A Bayesian incorporated linear non-Gaussian acyclic model for multiple directed graph estimation to study brain emotion circuit development in adolescence

Aiying Zhang1, Gemeng Zhang2, Biao Cai2, Tony W. Wilson3, Julia M. Stephen4, Vince D. Calhoun5, and Yu-Ping Wang2

1School of Data Science, University of Virginia, Charlottesville, VA, USA
2Department of Biomedical Engineering, Tulane University, New Orleans, LA, USA
3Institute for Human Neuroscience, Boys Town National Research Hospital, Boys Town, NE, USA
4Mind Research Network, Albuquerque, NM, USA
5Tri-institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, GA, USA

Keywords: Adolescence, Brain development, fMRI, Emotion processing, Bayesian network

ABSTRACT

Emotion perception is essential to affective and cognitive development which involves distributed brain circuits. Emotion identification skills emerge in infancy and continue to develop throughout childhood and adolescence. Understanding the development of the brain’s emotion circuitry may help us explain the emotional changes during adolescence. In this work, we aim to deepen our understanding of emotion-related functional connectivity (FC) from association to causation. We proposed a Bayesian incorporated linear non-Gaussian acyclic model (BiLiNGAM), which incorporated association model into the estimation pipeline. Simulation results indicated stable and accurate performance over various settings, especially when the sample size was small. We used fMRI data from the Philadelphia Neurodevelopmental Cohort (PNC) to validate the approach. It included 855 individuals aged 8–22 years who were divided into five different adolescent stages. Our network analysis revealed the development of emotion-related intra- and intermodular connectivity and pinpointed several emotion-related hubs. We further categorized the hubs into two types: in-hubs and out-hubs, as the center of receiving and distributing information, respectively. In addition, several unique developmental hub structures and group-specific patterns were discovered. Our findings help provide a directed FC template of brain network organization underlying emotion processing during adolescence.

AUTHOR SUMMARY

Our study introduces a novel method for analyzing directed graphs across multiple groups and demonstrates its effectiveness through a series of simulation studies. This method is applied to investigate the development of directed functional connectivity for emotion processing across diverse adolescent periods. Our findings unveil a notable increase in interfucational connectivity with age, specifically involved with the executive control and memory retrieval, indicating the maturation of emotion processing function. Additionally, significant development of intraconnectivity in the subcortical areas emerges in early adolescence,
whereas development of cerebellum emerges in the very end of adolescence. These insights offer valuable contributions to our understanding of the dynamic neural processes underlying emotion regulation during adolescence.

INTRODUCTION

Emotion perception is essential to affective and cognitive development, which plays a pivotal role in social interactions and mental health, especially during adolescence—a critical period characterized by significant emotional and cognitive transformations. The ability to identify distinct facial expressions is fundamental for recognizing others’ emotional states and is supported by distributed brain circuits that continue to develop throughout childhood and adolescence (Gee et al., 2022; Zhang, Padmanabhan, Gross, & Menon, 2019). Understanding the development of the brain’s emotion circuits during this period can shed light on the process of emotional maturation in adolescents and offer insights into potential neural deviations that may indicate a risk for psychiatric disorders.

Functional magnetic resonance imaging (fMRI) has been widely applied to study functional development in emotion circuits, because of its relative simplicity of use, noninvasive nature, and relatively high spatial resolution. In our previous work (Zhang, Cai, et al., 2020), we studied the fMRI images collected from the Philadelphia Neurodevelopmental Cohort (PNC) and delineated the trajectory of brain functional connectivity (FC) from late childhood (preadolescence) to early adulthood (postadolescence) during emotion identification tasks. The FC metrics that we used were defined by statistical associations (partial correlations, in particular) between measured brain regions. However, it has been pointed out that the statistical association may be problematic in that it only reveals the spatial connections but not causal information (Reid et al., 2019). Approaches that characterize statistical associations are likely a good starting point, but the directionality of brain connections (i.e., directed functional connectivity [dFC]) should provide more informative insights.

