). Based on the definition of the proximal operation, we can easily get the proximal operator of the weighted L\(_1\)-norm regularizer on \( W \) as follows:

$$\text{proximal}_{\gamma_{ij}} \|\cdot\|_1 (W) = \left[ \text{sgn} (W_{ij}) \times \max (abs (W_{ij}) - \lambda \times \gamma_{ij}), 0 \right]_{p \times p}, \quad (14)$$

which is exactly a weighted version of the soft threshold operation. Since the node degree of the brain network tends to follow the power law distribution (Barabási and Bonabeau, 2003; Eguíluz et al., 2005; Cecchi et al., 2007; Lin and Ihler, 2011), we assume that the hub nodes cover more useful connections (closely related to the neural disorders), while the non-hub nodes cover weak or noisy connections. Therefore, compared with the PC\textsubscript{sparsity} method that equally treats each edge (or link) of the FBN, the PC\textsubscript{scale-free} method penalizes more on the nodes with small degree, and penalizes less on the nodes with large degree. According to Equation (14), a big \( \gamma_{ij} \) may increase the possibility that \( W_{ij} \) shrinks to zero, which in turn tend to result in a sparse vector \( W_i = (W_{i1}, W_{i2}, \cdots, W_{ip}) \), and then a small degree of node \( i \). Conversely, a small \( \gamma_{ij} \) may result in a big degree of node \( i \). In other words, the parameter \( \gamma_{ij} \) should have an inverse relationship with the node degree (Peng et al., 2009; Lin and Ihler, 2011). Thus, we assume that \( \gamma_{ij} \) has the following form:

$$\gamma_{ij} = e^{-\left(\frac{1}{\beta} \frac{\|W_{ij}\|_1}{\sum_{j=1}^p \|W_{ij}\|_1} + \frac{1}{\sum_{i=1}^n \|W_{ij}\|_1} \right)}, \quad (15)$$

where \( \varepsilon \) is a small number for preventing the denominator in Equation (15) to be zero. In our experiment, we simply set \( \varepsilon = 0.0001 \). As a consequence, we get the following alternating optimization algorithm. In each iteration, with a fixed \( W \), the parameter \( \gamma_{ij} \) can be easily calculated by Equation (15), and then by fixing the value of each parameter \( \gamma_{ij} \), we update \( W \) by solving Equation (13). We summarize the algorithm for solving Equation (13) in the following Table 3.

### Table 3: Algorithm of PC-based FBN estimation with a scale-free prior.

| Input: \( X \) //observed data |
|---------------------------------|
| Output: \( W \) //functional brain network |
| Initialize: \( W \); |
| while not converge |
| \( W \leftarrow W - \alpha(W - X^TX); \)
| \( W \leftarrow \text{proximal}_{\gamma_{ij}} (W) = [\text{sgn} (W_{ij}) \times \max (abs (W_{ij}) - \lambda \times \gamma_{ij}), 0]_{p \times p}; \) |
| end |

#### Experimental Setting

After obtaining the FBNs of all subjects, the main task comes to use the constructed FBNs to train a classifier for identifying ASDs from NCs. Since the FBN matrix is symmetric, we just use the upper triangular connections for training the classification. We then conduct a feature filtering operation before training the classification. Specifically, the classification pipeline includes the following two main steps.

#### Step 1: FBN construction based on PC\textsubscript{threshold}, SR, PC\textsubscript{sparsity}, and PC\textsubscript{scale-free} respectively

Note that each FBN construction method involves a free parameter, e.g., the threshold parameter in PC\textsubscript{threshold} and the regularized parameter in the other methods. Therefore, in this step, we construct multiple FBNs based on different parametric values, and then select the optimal FBN (for each method) based on a separate parameter selection procedure, as shown in Figure 1.

1In order to improve the flexibility of PC and conduct fair comparison, we introduce a hard-thresholding parameter in PC by reducing a proportion of weak connections.
> *Step 2: Feature selection and classification using t-test (with $p < 0.05$) and linear SVM (with default parameter $C = 1$), respectively. As pointed out in Wee et al. (2014), both the feature selection and classifier design have a big influence on the final accuracy. However, in this paper, we only adopt the simplest feature selection method and the most popular used SVM classifier (Chang and Lin, 2007), since our main focus is FBN estimation. In other words, it would be difficult to conclude whether the FBN construction methods or the feature selection/classification methods contribute to the ultimate performance.

