



Estimating High-Order Brain Functional Networks in Bayesian View for Autism Spectrum Disorder Identification

Xiao Jiang^{1,2†}, Yueying Zhou^{3†}, Yining Zhang¹, Limei Zhang^{1,4}, Lishan Qiao^{1,4*} and Renato De Leone^{2*}

¹ School of Mathematics Science, Liaocheng University, Liaocheng, China, ² School of Science and Technology, University of Camerino, Camerino, Italy, ³ College of Computer Science and Technology, Nanjing University of Aeronautics, Nanjing, China, ⁴ School of Computer Science and Technology, Shandong Jianzhu University, Jinan, China

OPEN ACCESS

Edited by:

Zhengxia Wang,
Hainan University, China

Reviewed by:

Benzheng Wei,
Shandong University of Traditional
Chinese Medicine, China

Linling Li,
Shenzhen University, China
Jun Wang,
Shanghai University, China

*Correspondence:

Lishan Qiao
qiaolishan@lcu.edu.cn
Renato De Leone
renato.deleone@unicam.it

[†] These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 10 February 2022

Accepted: 31 March 2022

Published: 27 April 2022

Citation:

Jiang X, Zhou Y, Zhang Y,
Zhang L, Qiao L and De Leone R
(2022) Estimating High-Order Brain
Functional Networks in Bayesian View
for Autism Spectrum Disorder
Identification.
Front. Neurosci. 16:872848.
doi: 10.3389/fnins.2022.872848

Brain functional network (BFN) has become an increasingly important tool to understand the inherent organization of the brain and explore informative biomarkers of neurological disorders. Pearson's correlation (PC) is the most widely accepted method for constructing BFNs and provides a basis for designing new BFN estimation schemes. Particularly, a recent study proposes to use two sequential PC operations, namely, correlation's correlation (CC), for constructing the high-order BFN. Despite its empirical effectiveness in identifying neurological disorders and detecting subtle changes of connections in different subject groups, CC is defined intuitively without a solid and sustainable theoretical foundation. For understanding CC more rigorously and providing a systematic BFN learning framework, in this paper, we reformulate it in the Bayesian view with a prior of matrix-variate normal distribution. As a result, we obtain a probabilistic explanation of CC. In addition, we develop a Bayesian high-order method (BHM) to automatically and simultaneously estimate the high- and low-order BFN based on the probabilistic framework. An efficient optimization algorithm is also proposed. Finally, we evaluate BHM in identifying subjects with autism spectrum disorder (ASD) from typical controls based on the estimated BFNs. Experimental results suggest that the automatically learned high- and low-order BFNs yield a superior performance over the artificially defined BFNs via conventional CC and PC.

Keywords: brain functional network, high-order network, Pearson's correlation, Bayesian statistics, matrix-variate normal distribution, autism spectrum disorder

INTRODUCTION

Resting-state functional magnetic resonance imaging (rs-fMRI)-based brain functional network (BFN) analysis without a specific task, has shown a great potential to discover biomarkers for identifying neurological/mental disorders, such as autism spectrum disorder (ASD) (Wang et al., 2019), major depressive disorder (MDD) (Long et al., 2020), schizophrenia (Ariana and Cohen, 2013), Parkinson's disease (PD) (Baggio et al., 2014), Alzheimer's disease (AD) (Hahn et al., 2013), and its early stage, namely, mild cognitive impairment (MCI) (Jiang et al., 2019). However,

the identification of brain disorders based on the BFN remains a critical challenge, since its great performance depends on multiple interactive factors including reasonable brain parcellation, well-parametrized network estimation, discriminative feature selection/extraction, and powerful classifier design (Dadi et al., 2019; Pervaiz et al., 2020). Instead of considering all these aspects that have been empirically evaluated in recent studies (Dadi et al., 2019; Pervaiz et al., 2020), in this paper, we mainly focus on the BFN estimation issue. Recently, more advanced studies (Yu et al., 2017; Zhang et al., 2017; Mahjoub et al., 2018) have proposed the brain functional connectivity representations for estimating BFN at different connectivity levels, including low-order, high-order, etc. Low-order methods are designed to characterize the synchronization of blood oxygen level dependent (BOLD) signals and are insufficient to characterize a high level of interaction. Recent literature (Chen et al., 2016; Zhang et al., 2016) presented high-order methods to measure the relationship between the BFN connectivity. This paper specifically aims to capture the brain connectivity that is supposed to exist in a higher-order form.

Owing to its non-invasiveness and easy reproducibility, rs-fMRI (Wang et al., 2020) has become a widely used technique to estimate BFN whose nodes correspond to spatial regions of interest (ROIs) and edges describe the relationship (e.g., similarity, correlation, synchronism, etc.) between the rs-fMRI signals associated with these ROIs. In the past decades, researchers have developed many BFN estimation methods, including Pearson's correlation (PC) (Biswal et al., 1995; Eguiluz et al., 2005), partial correlation (Marrelec et al., 2006), regularized full/partial correlation (Friedman et al., 2008; Jie et al., 2009; Li et al., 2017), structural equation modeling (McIntosh and Gonzalez-Lima, 1994), and dynamic causal modeling (Friston et al., 2003), etc. According to a recent comparative study (Smith et al., 2013), the correlation-based approaches are "quite successful" for estimating informative BFNs. Particularly, PC is the fundamental and most widely used correlation-based method for BFN estimation. Despite its empirical effectiveness, PC only considers a pair of ROIs at a time, and thus suffers the confounding effect from other ROIs. Partial correlation can tackle this problem by regressing out the confounding variables. However, that may lead to an ill-posed estimation since the partial correlation is usually calculated by inverting a covariance matrix that may be singular. In practice, a regularizer is generally introduced into the partial correlation model, which not only deals with the ill-posed problem but also provides a natural way to introduce topological priors of the brain network into the estimation models. Specifically, L_1 -norm is commonly used to encode the sparsity prior of the BFN (Lee et al., 2011), a weighted version of the L_1 -norm to capture the hub structure (prior) (Li et al., 2017), the $L_{2,1}$ -norm to model group sparsity (or population prior) that imposes all the subjects share the same BFN topology (Zhang, 2010), and a combination of L_1 -norm with trace-norm to encode the modularity (prior) of the BFN (Qiao et al., 2016), to just name a few.

No matter which prior or regularizer is introduced, most of the correlation-based methods only estimate low-order BFNs whose edges are the full or partial correlation of the rs-fMRI time series.

Beyond these traditional low-order correlations, researchers found that some forms of high-order correlations may contain useful feature information for BFN analysis and classification (Chen et al., 2016; Zhang et al., 2016; Zhou et al., 2018a,b). For example, Chen et al. (2016) defined the high-order correlation as the dependency between functional connectivity fluctuation, with clustered mean correlation time series as input. Different from characterizing a temporal correlation, Zhang et al. (2016) proposed to construct the high-order BFN to examine spatial properties of the functional connectivity network. Specifically, such a scheme is achieved by two sequential PC operations, where the first PC operation is used to construct a traditional low-order BFN, and the ensuing PC operation is conducted on the edge weights of the estimated BFN to generate the high-order BFN. Despite encoding the network information from different dimensions, the above methods are uniformly called correlation's correlation (CC) (Zhang et al., 2017) since they both involve two PC operations in the high-order BFN construction. However, the CC-based high-order BFNs are estimated intuitively and heuristically without the support of any strong theoretical basis.

Toward a better understanding of CC, in this paper, we reformulate it in the Bayesian framework with a prior that the low-order BFN follows the matrix-variate normal (MVN) distribution. As a result, we obtain a probabilistic explanation for CC and develop a new method that both learns low- and high-order BFNs from data based on the rigorous theoretical framework. In brief, we summarize the main contributions of this paper as follows.

1. We reformulate PC from a statistical point of view. Based on this, a regularized statistical framework is derived by introducing Gaussian distribution to the error term, which provides a more flexible modeling idea.
2. A mathematical model for a high-order learning method based on CC is developed by assuming the adjacency matrix of low-order brain networks follows an aprior normal distribution.
3. Based on the probabilistic framework derived above, an automatic learning model, namely, BHM, is proposed. Compared with the traditional high-order network learning method (i.e., CC), the model simultaneously learns low-order and high-order brain networks. In the learning process, the direct information of the low-order network and the indirect information of the high-order network complement each other toward more reliable/discriminative brain networks.
4. Finally, we empirically verify that the automatically learned BFNs outperform the artificially defined ones *via* CC and other baselines in the identification of ASD, even with a simple feature selection method and classifier.

For a consistent expression throughout the paper, we first describe the basic notations as follows. Scalars involving the variables, parameters, and constants are denoted by *italic* lowercase letters, e.g., x . Vectors are denoted by bold lowercase letters and the elements inside are stored in a column, e.g., $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$. Matrices are denoted by bold uppercase letters such as \mathbf{X} .

The rest of the paper is organized as follows. In Section “Related Works,” we review the related works including PC, sparse representation (SR), and CC. In Section “High-Order Correlation Learning,” we first introduce a theoretical framework for explaining CC and then develop a new framework for learning high-order BFN by reformatting CC in a view of the maximal posterior probability. In Section “Experiments and Results,” we conduct experiments to evaluate the discrimination of the automatically learned high-order BFNs. In Section “Discussions,” we discuss the main findings. Finally, the conclusion are reported in Section “Conclusion.”