Directed acyclic graph (DAG) models, also known as Bayesian networks, are designed to model causal relationships in complex systems. Many fMRI studies have utilized the DAG models to characterize the dFC and elucidate their differences between different population groups (Henry & Gates, 2017; Smith et al., 2011). Iyer et al. (2013) validated the Peter Spirtes and Clark Glymour algorithm (PC algorithm, Spirtes et al. (2000)) in dFC estimation; Hanson, Hanson, Ramsey, and Glymour (2013) applied the greedy equivalent search (GES) method (Chickering, 2002) and investigated the differences in brain integration between neurotypical controls and patients with autism spectrum disorders; Liu et al. (2015) employed the linear non-Gaussian acyclic model (LiNGAM) (Shimizu, Hoyer, Hyvärinen & Kerminen, 2006) to account for the non-Gaussianity of fMRI data and revealed network differences between patients with bipolar disorders and those with major depression; Zhang et al. (2022) improved the Non-combinatorial Optimization via Trace Exponential and Augmented lagRangian for Structure learning (Notears, Zheng, Aragam, Ravikumar, and Xing (2018)) to detect abnormal dFC in schizophrenia.

Despite the success of DAGs in case-control comparisons, many biomedical applications involve data from multiple groups. Our study, for instance, aims to estimate multiple DAGs to reveal the development of emotion-related dFC across different adolescent periods. This raises an important statistical question, namely how to jointly estimate related graphical models in order to effectively make use of the available data (Wang, Segarra, & Uhler, 2018). Oates,
Smith, Mukherjee, and Cussens (2016) presented an ILP approach for joint estimation over multiple DAGs for small graphs. Later, Wang et al. (2018) proposed jointGES for Gaussian distributed data under high-dimensional cases. Specifically for non-Gaussian distributed fMRI data (Ramsey, Sanchez-Romero, & Glymour, 2014), Shimizu (2012) developed a joint estimation method for multiple LiNGAMs; Ramsey et al. (2010) proposed the independent multisample GES (iMaGES) algorithm to account for graphical differences and proved usefulness in understanding emotion regulation during the child-to-adolescent development (Gee et al., 2022). However, most methods assume that the DAGs across groups share the same directed connectivity structure while the connection strengths can vary (Ramsey, Hanson, & Glymour, 2011; Shimizu, 2012), which is not always true in reality. To relax the common-structure limitation, we propose the Bayesian incorporated linear non-Gaussian acyclic model (BiLiNGAM). The BiLiNGAM leverages insights from the association networks as prior information to narrow down the set of plausible DAG spaces. Specifically, we implement a joint Bayesian estimation approach (Zhang, Cai, et al., 2020) to acquiring the priors for multiple groups, optimizing for group similarities while accommodating their unique characteristics. A series of simulation studies have been conducted to illustrate the advantages of BiLiNGAM in terms of convergence speed and accuracy, especially in high dimensional datasets.

We applied our proposed model to fMRI data collected during an emotion identification task from participants in the Philadelphia Neurodevelopmental Cohort (PNC). Our analysis aimed to discern emotion-related dFC patterns across five distinct adolescent groups, shedding light on the nuanced changes within the emotion circuitry during this critical developmental period. By examining both common and distinctive dFC patterns, we delved into the evolving trajectories of functional network modules and identified pivotal hub regions of interest (ROIs) in the adolescent brain.

**METHODS AND MATERIALS**

In this section, we first introduce the proposed BiLiNGAM model, specifically designed for estimating multiple DAGs that are distinct yet related, particularly in high-dimensional settings. Then we describe the simulation studies to demonstrate the advantages of BiLiNGAM model and the application to investigate emotion-related dFC patterns across different adolescence periods.

**Background: DAG and LiNGAM**

A directed acyclic graph (DAG) is a probabilistic graphical model that represents a set of variables and their conditional dependencies via the directed edges. Formally, a DAG is denoted by $G = (V, E)$ where $V = \{1, 2, \ldots, p\}$ represents the set of nodes and $E \subset V \times V$ represents the set of directed edges. The nodes in set $V$ correspond to the set of variables $X = (X_1, X_2, \ldots, X_p)^T$, which are ROIs in our case. An edge $(i, j)$ is directed if $(i, j) \in E$ but $(j, i) \notin E$, and we denote it as $i \rightarrow j$. A DAG is acyclic, meaning that there are no circular dependencies within the graph.