The detailed experimental procedure (including a subprocedure for parameter selection) is shown in Figure 1. Due to the small sample size, we use the leave one out (LOO) cross validation strategy to verify the performance of the methods, in which only one subject is left out for testing while the others are used to train the models and get the optimal parameters. For the choice of the optimal parameters, an inner LOO cross-validation is further conducted on the training data by grid-search strategy. More specifically, for the regularized parameter $\lambda$, the candidate values range in $[0.05, 0.1, \cdots, 0.95, 1]$; for the hard threshold of $PC_{\text{threshold}}$, we use 20 sparsity levels ranging in
RESULTS

Network Visualization

For visual comparison of the FBN constructed by PC\_threshold, SR, PC\_sparsity and PC\_scale-free methods, we first show the FBN adjacency matrices\(^3\) \(W\) constructed by different methods in Figure 2.

It can be observed from Figure 2 that both (Figure 2B) PC\_threshold and (Figure 2C) PC\_sparsity can remove the noisy or weak connections from the dense FBN constructed directly by the original PC. Moreover, the topology of the FBN estimated by PC\_sparsity is similar to that of PC\_threshold, because (1) both methods employ the same data-fidelity term, and (2) the sparsification strategy behind PC\_sparsity (i.e., the soft-thresholding scheme) is based on the result of PC\_threshold (i.e., the hard-thresholding scheme). In contrast, the FBN constructed by SR has a topology highly different from those of PC\_threshold and PC\_sparsity, since it uses a different data-fidelity term (i.e., the first term in Equation (4)). More interestingly, compared with PC\_threshold and PC\_sparsity, the FBN estimated by (Figure 2E) PC\_scale-free has a clearer hub structure, due to the use of a weighted \(L_1\)-norm regularizer.

For showing the hub structure more clearly, we plot the brain connections estimated by PC\_scale-free in Figure 3, where the width of each arc represents the weight of the connection between two endpoints. Furthermore, we color the connections from the hub nodes, while showing other connections in gray for better visualization. In Figure 3, it can be interestingly observed that (1) the hub nodes are only a small proportion of the whole brain regions, illustrating the scale-free characteristic of the constructed FBN; (2) the hub nodes mainly locate at the brain regions, including the Cerebellum, Frontal, Rolandic, and Lingual, etc.

In order to visualize the relationship between the parameter \(\lambda\) in the PC\_scale-free model and the node degree, we simply count the number of the nodes from all participants in this dataset based on different node degree, and plot its cCDF (complementary cumulative distribution function) under log-log coordinates. The distribution of node degree results based on different values of parameter \(\lambda\) are shown in Figure 4.

Based on the results in Figure 4, we can find that, with the increase of the parametric value, the node degree distribution tends to be more scale-free.

For verifying the effectiveness of the regularizer and quantifying the scale-free topology of FBN constructed by PC\_scale-free and PC\_sparsity, we employ the s-metric (Li et al., 2005)

---

\(^3\)The adjacency matrix is an algebraic expression of a graph (or network). The elements of the matrix indicate the connection strength of the node pairs in the graph. Here, for the convenience of comparison among different methods, all the weights are normalized to the interval [-1 1].
to compute the corresponding scale-free measures.