RELATED WORKS

In this section, we review three related works: PC, SR, and CC. As discussed previously, PC and SR are used to construct the traditional **low-order** BFNs, while CC as a two-step sequential PC method is used to estimate the **high-order** BFNs.

Pearson’s Correlation

Suppose \mathbf{x}_i is the multivariate random variable (random vector) associated with the i^{th} ROI. Then, the observed rs-fMRI signals¹ $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{ni})^T, i = 1, 2, \dots, p$ can be considered as a sampling of the multivariate random variable (or population) \mathbf{x}_i , where p is the number of ROIs and n is the number of time points. Since our goal is to estimate the edge weights of the BFN, the simplest and empirically effective way is to calculate the sample PC coefficient $\hat{w}_{ij} (i, j = 1, 2, \dots, p)$ of pair-wise ROIs, as follows.

$$\hat{w}_{ij} = \frac{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^T (\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\sqrt{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^T (\mathbf{x}_i - \bar{\mathbf{x}}_i)} \sqrt{(\mathbf{x}_j - \bar{\mathbf{x}}_j)^T (\mathbf{x}_j - \bar{\mathbf{x}}_j)}} \quad (1)$$

where $\bar{\mathbf{x}}_i$ is the mean vector corresponding to \mathbf{x}_i . Under Gaussian assumption, Eq. 1 gives an asymptotically unbiased estimation for the population PC coefficient. Without loss of generality, we redefine $\mathbf{x}_i \triangleq (\mathbf{x}_i - \bar{\mathbf{x}}_i) / \sqrt{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^T (\mathbf{x}_i - \bar{\mathbf{x}}_i)}$. Then, the estimator of population PC coefficient can be simplified as follows:

$$\hat{w}_{ij} = \mathbf{x}_i^T \mathbf{x}_j \text{ or } \hat{\mathbf{W}} = \mathbf{X}^T \mathbf{X} \quad (2)$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p]$ is the rs-fMRI data matrix whose columns are the rs-fMRI time series associated with different ROIs. Therein, $\hat{\mathbf{W}}$ is the generalized estimator in matrix form.

Sparse Representation

Sparse representation is one of the commonly used methods for calculating partial correlation among ROIs. A regularization term encoding sparsity prior is introduced into the BFN

estimation model. Specifically, the mathematical model of SR is given by:

$$\min_{\mathbf{W}} \sum_{i=1}^p \left\| \mathbf{x}_i - \sum_{j \neq i} w_{ij} \mathbf{x}_j \right\|^2 + \lambda \sum_{j \neq i} |w_{ij}| \quad (3)$$

where \mathbf{W} is the edge weight matrix of BFN.

Similar to PC, we can rewrite SR in matrix form:

$$\begin{aligned} \min_{\mathbf{W}} \|\mathbf{X} - \mathbf{X}\mathbf{W}\|_F^2 &= \lambda \|\mathbf{W}\|_1 \\ \text{s.t. } w_{ii} &= 0, \forall i = 1, 2, \dots, p \end{aligned} \quad (4)$$

where the constraint $w_{ii} = 0$ plays a role in removing \mathbf{x}_i from \mathbf{X} to avoid the trivial solution.

Correlation’s Correlation

Despite its popularity and effectiveness, the traditional PC can only construct the low-order BFN. That is, the connection between two ROIs is determined by the correlation of the corresponding rs-fMRI time series. However, in practice, a connection can be described in *both* low-order *and* high-order views. For example, we can directly define a connection between ROI i and j if there is a relationship. Besides, if ROIs i and j are connected to the same brain region, we can infer with a great possibility that there is a connection between ROI i and j . The former corresponds to the correlation in the traditional low-order view, while the latter can be considered as CC in a high-order perspective. This results can be achieved by a two-step procedure. First, the low-order BFN is estimated *via* PC. According to the formula in Eqs 1 or 2, the adjacency matrix $\hat{\mathbf{W}} = (\hat{w}_{ij})_{p \times p}$ of the PC-based BFN can be calculated as follows.

$$\hat{h}_{ij} = \hat{\mathbf{w}}_i^T \hat{\mathbf{w}}_j \quad (5)$$

where $\hat{\mathbf{w}}_i$ and $\hat{\mathbf{w}}_j$ are the i^{th} and j^{th} columns of $\hat{\mathbf{W}}$, respectively. For simplicity, in Eq. 5, $\hat{\mathbf{w}}_i$ and $\hat{\mathbf{w}}_j$ has been centralized and normalized as the case in Eq. 2. As a result, the CC-based high-order BFN is defined as follows,

$$\hat{\mathbf{H}} = (\hat{h}_{ij})_{p \times p} = \hat{\mathbf{W}}^T \hat{\mathbf{W}} \quad (6)$$

HIGH-ORDER CORRELATION LEARNING

As described previously, CC constructs the high-order BFN based on two sequential correlation operations. Despite its empirical effectiveness in identifying neuro-disorders (Chen et al., 2016; Zhang et al., 2016), CC is a measure defined intuitively without a clear mathematical/probabilistic explanation. Therefore, in this section, we will construct a more rigorous mathematical model for CC, in order to provide a better understanding of the CC-based high-order BFN.

Since CC is based on the PC variant, in the following Section “Pearson’s Correlation-Based Brain Functional Network Learning Framework in Bayesian View,” we first reformulate PC

¹Note that the rs-fMRI signals have been preprocessed as described in Section “Data Acquisitions and Processing.”

into a more flexible BFN estimation framework. Then, based on the framework, we establish a theoretical model for CC in Section “Learning High-Order Brain Functional Network With a Matrix-Normal Penalty.” Finally, we design an algorithm for learning the high-order BFN based on the theoretical model of CC in Section “Algorithm”.

Pearson’s Correlation-Based Brain Functional Network Learning Framework in Bayesian View

As we know, PC is a measure of the linear correlation between pair-wise rs-fMRI time series associated with the ROIs. In other words, a time series $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{ni})^T$ can be linearly represented by other time series $\mathbf{x}_j = (x_{1j}, x_{2j}, \dots, x_{nj})^T$ as

$$x_i = a_{ij}x_j + \varepsilon_i \tag{7}$$

where a_{ij} is the representative coefficient, and $\varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i}, \dots, \varepsilon_{ni})^T$ is the random error vector. That is, for each variable, we have:

$$x_{ki} = a_{ij}x_{kj} = \varepsilon_{ki} \quad (k = 1, 2, \dots, n; i, j = 1, \dots, p) \tag{8}$$

Generally, we assume that the random variable ε_{ki} follows a normal distribution with mathematical expectation 0, i.e., $\varepsilon_{ki} \sim \mathcal{N}(0, \sigma^2)$. Therefore, given x_{kj} and a_{ij} are constants, x_{ki} follows the normal distribution $x_{ki} \sim \mathcal{N}(a_{ij}x_{kj}, \sigma^2)$. Then, The following formula can be obtained by the maximum likelihood estimation of x_{ki} (see **Appendix A** for details).

$$\max_{a_{ij}} -n \log(\sqrt{2\pi}\sigma) - \frac{\|\mathbf{x}_i - a_{ij}\mathbf{x}_j\|^2}{2\sigma^2} \tag{9}$$

Note that Eq. 9 can be equivalently written as the following least-squares problem:

$$\min_{a_{ij}} \|\mathbf{x}_i - a_{ij}\mathbf{x}_j\|^2 \tag{10}$$

The optimal solution to Problem (10) is given by $\hat{a}_{ij} = (\mathbf{x}_j^T \mathbf{x}_j)^{-1} \mathbf{x}_j^T \mathbf{x}_i = \mathbf{x}_i^T \mathbf{x}_j = \hat{w}_{ij}$, considering that all of time series $\mathbf{x}_i, i = 1, 2, \dots, p$ have been normalized by $\frac{\mathbf{x}_i}{\|\mathbf{x}_i\|}$. This means that the solution of Eqs 9, 10 is the same as PC shown in Eq. 2. Therefore, in what follows, we only use w_{ij} instead of a_{ij} for the consistency of mathematical notations.