$$S(W) = \sum d_i d_j,$$  
(16)

where $d_i$ means the degree of the node $i$, and $S(W)$ is the value of the s-metric for the network $W$. Since the s-metric relies on the number of the connects in FBN, and the network threshold affects the degree and scale-free measures significantly for these two methods. In this paper, as an example, we construct the FBN by $\text{PC}_{\text{scale\_free}}$ ($\lambda = 0.5$), and then find the FBN constructed by $\text{PC}_{\text{sparsity}}$ with the same number of connects. Based on Equation (14), the s-metric of the FBN constructed by $\text{PC}_{\text{sparsity}}$ is 18313091, and the one by $\text{PC}_{\text{scale\_free}}$ is 27862470. We note that the $\text{PC}_{\text{scale\_free}}$ has a higher s-metric value than $\text{PC}_{\text{sparsity}}$. Since the high s-metric value is achieved by connecting high degree nodes to each other, the FBN constructed by $\text{PC}_{\text{scale\_free}}$ can obtain more hub-nodes than the FBN constructed by $\text{PC}_{\text{sparsity}}$. Thus, the brain network constructed by $\text{PC}_{\text{scale\_free}}$ tends to be more “scale-free”.

**ASD Identification**

The ASD vs. NC classification results on ABIDE dataset are given in Table 4. The remodeling method ($\text{PC}_{\text{scale\_free}}$) achieves the best accuracy in this experiment. In addition, the results of Wee et al.’s method available from Wee et al. (2016) are also provided in Table 4 as a reference.

A set of quantitative measurements, including accuracy, sensitivity, and specificity, are used to evaluate the classification performance of four different methods ($\text{PC}_{\text{threshold}}$, SR, $\text{PC}_{\text{sparsity}}$ and $\text{PC}_{\text{scale\_free}}$). The mathematical definition of these three measures are given as follows:

$$\text{Accuracy} = \frac{\text{TruePositive} + \text{TrueNegative}}{\text{TruePositive} + \text{FalsePositive} + \text{TrueNegative} + \text{FalseNegative}},$$  
(17)

$$\text{Sensitivity} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}},$$  
(18)

$$\text{Specificity} = \frac{\text{TrueNegative}}{\text{TrueNegative} + \text{FalsePositive}}.$$  
(19)

Here, $\text{TruePositive}$ is the number of the positive subjects that are correctly classified in the ASD identification task. Similarly, $\text{TrueNegative}$, $\text{FalsePositive}$, and $\text{FalseNegative}$ are the numbers of their corresponding subjects, respectively.

**Sensitivity to Network Model Parameters**

The ultimate classification accuracy is particularly sensitive to the network model parameters. In Figure 5, we show the classification accuracy corresponding to different parametric combination (i.e., [0.05, 0.1, · · · , 0.95, 1] for $\text{PC}_{\text{sparsity}}$, and $\text{PC}_{\text{scale\_free}}$ [5%, 10%, · · · , 95%, 100%] for $\text{PC}_{\text{threshold}}$) in 4 different methods. In addition, the classification accuracy is computed by the LOO test on all of the subjects.

**DISCUSSION**

The FBN commonly has more “structures” than just sparsity (Smith et al., 2011; Sporns, 2011). In this work, we remodel the PC-based method into an optimization model for incorporating some of these structures such as, scale-free property. The proposed models were verified on the ABIDE dataset for ASD vs. NC classification. Based on the results, we give the following brief discussion.

1. The accuracy of the PC-based methods outperforms the SR method on our used dataset. A possible reason is that the SR implicitly involves an inverse operation on the covariance matrix, which tends to be ill-posed due to the limited sample size and high-dimensional features. In fact, a recent study (Qiao et al., 2016) also notes a similar problem that the performance of SR-based method drops significantly with the increase of the feature dimension. In contrast, the PC-based methods can be derived directly from the covariance matrix without the inverse operation, and thus works robustly and also generally scales well.

2. The performance of $\text{PC}_{\text{sparsity}}$ in our experiments is similar to that of the hard-threshold counterpart $\text{PC}_{\text{threshold}}$, because both methods share the same data-fidelity term and a similar sparsification scheme (i.e., hard threshold for $\text{PC}_{\text{threshold}}$ while soft threshold for $\text{PC}_{\text{sparsity}}$). The subtle difference of the results between these two methods may be due to the regularized parameters (e.g., hard threshold in PC and $\lambda$ in $\text{PC}_{\text{sparsity}}$). However, we argue that the model of $\text{PC}_{\text{sparsity}}$ is more flexible than $\text{PC}_{\text{threshold}}$. For example, it can be naturally extended to a weighted version, namely $\text{PC}_{\text{scale\_free}}$, for better performance.
Figure 4 | The distribution of node degree and the corresponding cCDF under log-log coordinates with respect to different parameter $\lambda$. (A) $\lambda = 0$, (B) $\lambda = 0.5$, (C) $\lambda = 1$, (D) $\lambda = 0$ (under log-log coordinates), (E) $\lambda = 0.5$ (under log-log coordinates), and (F) $\lambda = 1$ (under log-log coordinates).

(3) The proposed PC$_{\text{scale-free}}$ method achieves the best classification accuracy among all the methods. In our opinion, this is mainly due to its power for modeling the hub node in a network that may cover the useful connections closely related to neural disorders. Interestingly, it outperforms Wee et al.’s method (Wee et al., 2016), which used the same NYU dataset, even though the latter employs more sophisticated feature selection and classification strategy. In addition, the proposed PC$_{\text{scale-free}}$ method provides an empirical evidence that a suitable biological/physical prior can be used to guide the estimation of better FBNs.

In addition, we further conduct experiments for verifying the effectiveness of the proposed methods on a non-ASD fMRI dataset from ADNI, and find that the PC$_{\text{scale-free}}$ methods still achieve the best accuracy. Since the main focus of this paper is on ASD identification, we supply the details of the dataset and experimental results in a Supplementary Material. The results
show that the proposed method tends to generalize well on both ASD and non-ASD datasets. In other words, the idea for estimating FBN in this paper is general and independent of the used datasets. However, there are several limitations in the proposed methods that need to be improved in the future work.

(1) We use the $L_1$-norm (or weighed $L_1$-norm) as a regularizer to estimate sparse (or scale-free) FBNs for the subjects one by one. However, the FBNs of different subjects tend to share some similar structures (Wee et al., 2014; Yu et al., 2016) and thus the proposed method may lose such group information. Therefore, in the future work, we need to adopt the development and application of “group constraint” such as, Group LASSO (Yuan and Lin, 2006) for addressing this problem.

(2) In this paper, the ratio of male to female participants is substantially 5 to 1. According to a recent finding (Lai et al., 2013), the gender is one of the obvious sources of heterogeneity in ASD. Therefore, in the future work, we need to consider this issue for reducing the effect of heterogeneity.

CONCLUSION

Pearson’s correlation is the most commonly used scheme in estimating FBNs due to its simplicity, efficiency and robustness. However, the PC scheme is inflexible due to the difficulty of incorporating informative priors. In this paper, we remodel the PC into an optimization framework, based on which the biological priors or assumptions can be naturally introduced in the form of regularizers. More specifically, based on this framework, we propose two PC-based FBN estimation methods, namely $PC_{\text{sparsity}}$ and $PC_{\text{scale-free}}$, which can effectively encode sparse and scale-free priors, respectively. Finally, we use these constructed FBNs to classify the ASDs from NCs, and get an 81.52% accuracy, which outperforms the baseline and state-of-the-art methods. On the other hand, the topology of FBN is much more than just the sparsity and scale-free. Therefore, it is a potentially valuable topic to incorporate other biological/physical priors in constructing FBNs.

AUTHOR CONTRIBUTIONS

All authors developed remodeling algorithm, architecture. WL, ZW, LZ, and LQ designed the evaluation experiments. DS preprocessed the fMRI. All authors contributed to preparation of the article, figures, and charts.

ACKNOWLEDGMENTS

This work was partly supported by National Natural Science Foundation of China (61300154, 61402215, 11401060), Natural Science Foundation of Shandong Province (2014ZRB019E0, 2014ZRB019V9C), Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJ1500501, KJ1600512, KJ1600518), Chongqing Graduate Student Research Innovation Project (CYS16183), and NIH grants (AG041721, MH107815, EB006733, EB008374, EB009634).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fninf.2017.00055/full#supplementary-material

REFERENCES

Achard, S., Salvador, R., Whitcher, B., Buckner, L., Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. J. Neurosci. 26, 63–72. doi: 10.1523/JNEUROSCI.3874-05.2006

Allen, G., and Courchesne, E. (2003). Differential effects of developmental cerebellar abnormality on cognitive and motor functions in the cerebellum: an fMRI study of autism. Am. J. Psychiatry 160, 262–273. doi: 10.1176/appi.ajp.160.2.262

Anderson, J. S., Druzel, T. J., Froehlich, A., Dubray, M. B., Lange, N., Alexander, A. L., et al. (2011). Decreased interhemispheric functional connectivity in autism. Cereb. Cortex 21, 1134–1146. doi: 10.1093/cercor/bhq190

Baio, J. (2014). Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010. MMWR 63, 1–21.