To provide a more flexible framework for BFN estimation, we further generalize PC in Bayesian view by introducing a prior distribution on w_{ij} . Although various distributions can be used as the prior, here we first consider the standard normal distribution, i.e., $w_{ij} \sim \mathcal{N}(0, 1)$, since it provides a basis for understanding more complex cases. However, in practice, the entries w_{ij} in \mathbf{W} may not be apriori independent of each other, but exist a relationship. Due to $w_{ij} \sim \mathcal{N}(0, 1)$, the edge weight w_{ij} of BFN has the following prior probabilistic density:

$$P(w_{ij}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w_{ij}^2}{2}} \tag{11}$$

Next, a maximal posterior estimation of w_{ij} (see **Appendix B** for details) can be obtained as follows,

$$\max_{w_{ij}} -(n + 1) \log(\sqrt{2\pi}) = (-n) \log - \frac{\|\mathbf{x}_i - w_{ij}\mathbf{x}_j\|^2 + \sigma^2 w_{ij}^2}{2\sigma^2} \tag{12}$$

The above problem is equivalent to the regularized least-squares problem:

$$\min_{w_{ij}} \|\mathbf{x}_i - w_{ij}\mathbf{x}_j\|^2 + \lambda w_{ij}^2 \tag{13}$$

where $\lambda = \sigma^2$ in the case of standard normal distribution. In practice, $\lambda \triangleq \sigma^2/\sigma_0^2$ is a hyper-parameter that controls the balance between the two terms in Eq. 13, where σ_0^2 corresponds to the variance of normal distribution of the edge weight w_{ij} . Setting the gradient of the objective function to zero, we obtain the optimal solution of Eq. 13 as follows:

$$\hat{w}_{ij} = (\mathbf{x}_j^T \mathbf{x}_j + \lambda)^{-1} \mathbf{x}_j^T \mathbf{x}_i = (1 + \lambda)^{-1} \mathbf{x}_i^T \mathbf{x}_j \tag{14}$$

We find that it is a shrinkage of the original estimation of PC which helps remove the weak connections in BFN.

Note that Eq. 13 only considers finding one edge weight at a time. Without loss of generality, with the assumption that the variables w_{ij} in \mathbf{W} are independent, we can rewrite Eq. 13 in the following matrix form:

$$\min_{\mathbf{W}} \|\mathbf{W} - \mathbf{X}^T \mathbf{X}\|_F^2 + \lambda \text{tr}(\mathbf{W}\mathbf{W}^T) \tag{15}$$

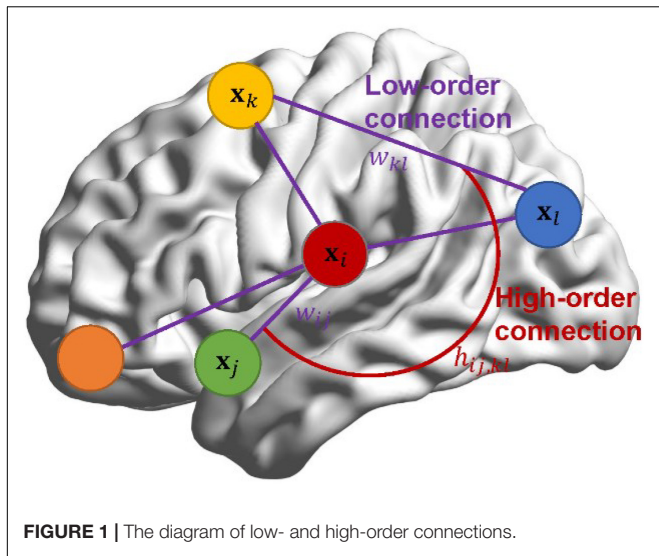
where $\text{tr}(\mathbf{W}\mathbf{W}^T) = \sum_{i,j} w_{ij}^2$ is the trace operator of $\mathbf{W}\mathbf{W}^T$. As a result, we achieve the estimation of BFN in a batching way as follows,

$$\hat{\mathbf{W}} = (1 + \lambda)^{-1} \mathbf{X}^T \mathbf{X} \tag{16}$$

This formula is essentially a generalization of Eq. 2 with a shrinkage factor $(1 + \lambda)^{-1}$. When $\lambda \rightarrow 0$, Eq. 16 reduces to the traditional PC.

Learning High-Order Brain Functional Network With a Matrix-Normal Penalty

In Section “Pearson’s Correlation-Based Brain Functional Network Learning Framework in Bayesian View,” we reformulate PC and then generalize it in Bayesian view by introducing a standard normal prior $w_{ij} \sim \mathcal{N}(0, 1)$ for each pair of ROIs $(i, j), i, j = 1, 2, \dots, p$. However, in practice, the entries w_{ij} in \mathbf{W} may not be apriori independent of each other, but exist a relationship. Even so, Section “Pearson’s Correlation-Based Brain Functional Network Learning Framework in Bayesian View” provides a flexible probabilistic framework to develop new brain network estimation methods. Inspired by this point, we lay down theoretical support for CC from the Bayesian perspective. More importantly, instead of assuming that w_{ij} in \mathbf{W} are independent, we propose a Bayesian high-order model (BHM) for BFN estimation by introducing the prior



of matrix-variate normal distribution to the low-order BFN \mathbf{W} . BHM learns the high-order relationship from the data automatically, rather than manually define as the case in CC. Interestingly, such a scheme can simultaneously learn low-order and high-order BFNs by considering the spatial structure of network connections. To distinguish the low-order and high-order correlations, an illustration is shown in **Figure 1**. The connections among w_{ij} can be considered as a high-order correlation $h_{ij,kl}$, while w_{ij} denotes the traditional low-order linear correlation between ROIs.

Model

Since we can vectorize the low-order edge weight matrix \mathbf{W} into a $p^2 = 1$ vector, the low- and high- order correlation can be modeled by a multivariate normal distribution, $\text{vec}(\mathbf{W} \sim \mathcal{N}(\mathbf{O}, \Omega))$, where \mathbf{W} encodes the low-order relationship and $\Omega \in R^{p^2 \times p^2}$ is the covariance matrix for modeling the relationship between the entries in \mathbf{W} . Despite the theoretical feasibility for encoding the high-order relationship, \mathbf{W} -vectorization ignores its spatial structure as a matrix. Even worse, the estimation of Ω is extremely challenging due to its high dimension. Such a scale *not only* goes beyond the storage ability of the general memory (Specifically, in our experiment, p is 160, which takes up about 4.9 GB storage), *but also* may lead to the overfitting problem. Therefore, we further assume that the covariance matrix Ω has the Kronecker product decomposition (Gupta and Nagar, 2000), i.e., $\Omega = \Omega_1 \otimes \Omega_2$, where Ω_1 and Ω_2 denote the **row** and **column** covariance matrices, respectively. That is, \mathbf{W} follows the distribution $\mathbf{W} \sim \mathcal{MN}(\mathbf{O}, \Omega_1 \otimes \Omega_2)$. As described earlier, in this paper, we mainly focus on correlation-based methods that generally result in the symmetric BFN. Therefore, the row and column covariance matrices of \mathbf{W} are the same, i.e., $\Omega_1 = \Omega_2$, and without loss of generality, we define $\Omega \triangleq \Omega_1 = \Omega_2$. As a result, the matrix-variate normal distribution (Gupta and Nagar, 2000) of the low-order BFN \mathbf{W} has a probability density:

$$P(\mathbf{W}) = (2\pi)^{-\frac{1}{2}p^2} |\Omega|^{-p} e^{\text{tr}\{-\frac{1}{2}\Omega^{-1}\mathbf{W}\Omega^{-1}\mathbf{W}^T\}} \quad (17)$$

Similar to the formulation in Eqs 11–16, we take Eq. 17 as a prior distribution of low-order network \mathbf{W} . Then, we can formulate the posterior probability of \mathbf{W} based on the Bayesian rule (see **Appendix C** for details). By maximizing the posterior probability, the low- and high-order BFN mutual learning model can be obtained as follows:

$$J(W, \Omega) = \min_{W, \Omega} \left[\left\| W - \mathbf{X}^T \mathbf{X} \right\|_F^2 + \lambda \left[\frac{1}{2} \text{tr} \left(\Omega^{-1} W \Omega^{-1} W^T \right) + p \log (|\Omega|) \right] \right] \quad (18)$$

where \mathbf{W} is the Bayesian low-order BFN, Ω corresponds to the Bayesian high-order BFN, p is the number of ROIs, and λ is a hyper-parameter that controls the balance between the two terms in the objective function.

Algorithm

The alternating optimization (AO) scheme (Bezdek and Hathaway, 2002) is employed to solve Problem (18). More specifically, we first initialize the low-order BFN using the PC estimator, i.e., $\mathbf{W} = \hat{\mathbf{W}} = \mathbf{X}^T \mathbf{X}$, and then alternatively optimize \mathbf{W} and Ω .

Step 1 Fix \mathbf{W} and solve Ω . The optimization problem is

$$\min_{\Omega} \text{tr} \left(\Omega^{-1} \mathbf{W} \Omega^{-1} \mathbf{W}^T \right) + p \log (|\Omega|) \quad (19)$$

which can be solved by the following iterative formula (Dutilleul, 1999; Zhang and Schneider, 2010)

$$\Omega = \mathbf{W}^T \Omega^{-1} \mathbf{W} \quad (20)$$

Note that, by initializing $\Omega = \mathbf{I}$ in Eq. 20, at the first iteration, we obtain $\Omega = \mathbf{W}^T \mathbf{W}$, which reduces to the traditional CC, as $\hat{\mathbf{H}}$ defined in Eq. 6. In other words, the traditional CC is only a rough estimation of the theoretical value at the first iteration. We can continue the iteration toward a more accurate estimation of Ω . In fact, with the estimated Ω , we can further update \mathbf{W} according to the AO scheme. In practice, we generally add a small quantity $\delta \mathbf{I}$ to Eq. 20 for a more stable numerical solution where δ is a small positive constant.

Step 2 Fix Ω and solve \mathbf{W} . The optimization problem is

$$\min_{\mathbf{W}} \left[\left\| \mathbf{W} - \mathbf{X}^T \mathbf{X} \right\|_F^2 + \frac{\lambda}{2} \text{tr} \left(\Omega^{-1} \mathbf{W} \Omega^{-1} \mathbf{W}^T \right) \right] \quad (21)$$

With the fixed Ω , the gradient of Eq. 21 with respect to \mathbf{W} is

$$2\mathbf{W} - 2\mathbf{X}^T \mathbf{X} + 2\lambda \Omega^{-1} \mathbf{W} \Omega^{-1} \quad (22)$$

Setting the gradient equal to zero, we obtain:

$$\mathbf{W} = \mathbf{X}^T \mathbf{X} - \Omega^{-1} \mathbf{W} \Omega^{-1} \quad (23)$$

We summarize the algorithm for solving Problem (18) in **Algorithm 1**.

ALGORITHM 1 | Estimating BFN with BHM model.

```

Input:  $\mathbf{X}$  – data matrix
       $\epsilon$  – iteration stopping threshold
      MaxIter – maximum number of iterations
Output:  $\mathbf{W}$  – Bayesian low-order BFN
        $\Omega$  – Bayesian high-order BFN

Initialize  $\mathbf{W} = \mathbf{X}^T \mathbf{X}$ 
For  $k = 1, 2, \dots, \text{MaxIter}$ 
  Calculate  $\Omega = \mathbf{W}^T \Omega^{-1} \mathbf{W} = \delta \mathbf{I}$ ;
  Update  $\mathbf{W} = \mathbf{X}^T \mathbf{X} - \Omega^{-1} \mathbf{W} \Omega^{-1}$ ;
  Calculate the objective function of Eq. 18:  $J_k = J(\mathbf{W}, \Omega)$ ;
  If  $|J_k - J_{k-1}| = \epsilon$ 
    Break and Return  $\mathbf{W}, \Omega$ ;
  End If
End For
    
```

EXPERIMENTS AND RESULTS

Data Acquisitions and Processing

To evaluate the effectiveness of the proposed BHM, we conduct experiments on Autism Brain Imaging Data Exchange (ABIDE) database. The objective is to identify subjects with ASD from typical controls (TCs). Considering the heterogeneity of multi-site data, we only use data from the NYU site in our study. The dataset includes 184 subjects (79 ASD patients and 105 TCs). The detailed scan procedures and protocols are described on the ABIDE website.² The demographic information of all participants is summarized and displayed in **Table 1**.

All rs-fMRI images were acquired using a standard echo-planar imaging sequence on a clinical routine 3T Siemens Allegra scanner. During the 6-min rs-fMRI scanning procedure, most 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. A few participants close their eyes. The functional scan parameters include the flip angle = 90°, 33 slices, TR/TE = 2000/15 ms with 180 volumes, FOV = 240 mm and voxel size = 3 × 3 × 4 mm³. The rs-fMRI data were preprocessed by DPARSF³ software. Specifically, to avoid the interference of early signal instability, the first 5 rs-fMRI volumes of each subject were discarded. The remaining volumes were calibrated as follows: (1) Slice timing correction and head motion correction; (2) Regression of nuisance signals (ventricle, white matter) and head-motion with Friston 24-parameter model (Friston et al., 1996); (3)

²http://fcon_1000.projects.nitrc.org/indi/abide/

³<http://rfmri.org/dpabi>

TABLE 1 | Demographic information of the used dataset.

	ASD (n = 79)	TC (n = 105)
Gender (M/F)	68/11	79/26
Age (year SD)	14.51 ± 6.23	15.80 ± 3.23
FIQ (mean SD)	107.91 ± 16.62	113.15 ± 13.12
ADOS (mean SD)	11.3 ± 4.08	–

ASD, autism spectrum disorders; TC, typical control; FIQ, full intelligence quotient; ADOS, autism diagnostic observation schedule.

Normalization and register to MNI space with resolution of 3 × 3 × 3 mm³; (4) Segmentation using DATTEL; (5) Spatial smoothing by a kernel of 6 mm. After that, since our focus is functional connectivity, the rs-fMRI time series signals were partitioned into 160 ROIs, according to the functional atlas Dosenbach 160 (Dosenbach et al., 2010). Finally, the mean time series of the ROI were put into a data matrix $\mathbf{X} \in R^{175 \times 160}$, which will be used for the subsequent BFN estimation.

Brain Functional Network Construction, Feature Selection, and Classification

With the preprocessed rs-fMRI data, we estimate the low- and high-order BFNs using the proposed method, i.e., Bayesian low-order Network \mathbf{W} and Bayesian high-order Network Ω , respectively. For comparison, we also choose PC, SR, and traditional CC as baseline methods to construct BFNs.

Once the BFNs are constructed, the next step is feature selection and classification. In our study, we directly use edge weights of the estimated BFN as features for ASD identification. Despite its simplicity (without complex feature design), such a scheme easily causes the curse of dimensionality due to limited sample size. As described previously, the number of ROIs is 160 and thus the estimated feature edges are $160 \times (160 - 1) / 2 = 12720$, which is far greater than the sample size (i.e., the number of subjects 184). To alleviate the problem of small sample size, we adopt a two-sample *t*-test with an empirically fixed *p* values to select features before ASD classification. In our experiments, we evaluate five candidate parametric values of *p*, that is [0.001, 0.005, 0.01, 0.05, 0.1]. The specific parameter analysis results are given in Section “Sensitivity to Network Modeling Parameters.”

To perform the following classification task, we use a linear support vector machine (SVM) (Chang and Lin, 2011) with default *C* = 1 as the classifier. To evaluate the model, we adopt leave-one-out cross-validation (LOOCV) in our experiments due to the limited data samples. Specifically, a LOOCV works in each run and only one sample is used to test while the rest are used to train a classifier. The final performance is obtained by the averaged results of all the runs. Note that the model parameters are involved in certain methods, including SR and BHM. Therefore, we additionally adopt an inner LOOCV procedure on the training data to obtain the optimal parametric value. Specifically, for SR, the regularization parameter λ is set to [2^{-2} , 2^{-1} , 2^0 , 2^1 , 2^2]. For the proposed BHM, the regularization parameter λ is set to [0.0001, 0.001, 0.01, 0.1, 1]. To be consistent with the number of parameters in other methods, the coefficient δ of the perturbation involved in Eq. 20 is set to 0.1 empirically.

Classification Results

To evaluate the classification results of different methods, we use accuracy (ACC), sensitivity (SEN), specificity (SPE) as performance metrics. The definition of these quantities are reported in **Table 2**. Note that, in this work, we treat ASD patients as the positive class while the NCs as the negative class.

TABLE 2 | Different performance metrics.

Performance metrics	Abbreviations	Definitions
Accuracy	ACC	$\frac{TP + TN}{TP + FP + TN + FN}$
Sensitivity	SEN	$\frac{TP}{TP + FN}$
Specificity	SPE	$\frac{TN}{TN + FP}$

TP, TN, FP, and FN indicate true positive, true negative, false positive, and false negative, respectively.

TABLE 3 | The classification results based on five different methods for ASD identification.

Methods	ACC	SEN	SPE
PC	0.6359	0.6222	0.6383
SR	0.6033	0.2658	0.8571
CC	0.6630	0.5570	0.7429
BHM-W	0.6576	0.5316	0.7143
BHM-Ω	0.7283	0.6329	0.8000

In **Table 3**, we report the ASD classification results of five methods. It can be observed that the Bayesian low-order network (BHM-W) and Bayesian high-order network (BHM-Ω) constructed by the proposed BHM perform better than the BFNs constructed by the traditional PC and CC, respectively. Moreover, BHM-Ω achieves the best performance. Besides, the high-order BFNs (traditional CC and BHM-Ω) are associated with better recognition performance when they are compared with the baseline methods PC, SR, and BHM-W. This means that the high-order network structure can provide more helpful information for BFN analysis to some extent. Furthermore, for two corresponding low-order methods, the performance of the traditional PC and BHM-W are approximately similar whereas BHM-W has slightly better accuracy than the traditional PC. This may benefit from the guidance information provided by the Bayesian high-order network Ω in the optimization process.