Barabási, A. L., and Bonabeau, E. (2003). Scale-free networks. Sci. Am. 288, 60–69. doi: 10.1038/scientificamerican0503-60

Bertsekas, D. P. (2015). Incremental gradient, subgradient, and proximal methods for convex optimization: a survey. Optimization 2, 691–717.
Li, L., Alderson, D., Doyle, J. C., and Willinger, W. (2005). Towards a theory of scale-free graph's definition, properties, and implications. Internet Math. 2, 431–523. doi: 10.1080/154279505.2001.1012911

Li, F., Wee, C. Y., Chen, H., and Shen, D. (2012). Inter-modality relationship constrained multi-task feature selection for AD/VD classification. Med. Image Comput. Comput. Assist. Interv. 16, 308–315.

Lin, Q., and Iliner, A. T. (2011). Learning scale free networks by reweighted L1 regularization. J. Mach. Learn. Res. 15, 40–48.

Lo, Y. C., Soong, W. T., Gau, S. F., Wu, Y. Y., Lai, M. C., Yeh, F. C., et al. (2011). The loss of asymmetry and reduced interhemispheric connectivity in adolescents with autism: a study using diffusion spectrum imaging tractography. Psychiatry Res. 192, 60–66. doi: 10.1016/j.psychres.2010.09.008

Lord, C., Cook, E. H., Leventhal, L. B., and Aamal, D. G. (2000). Autism spectrum disorders. Neuron 21, 86–92. doi: 10.1016/S0896-6273(00)00115-X

Lord, C., and Jones, R. M. (2012). Annual research review: re-thinking the classification of autism spectrum disorders. J. Child Psychol. Psychiatry Allied Discipl. 53, 490–509. doi: 10.1111/j.1469-7610.2012.02547.x

Mason, M. F., Norton, M. I., Horn, J. D. V., Wegner, D. M., Grafton, S. T., and Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. Science 315, 393–395. doi: 10.1126/science.1113195

Meinsma, N., and Buhlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. Ann. Stat. 34, 1436–1462. doi: 10.1214/009053606000000281

Nielson, J. A., Zielinski, B. A., Fletcher, P. T., Alexander, A. L., Lange, N., Bigler, E. D., et al. (2013). Multisite functional connectivity MRI classification of autism: ABIDE results. Front. Hum. Neurosci. 7:599. doi: 10.3389/fnhum.2013.00599

O’Roak, B. J., Vives, L., Fu, W., Egertson, J. D., Stanaway, I. B., Phelps, I. G., et al. (2012). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 338, 1619–1622. doi: 10.1126/science.1217764

Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. J. Am. Stat. Assoc. 104, 735–746. doi: 10.1198/jasa.2009.0126

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2013). Methods to detect, characterize, and remove motion artifact in resting state fMRI. NeuroImage 84, 320–341. doi: 10.1016/j.neuroimage.2013.08.048

Qiao, L., Zhang, H., Kim, M., Teng, S., Zhang, L., and Shen, D. (2016). Estimating functional brain networks by incorporating a modularity prior. NeuroImage 141, 399–407. doi: 10.1016/j.neuroimage.2016.07.058

Rosa, M. J., Portugal, L., Hahn, T., Fallgatter, A. J., Garrido, M. I., and Shawe-Taylor, J., et al. (2015). Sparse network-based models for patient classification using fMRI. NeuroImage 105, 493–506. doi: 10.1016/j.neuroimage.2014.11.021

Segal, D. L. (2013). “Diagnostic and statistical manual of mental disorders (DSM-IV-TR)” in Corrioni Encyclopedia of Psychology, eds I. Weiner and W. E. Craighead (Hoboken NJ: John Wiley & Sons, Inc.), 495–497.