DISCUSSION

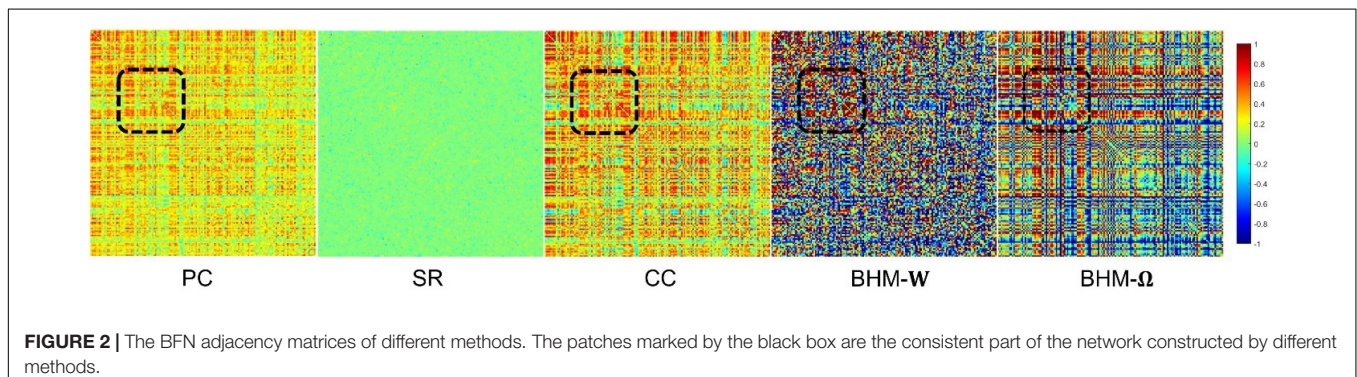
Brain Functional Network Visualization

To evaluate the BFNs estimated by different methods, we randomly select a subject and visualize the BFNs constructed

by PC, SR, CC, BHM, as shown in **Figure 2**. Specifically, the different colors of **Figure 2** indicate different weights of the edge weights matrix (i.e., the BFN), ranging from -1 to 1. It is observed that: (1) Compared with SR, the BFNs estimated by the correlation-based methods (i.e., PC, CC, BHM) are denser since the sparsity prior is introduced into SR. (2) There are fewer areas of cold colors in the PC network heatmap, implying that the edges with the negative weights are less. (3) Compared with PC, CC's network heatmap shows a sharper distinction between the areas of warm and cold colors, indicating that the positive edge weights of CC's BFN are larger and the negative edge weights are smaller. (4) The BHM-W estimated from the Bayesian perspective has a greater distinction between positive and negative edge weights than PC. (5) The BHM-Ω as a Bayesian version of the traditional CC tends to produce a greater distinction between positive and negative edge weights than CC. Combined with the fact that the high classification accuracy of the BHM method, we can infer that the negative edge weights of BFN also have important information for classification. (6) The BFNs based on the correlation methods show a degree of consistency, as shown in the black box in **Figure 2**. Similar structures appear in the four BFNs estimated by PC, CC, BHM, which can provide certain support for the reliability of the three correlation methods. In addition, we cluster the brain regions using spectral clustering (Ng et al., 2001) and visualize the clustered adjacency matrices of these 5 methods in **Figure 3**. It is observed that the BFN estimated by BHM-Ω shows a more significant modular structure.

Sensitivity to Network Modeling Parameters

As stated earlier, some methods including SR and BHM involve optional model parameters. Different parameter values may have a significant impact on the results. Therefore, we calculate the accuracy of different methods under different parameter values, as shown in **Figure 4**. It is worth noting that, traditional PC and CC models do not involve optional parameters. However, for comparison, we fix their values with the final classification accuracy in **Figure 4** for visualization. We can observe that BHM-Ω and BHM-W are quite sensitive to the parameters. When λ in BHM is set to a large real number, the accuracy of BHM-Ω decreases, which may be because the large value of λ, the algorithm has difficulties to converge. Besides, the accuracy of



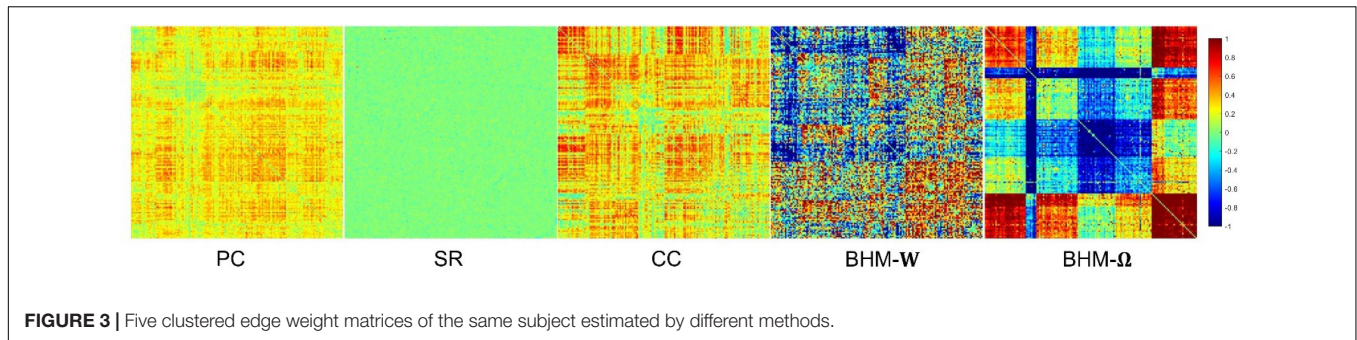


FIGURE 3 | Five clustered edge weight matrices of the same subject estimated by different methods.

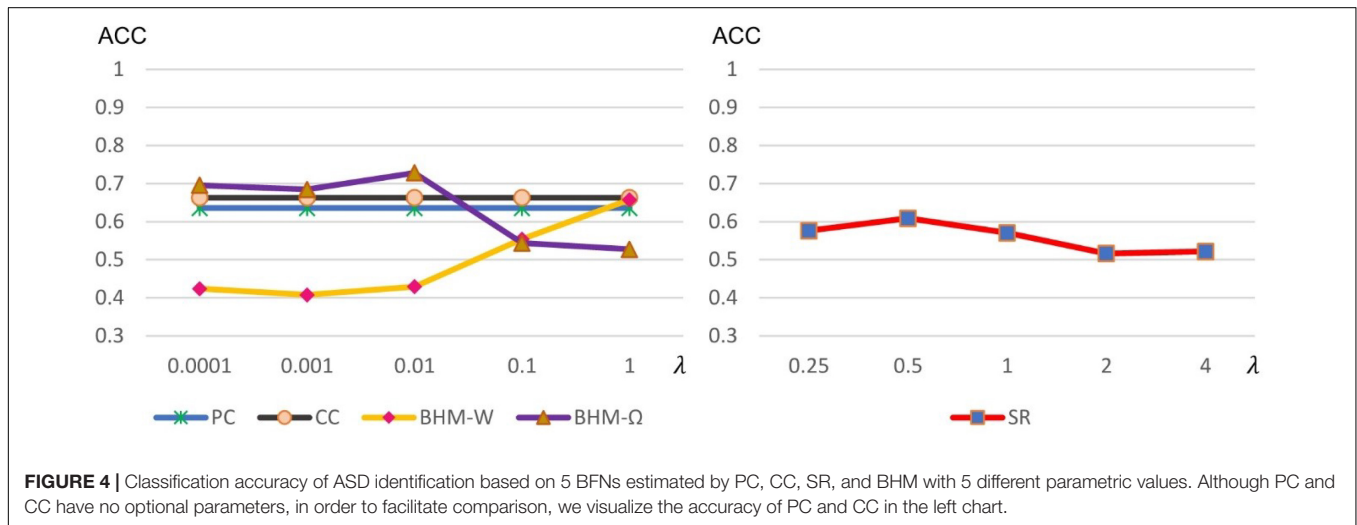


FIGURE 4 | Classification accuracy of ASD identification based on 5 BFNs estimated by PC, CC, SR, and BHM with 5 different parametric values. Although PC and CC have no optional parameters, in order to facilitate comparison, we visualize the accuracy of PC and CC in the left chart.

BHM-W increases as λ increases. For this, we empirically tested a larger lambda range [2^{-5} , 2^{-4} , \dots , 2^0 , \dots , 2^4 , 2^5] and find that as λ continues to increase, the accuracy decreases, which is consistent with the performance of BHM- Ω . Moreover, SR is not sensitive to different parameter values, but its accuracy performs average in general.

Considering that different p -values significantly influence the results, we show the classification accuracies of 5 methods under different p -values in **Figure 5**. Note that all 5 methods are sensitive to different p -values. We selected the optimal parameter value for feature selection, so that different methods can get the best classification performance.

Top Discriminative Features

In this work, for the ASD classification task, we use the edge weights of the estimated BFN as features. With the empirically optimal parameter, we construct the BFNs using the proposed BHM, then apply a two-sample t -test to rearrange the features according to the p -values. Particularly, we choose the BHM- Ω since it outperforms the BFNs estimated by the other methods. As a result, we obtain the discriminative edge connections with a threshold value $p < 0.001$ as shown in **Figure 6**. Here, the thickness of each arc represents the discriminative power that is inversely proportional to the corresponding p -value. The colors of each arc are assigned randomly for better visualization.

In **Figure 6**, the top discriminative features and the corresponding brain regions, that may contribute to ASD identification include occipital lobe, post-cingulate, dorsal frontal cortex, inferior parietal lobule, precuneus, anterior prefrontal cortex, lateral cerebellum, temporal lobe, fusiform gyrus, mid insula, etc. in order of discriminant ability. The findings are consistent with previous studies (Nickl-Jockschat et al., 2012;

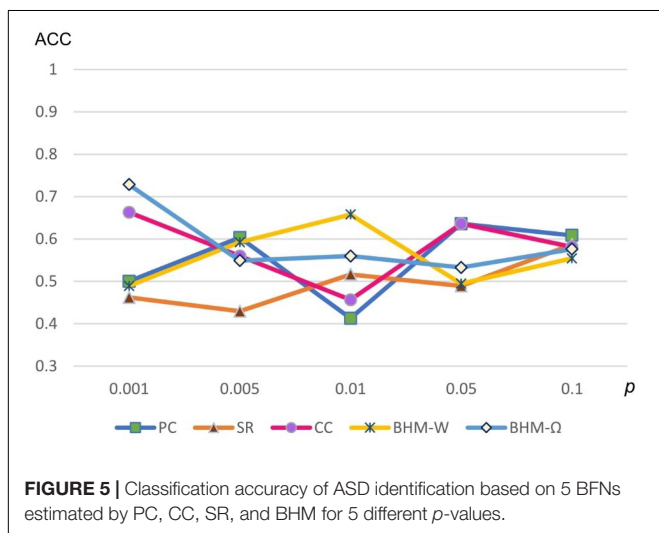


FIGURE 5 | Classification accuracy of ASD identification based on 5 BFNs estimated by PC, CC, SR, and BHM for 5 different p -values.

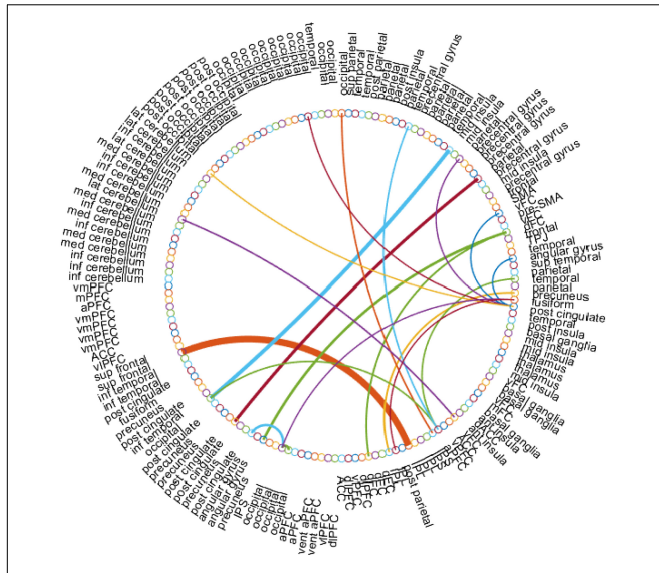


FIGURE 6 | The most discriminative edge features of the BHM-Ω involved in the ASD classification task by using a t-test ($p < 0.001$). This figure is created by the circularGraph tool, which is designed by Paul Kassebaum and can be downloaded from <http://www.mathworks.com/matlabcentral/fileexchange/48576-circulargraph>.

Hashem et al., 2020; Lau et al., 2020). We visualize the ROIs using the Brainmesh of Ch2 with Cerebellum in **Figure 7**, where the size of node spheres depends on the original value in the node file provided by the Dosenbach 160 template.

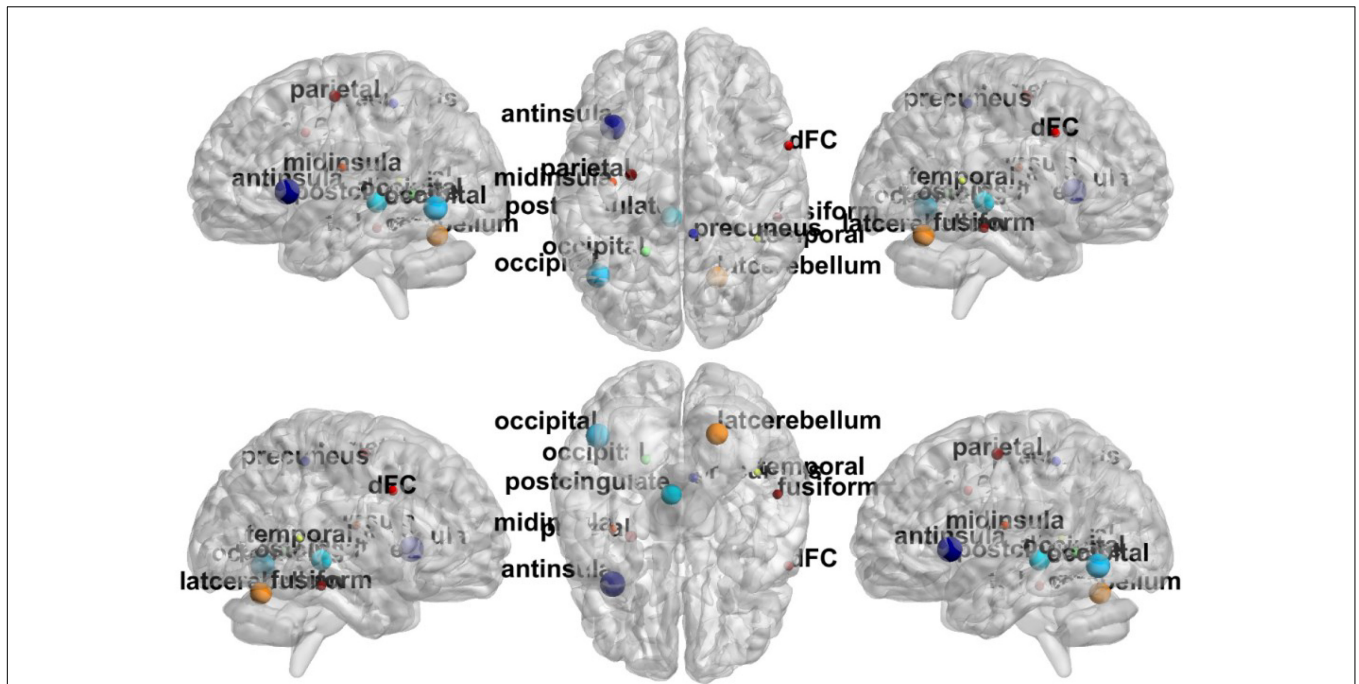


FIGURE 7 | The full view of most relevant ROI associated with the ASD classification task based on BHM-Ω. This visualization is created using the BrainNet Viewer (<https://www.nitrc.org/projects/bnv/>).

Other Distribution Priors

As described before, we first give an equivalent probability explanation for PC by introducing a normal distribution for the rs-fMRI signal values. Then we reformulate PC with Bayesian rule, thus getting two perspectives of PC. This provides a platform for generalizing PC to CC by assuming that the edge weight matrix **W** follows the MVN distribution prior. As a result, we derive a probabilistic explanation of CC and develop a high-order BFN estimation framework that allows the introduction of different priors (or regularizers).

Besides the introduced normal distribution prior on **W** for BFN estimation, we can also introduce other priors on **W**. For example, considering Laplacian distribution prior for w_{ij} , e.g., $P(w_{ij}) = \frac{1}{2\beta} e^{-\frac{|w_{ij}|}{\beta}}$ where β is a scale parameter. In this way, the regularized least square problem is

$$\min_{w_{ij}} \sum_{ij} (||\mathbf{x}_i - w_{ij}\mathbf{x}_j||^2 + \lambda |w_{ij}|) \tag{24}$$

We can find that the Laplacian distribution generates sparse BFN due to the regularizer w_{ij} . Besides, we can get the optimal solution by the soft thresholding, as follows:

$$\hat{w}_{ij} = \begin{cases} \mathbf{x}_i^T \mathbf{x}_j - \lambda, & \mathbf{x}_i^T \mathbf{x}_j \geq \lambda \\ 0, & \mathbf{x}_i^T \mathbf{x}_j < \lambda \end{cases} \tag{25}$$

Although different prior distributions can be tried to introduce the proposed probabilistic framework, we do not

formulate their models in detail since this paper focuses on the formulation of CC.

CONCLUSION

In this paper, we propose a probabilistic high-order BFN learning framework with a matrix normal penalty for ASD identification. As pointed out previously, CC is intuitively defined based on two sequential PC operations and falls short of a rigorous mathematical basis. To address this issue, we first reformulate PC with Bayesian rule and then generalize PC to CC by assuming that the edge weight matrix follows a matrix-variate normal distribution prior. This work lays the theoretical foundation for CC methods, leading to a better understanding of high-order BFN learning. In this base, we develop a Bayesian High-order Model to simultaneously estimate the high- and low-order BFN. To efficiently solve the proposed objective function, an alternating optimization algorithm is proposed. Extensive experiments on the NYU site of ABIDE dataset demonstrate the effectiveness of the proposed method, in comparison to the baseline methods. Especially for the BHM- Ω , it achieves the best performance. Note that we only construct high-order BFN based on PC. In principle, any correlation-based BFN estimation method (e.g., SR) can be embedded in the proposed probabilistic framework. In the future, we plan to validate the scheme on the other correlation-based BFN models.