Shen, Y., Fang, J., and Li, H. (2013). Exact reconstruction analysis of log-sum network recovery under compressed sensing. IEEE Signal Process. Lett. 20, 1223–1226. doi: 10.1109/LSP.2013.2285579

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., et al. (2011). Network modelling methods for fMRI. NeuroImage 54, 875–891. doi: 10.1016/j.neuroimage.2010.08.063

Smith, S. M., Vidaurre, D., Beckmann, C. F., Glasser, M. F., Jenkinson, M., Miller, K. L., et al. (2013). Functional connectomics from resting-state fMRI. Trends Cogn. Sci. 17, 666–682. doi: 10.1016/j.tics.2013.09.016

Sporn, O. (2011). Networks of the Brain. Cambridge, MA: MIT Press.

Stam, C. J. (2014). Modern network science of neurological disorders. Nature Rev. Neurosci. 15, 683–695. doi: 10.1038/nrn3801

Supek, K., Menon, V., Rubin, D., Musen, M., and Greicius, M. D. (2008). Network analysis of intrinsic functional brain connectivity in Alzheimer's Disease. PLoS Comput. Biol. 4:e1000100. doi: 10.3389/jcnisci.1000100

Theije, C. G. M. D., Wu, J., Silva, S. L. D., Kamphuis, P. J., Garsen, J., Korte, S. M., et al. (2011). Pathways underlying the gut-to-brain connection in autism spectrum disorders as future targets for disease management. Eur. J. Pharmacol. 668(Suppl 1), 570–580. doi: 10.1016/j.ejphar.2017.07.013

Tomikawa, A., and Sugiyama, M. (2009). Dual augmented lagrangian method for efficient sparse reconstruction. IEEE Signal Process. Lett. 16, 1067–1070. doi: 10.1109/LSP.2009.2030111
Tononi, G., Sporns, O., and Edelman, G. M. (1999). Measures of degeneracy and redundancy in biological networks. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3257–3262. doi: 10.1073/pnas.96.6.3257

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978

Wang, K., Zhang, H., Ma, D., Bucan, M., Glessner, J. T., Abrahams, B. S., et al. (2009). Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 459, 528–533. doi: 10.1038/nature07999

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature* 393, 440–442. doi: 10.1038/30918

Wee, C. Y., Yap, P. T., and Shen, D. (2012). Identification of MCI individuals using structural and functional connectivity networks. *Neuroimage* 59, 2045–2056. doi: 10.1016/j.neuroimage.2011.10.015

Wee, C. Y., Yap, P. T., Zhang, D., Wang, L., and Shen, D. (2014). Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification. *Brain Struct. Funct.* 219, 641–656. doi: 10.1007/s00429-013-0524-8

White, J. G., Southgate, E., Thomson, J. N., and Brenner, S. (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond.* 314, 1–340. doi: 10.1098/rstb.1986.0056

Yan, C. G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R. C., Martino, A. D., et al. (2013). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *Neuroimage* 76, 183–201. doi: 10.1016/j.neuroimage.2013.03.004

Yu, R., Zhang, H., An, L., Chen, X., Wei, Z., and Shen, D. (2016). "Correlation-weighted sparse group representation for brain network construction in MCI classification" in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* (Springer International Publishing).

Yuan, M., and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc.* 68, 49–67. doi: 10.1111/j.1467-9868.2005.00532.x

Zhao, Y. B. (2013). New and improved conditions for uniqueness of sparsest solutions of underdetermined linear systems. *Appl. Mathem. Comput.* 224, 58–73. doi: 10.1016/j.amc.2013.08.010

Zhou, C., Zemanová, L., Zamora, G., Hilgetag, C. C., and Kurths, J. (2006). Hierarchical organization unveiled by functional connectivity in complex brain networks. *Phys. Rev. Lett.* 97:238103. doi: 10.1103/PhysRevLett.97.238103

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Li, Wang, Qiao and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.