REFERENCES

- Ariana, A., and Cohen, M. S. (2013). Decreased small-world functional network connectivity and clustering across resting state networks in Schizophrenia: an fMRI classification tutorial. *Front. Human Neurosci.* 7:520. doi: 10.3389/fnhum.2013.00520
- Baggio, H. C., Sala-Llonch, R., Segura, B., Marti, M.-J., Valldeoriola, F., Compta, Y., et al. (2014). Functional brain networks and cognitive deficits in Parkinson's disease. *Human Brain Mapp.* 35, 4620–4634. doi: 10.1002/hbm.22499
- Bezdek, J. C., and Hathaway, R. J. (2002). Some notes on alternating optimization. *Lect. Notes Comp. Sci.* 2275, 187–195.
- Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409
- Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 1–27. doi: 10.1145/1961189.1961199
- Chen, X., Zhang, H., Gao, Y., Wee, C.-Y., and Li, G. (2016). High-order resting-state functional connectivity network for MCI classification. *Human Brain Mapp.* 37, 3282–3296. doi: 10.1002/hbm.23240
- Dadi, K., Rahim, M., Abraham, A., Chyzykh, D., Milham, M., Thirion, B., et al. (2019). Benchmarking functional connectome-based predictive models for resting-state fMRI. *NeuroImage* 192, 115–134. doi: 10.1016/j.neuroimage.2019.02.062
- Dosenbach, N. U. F., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., et al. (2010). Prediction of individual brain maturity using fMRI. *Science* 329, 1358–1361. doi: 10.1126/science.1194144
- Dutilleul, P. (1999). The MLE algorithm for the matrix normal distribution. *J. Stat. Comp. Simul.* 64, 105–123. doi: 10.1080/00949659908811970
- Eguiluz, V. M., Chialvo, D. R., Cecchi, G. A., Baliki, M., and Apkarian, A. V. (2005). Scale-free brain functional networks. *Phys. Rev. Lett.* 94:18102. doi: 10.1103/PhysRevLett.94.018102
- Friedman, J. H., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9, 432–441. doi: 10.1093/biostatistics/kxm045

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

LQ proposed the idea of Bayesian high-order model. RD steered the structure of the manuscript and polished the language. XJ validated the models' derivation, coded and performed the experiments, and planned and wrote the manuscript. YuZ set the procedures of ASD identification experiments. YiZ and LZ designed a core derivation of the models. All authors developed the estimation algorithm and contributed to the preparation of the manuscript.

FUNDING

This work was partly supported by the National Natural Science Foundation of China (Nos. 61976110, 62176112, and 11931008), Natural Science Foundation of Shandong Province (No. ZR202102270451), and The Open Project of Liaocheng University Animal Husbandry Discipline (No. 319312101-01).

- Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *NeuroImage* 19, 1273–1302. doi: 10.1016/s1053-8119(03)00202-7
- Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S., and Turner, R. (1996). Movement-related effects in fMRI time-series. *Magn. Reson. Med.* 35, 346–355. doi: 10.1002/mrm.1910350312
- Gupta, A. K., and Nagar, D. K. (2000). *Matrix Variate Distributions*. London: Chapman and Hall.
- Hahn, K., Myers, N., Prigarin, S., Rodenacker, K., Kurz, A., Förstl, H., et al. (2013). Selectively and progressively disrupted structural connectivity of functional brain networks in Alzheimer's disease — Revealed by a novel framework to analyze edge distributions of networks detecting disruptions with strong statistical evidence. *NeuroImage* 81, 96–109. doi: 10.1016/j.neuroimage.2013.05.011
- Hashem, S., Nisar, S., Bhat, A. A., Yadav, S. K., Azeem, M. W., Bagga, P., et al. (2020). Genetics of structural and functional brain changes in autism spectrum disorder. *Transl. Psychiatry* 10, 229–229. doi: 10.1038/s41398-020-00921-3
- Jiang, X., Zhang, L., Qiao, L., and Shen, D. (2019). Estimating functional connectivity networks via low-rank tensor approximation with applications to MCI identification. *IEEE Trans. Bio-Med. Eng.* 67, 1912–1920. doi: 10.1109/TBME.2019.2950712
- Jie, P., Pei, W., Zhou, N., and Ji, Z. (2009). Partial correlation estimation by joint sparse regression models. *J. Am. Stat. Assoc.* 104, 735–746. doi: 10.1198/jasa.2009.0126
- Lau, W. K. W., Leung, M.-K., and Zhang, R. (2020). Hypofunctional connectivity between the posterior cingulate cortex and ventromedial prefrontal cortex in autism: Evidence from coordinate-based imaging meta-analysis. *Prog. Neuro-Psychopharm. Biol. Psychiatry* 103:109986. doi: 10.1016/j.pnpbp.2020.109986
- Lee, H., Lee, D. S., Kang, H., Kim, B., and Chung, M. K. (2011). Sparse brain network recovery under compressed sensing. *IEEE Trans. Med. Imag.* 30, 1154–1165. doi: 10.1109/TMI.2011.2140380
- Li, W., Wang, Z., Zhang, L., Qiao, L., and Shen, D. (2017). Remodeling pearson's correlation for functional brain network estimation and autism spectrum disorder identification. *Front. Neuroinform.* 11:55. doi: 10.3389/fninf.2017.00055

- Long, Y., Cao, H., Yan, C., Chen, X., Li, L., Castellanos, F. X., et al. (2020). Altered resting-state dynamic functional brain networks in major depressive disorder: findings from the REST-meta-MDD consortium. *NeuroImage. Clin.* 26:102163. doi: 10.1016/j.nicl.2020.102163
- Mahjoub, I., Mahjoub, M. A., Rezik, I., Weiner, M., Aisen, P., Petersen, R., et al. (2018). Brain multiplexes reveal morphological connective biomarkers fingerprinting late brain dementia states. *Sci. Rep.* 8:4103. doi: 10.1038/s41598-018-21568-7
- Marrelec, G., Krainik, A., Duffau, H., Péligrini-Issac, M., Lehericy, S., Doyon, J., et al. (2006). Partial correlation for functional brain interactivity investigation in functional MRI. *NeuroImage* 32, 228–237. doi: 10.1016/j.neuroimage.2005.12.057
- McIntosh, A. R., and Gonzalez-Lima, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Human Brain Mapp.* 2, 2–22. doi: 10.1002/hbm.460020104
- Ng, A. Y., Jordan, M. I., and Weiss, Y. (2001). On spectral clustering: analysis and an algorithm. *Adv. Neural Inform. Proc. Syst.* 14, 849–856.
- Nickl-Jockschat, T., Habel, U., Michel, T. M., Manning, J., Laird, A. R., Fox, P. T., et al. (2012). Brain structure anomalies in autism spectrum disorder—a meta-analysis of VBM studies using anatomic likelihood estimation. *Human Brain Mapp.* 33, 1470–1489. doi: 10.1002/hbm.21299
- Pervaiz, U., Vidaurre, D., Woolrich, M. W., and Smith, S. M. (2020). Optimising network modelling methods for fMRI. *NeuroImage* 211:116604. doi: 10.1016/j.neuroimage.2020.116604
- 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
- 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.
- Wang, J., Wang, Q., Zhang, H., Chen, J., Wang, S., and Shen, D. (2019). Sparse multiview task-centralized ensemble learning for ASD diagnosis based on age- and sex-related functional connectivity patterns. *IEEE Trans. Cybernet.* 49, 3141–3154. doi: 10.1109/TCYB.2018.2839693
- Wang, J., Zhang, L., Wang, Q., Chen, L., Shi, J., Chen, X., et al. (2020). Multi-class ASD classification based on functional connectivity and functional correlation tensor via multi-source domain adaptation and multi-view sparse representation. *IEEE Trans. Med. Imag.* 39, 3137–3147. doi: 10.1109/TMI.2020.2987817
- Yu, R., Zhang, H., An, L., Chen, X., Wei, Z., and Shen, D. (2017). Connectivity strength-weighted sparse group representation-based brain network construction for MCI classification. *Hum. Brain Mapp.* 38, 2370–2383. doi: 10.1002/hbm.23524
- Zhang, H. T. (2010). The benefit of group sparsity. *Ann. Stat.* 38, 1978–2004.
- Zhang, H., Chen, X., Shi, F., Li, G., Kim, M., Giannakopoulos, P., et al. (2016). Topographical information-based high-order functional connectivity and its application in abnormality detection for Mild Cognitive Impairment. *J. Alzheimer's Dis.* 54, 1095–1112. doi: 10.3233/JAD-160092
- Zhang, Y., and Schneider, J. (2010). Learning multiple tasks with a sparse matrix-normal penalty. *Adv. Neural Inform. Proc. Syst.* 23, 2550–2558.
- Zhang, Y., Zhang, H., Chen, X., Lee, S. W., and Shen, D. (2017). Hybrid high-order functional connectivity networks using resting-state functional MRI for Mild Cognitive Impairment diagnosis. *Sci. Rep.* 7:6530. doi: 10.1038/s41598-017-06509-0
- Zhou, Y., Qiao, L., Li, W., Zhang, L., and Shen, D. (2018a). Simultaneous estimation of low- and high-order functional connectivity for identifying Mild Cognitive Impairment. *Front. Neuroinform.* 12:3. doi: 10.3389/fninf.2018.00003
- Zhou, Y., Zhang, L., Teng, S., Qiao, L., and Shen, D. (2018b). Improving sparsity and modularity of high-order functional connectivity networks for MCI and ASD identification. *Front. Neurosci.* 12:959. doi: 10.3389/fnins.2018.00959

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Jiang, Zhou, Zhang, Zhang, Qiao and De Leone. 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) and the copyright owner(s) 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.

APPENDIX

This Appendix consists of three appendices: Appendix A gives a detailed explanation of Eq. 9, Appendix B presents a detail for Eq. 12 and Appendix C gives a detailed derivation of Eq. 18. To keep the process of derivation smooth, we write the formulas that appeared above with the original number in the appendix, while the new formulas in the process of derivation was renumbered.

Appendix A

Given x_{kj} and a_{ij} are constants, x_{ki} follows the normal distribution $x_{ki} \sim \mathcal{N}(a_{ij}x_{kj}, \sigma^2)$. The conditional distribution can be written as

$$P(x_{ki} | a_{ij}, x_{kj}, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_{ki}-a_{ij}x_{kj})^2}{2\sigma^2}} \tag{A1'}$$

Assuming that the variables x_{ki} of rs-fMRI time series \mathbf{x}_i , $i = 1, \dots, p$ are independent identically distributed, the likelihood function can be written as follows:

$$\begin{aligned} P(\mathbf{x}_i | a_{ij}, \mathbf{x}_j, \sigma^2) &= \prod_{k=1}^n P(x_{ki} | a_{ij}, x_{kj}, \sigma^2) \\ &= (\sqrt{2\pi})^{-n} e^{-\frac{\|\mathbf{x}_i - a_{ij}\mathbf{x}_j\|^2}{2\sigma^2}} \end{aligned} \tag{A2'}$$

To avoid overflow caused by multiplying operations in Eq. A2', we use the log-likelihood function and further maximize it:

$$\max_{a_{ij}} \log P(\mathbf{x}_i | a_{ij}, \mathbf{x}_j, \sigma^2) = \max_{a_{ij}} -n \log(\sqrt{2\pi}) - \frac{\|\mathbf{x}_i - a_{ij}\mathbf{x}_j\|^2}{2\sigma^2} \tag{A3'}$$

Appendix B

The edge weight w_{ij} of BFN has the following prior probabilistic density:

$$P(w_{ij}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w_{ij}^2}{2}} \tag{A11}$$

According to Bayesian rule, the posterior distribution of w_{ij} is proportional to $P(\mathbf{x}_i | w_{ij}, \mathbf{x}_j) P(w_{ij})$:

$$P(w_{ij} | \mathbf{x}_i, \mathbf{x}_j) \propto P(\mathbf{x}_i | w_{ij}, \mathbf{x}_j) P(w_{ij}) \tag{A4'}$$

where $P(\mathbf{x}_i | w_{ij}, \mathbf{x}_j)$ is the likelihood of w_{ij} for \mathbf{x}_i . Based on Eqs A2', 11

$$P(\mathbf{x}_i | w_{ij}, \mathbf{x}_j) P(w_{ij}) = (\sqrt{2\pi})^{-(n+1)} \sigma^{-n} e^{-\frac{\|\mathbf{x}_i - w_{ij}\mathbf{x}_j\|^2}{2\sigma^2} - \frac{w_{ij}^2}{2}} \tag{A5'}$$

Taking the logarithm on Eq. A4', we obtain

$$\log P(w_{ij} | \mathbf{x}_i, \mathbf{x}_j, \sigma^2) \propto -(n+1) \log(\sqrt{2\pi}) = (-n) \log \sigma - \frac{\|\mathbf{x}_i - w_{ij}\mathbf{x}_j\|^2}{2\sigma^2} - \frac{w_{ij}^2}{2} \tag{A6'}$$

Next, the log-posterior probability is maximized (i.e., maximal posterior estimation) as follows,

$$\max_{w_{ij}} -(n+1) \log(\sqrt{2\pi}) = (-n) \log \sigma - \frac{\|\mathbf{x}_i - w_{ij}\mathbf{x}_j\|^2}{2\sigma^2} = \sigma^2 w_{ij}^2 \tag{A12}$$

Appendix C

As shown in Section “Learning High-Order Brain Functional Network With a Matrix-Normal Penalty,” the proposed model is formulated as

$$J(\mathbf{W}, \mathbf{\Omega}) = \min_{\mathbf{W}, \mathbf{\Omega}} \left\| \mathbf{W} - \mathbf{X}^T \mathbf{X} \right\|_F^2 + \lambda \left[\frac{1}{2} \text{tr}(\mathbf{\Omega}^{-1} \mathbf{W} \mathbf{\Omega}^{-1} \mathbf{W}^T) + p \log(|\mathbf{\Omega}|) \right] \tag{A18}$$

As described before, we assume that any time series \mathbf{x}_i from the data set $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p]^T$ follows the conditional distribution as:

$$\begin{aligned}
 P(\mathbf{x}_i \mid a_{ij}, \mathbf{x}_j, \sigma^2) &= \prod_{k=1}^n P(x_{ki} \mid a_{ij}, x_{kj}, \sigma^2) \\
 &= (\sqrt{2\pi}\sigma)^{-n} e^{-\frac{\|\mathbf{x}_i - a_{ij}\mathbf{x}_j\|^2}{2\sigma^2}}
 \end{aligned} \tag{A2'}$$

Using the maximum likelihood estimation for a_{ij} , we get $a_{ij} = \mathbf{x}_i^T \mathbf{x}_j = w_{ij}$. Therefore, we rewrite Eq. A2' as follows.

$$P(\mathbf{x}_i \mid w_{ij}, \mathbf{x}_j, \sigma^2) = (\sqrt{2\pi}\sigma)^{-n} e^{-\frac{\|\mathbf{x}_i - w_{ij}\mathbf{x}_j\|^2}{2\sigma^2}} \tag{A7'}$$

We can further convert Eq. A7' to the matrix form:

$$P(\mathbf{X} \mid \mathbf{W}) = \prod_{i,j=1}^p P(\mathbf{x}_i \mid w_{ij}, \mathbf{x}_j, \sigma^2) = (\sqrt{2\pi}\sigma)^{-p^2 n} e^{-\frac{\sum_{i,j=1}^p \|\mathbf{x}_i - w_{ij}\mathbf{x}_j\|^2}{2\sigma^2}} \tag{A8'}$$

Furthermore, as mentioned in Section ‘‘Learning High-Order Brain Functional Network With a Matrix-Normal Penalty,’’ we assume that the low-order correlation matrix \mathbf{W} follows the MVN distribution with the probabilistic density

$$P(\mathbf{W}) = (2\pi)^{-\frac{1}{2}p^2} |\Omega|^{-p} e^{\text{tr}\{-\frac{1}{2}\Omega^{-1}\mathbf{W}\Omega^{-1}\mathbf{W}^T\}} \tag{A9'}$$

Note that the row\column variance matrix Ω is considered as the high-order correlation matrix between \mathbf{x}_i since it models the relationships between w_{ij} .

Based on the Bayesian rule, the posterior probability of \mathbf{W} is given by

$$P(\mathbf{W} \mid \mathbf{X}) \propto P(\mathbf{X} \mid \mathbf{W}) \cdot P(\mathbf{W}) \tag{A10'}$$

From Eqs A8', A9', we obtain

$$P(\mathbf{X} \mid \mathbf{W}) \cdot P(\mathbf{W}) = (2\pi)^{-\frac{p^2(n+1)}{2}} \sigma^{-p^2 n} |\Omega|^{-p} \cdot e^{\text{tr}\{-\frac{1}{2}\Omega^{-1}\mathbf{W}\Omega^{-1}\mathbf{W}^T\} - \frac{\sum_{i,j=1}^p \|\mathbf{x}_i - w_{ij}\mathbf{x}_j\|^2}{2\sigma^2}} \tag{A11'}$$

Taking the logarithm of the above likelihood function and then maximizing it, we get

$$\max_{\mathbf{W}, \Omega} -\frac{\sum_{i,j=1}^p \|\mathbf{x}_i - w_{ij}\mathbf{x}_j\|^2}{2\sigma^2} = \text{tr}\left\{-\frac{1}{2}\Omega^{-1}\mathbf{W}\Omega^{-1}\mathbf{W}^T\right\} - p\log(|\Omega|) \tag{A12'}$$

The above is equivalent to the optimization problem:

$$\min_{\mathbf{W}, \Omega} \frac{\sum_{i,j=1}^p \|\mathbf{x}_i - w_{ij}\mathbf{x}_j\|^2}{2\sigma^2} = \text{tr}\left\{\frac{1}{2}\Omega^{-1}\mathbf{W}\Omega^{-1}\mathbf{W}^T\right\} = p\log(|\Omega|) \tag{A13'}$$

Consider that the solution of the former is $w_{ij} = \mathbf{x}_i^T \mathbf{x}_j$. Note that $\min_{w_{ij}} \sum_{i,j=1}^p \|\mathbf{x}_i - w_{ij}\mathbf{x}_j\|^2$ can be rewritten in matrix form as $\min_{\mathbf{W}} \|\mathbf{W} - \mathbf{X}^T \mathbf{X}\|_F^2$. Therefore, problem (A13') can be transformed as

$$\min_{\mathbf{W}, \Omega} \|\mathbf{W} - \mathbf{X}^T \mathbf{X}\|_F^2 = \lambda \left[\frac{1}{2} \text{tr}(\Omega^{-1}\mathbf{W}\Omega^{-1}\mathbf{W}^T) = p\log(|\Omega|) \right] \tag{A14'}$$