The linear non-Gaussian acyclic model (LiNGAM) used the structural equation model (SEM) to estimate the DAG for non-Gaussian variables (Shimizu et al., 2006). Let $B = \{b_{ij}\} \in \mathbb{R}^{p \times p}$ be the weighted adjacency matrix specifying the edge weights of the underlying DAG $G$. The observed random vector $X = (X_1, X_2, \ldots, X_p) \in \mathbb{R}^p$ is assumed to be generated from the following linear SEM,

$$X = B^T X + \epsilon,$$
where \( \mathbf{e} = (\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_p) \) is a continuous random vector; the \( \mathbf{e}_i \)'s, \( \forall i = 1, 2, ..., p \) have non-Gaussian distributions with nonzero variances, and are independent of each other. A property of acyclicity is that there exists at least one permutation \( \pi \) of \( p \) variables such that \( b_{ij} = 0 \), \( \forall \pi(i) < \pi(j) \). In other words, the weight matrix \( \mathbf{B} \) can be reordered to a strictly lower triangular matrix according to the permutation \( \pi \). The goal of LiNGAM is to find the correct permutation and estimate the weight matrix \( \mathbf{B} \). Since the components of \( \mathbf{e} \) are independent and non-Gaussian, Shimizu et al. (2006) first proposed the independent component analysis (ICA) based algorithm known as ICA-LiNGAM. Later, the direct LiNGAM (Shimizu et al., 2011) was developed, which estimates the causal order of variables by successively subtracting the effect of each independent component from the given data. Compared to the ICA-based algorithm, the direct LiNGAM needs no initial guess or algorithmic parameters and has guaranteed convergence.

In fMRI studies, it has been demonstrated that the non-Gaussianity can provide valuable orientation information, leading to more accurate identification of directed graph structures compared to traditional Gaussian settings (Ramsey et al., 2014). However, when dealing with heterogeneous data with limited sample size, the LiNGAMs may not yield satisfactory estimations or fail under high dimensional setting. This highlights the need for additional modeling techniques to address these challenges and bridge the gap in application.

**Bayesian Incorporating Priors**

**Incorporating association networks as prior information.** To overcome the sample size limitation, we propose to incorporate prior knowledge to direct LiNGAM, thereby reducing the searching space and thus largely accelerating the convergence rate to true graph and improving the computation efficiency. Reid et al. (2019) has highlighted that generalizing insights from the FC studies measured by statistical associations can facilitate the inference of conditional dependency. Current FC methods mainly employ three types of association measures, namely, the Pearson correlation (Rasero et al., 2023), partial correlation (Zhang, Cai, et al., 2020), and distance correlation (Xiao et al., 2022). Pearson correlation describes the linear correlation of a pair of variables, partial correlation measures the association between two variables removing the effect of other variables, and distance correlation can reflect nonlinear associations between two variables. Within a linear model, a partial correlation-based approach is more appropriate to incorporate as the prior than the Pearson correlation. This is because in a complex system like the brain network, the Pearson correlation is much weaker marginally (Liang, Song, & Qiu, 2015), that is, all nodes (variables) are directly or indirectly correlated, making it difficult to distinguish significant connections from a dense network. On the contrary, partial correlations can explore direct associations between two nodes, controlling for the confounding variables, which can facilitate the inference of influential directions. In our previous study (Zhang, Zhang, et al., 2020), we incorporated the partial correlation network using the \( \psi \)-learning method (Liang et al., 2015) as a prior with LiNGAM and showed its superiority in both of convergence and accuracy.

**Bayesian joint prior estimation for multiple groups.** We proposed the joint Bayesian-incorporating \( \psi \)-learning to address heterogeneity across multiple groups, which consisted of three steps: Step 1, Gaussian transformation; Step 2, distinct and common graph construction; and Step 3, prior matrices acquisition. Many partial correlation–based methods are designed for Gaussian distributed data due to their mathematical simplicity (Zhang, Fang, Liang, Calhoun, & Wang, 2019); Liu, Lafferty, and Wasserman (2009) have proposed a nonparanormal transformation, which relaxes the Gaussian assumption to any continuous one, which is therefore applied in Step 1. For Step 2, we consider both the distinct and common structure in each group. The distinct graph estimation
A BiLiNGAM method to study emotion circuit development

Bayesian prior:
In Bayesian statistics, a prior probability distribution, often simply called a prior, represents the state of knowledge or belief about the parameters of interest before observing any data.

for each group is implemented through the \( \psi \)-learning method as in Zhang, Fang, et al. (2019). To strengthen the similarities over various groups, we adopt the same Bayesian incorporating joint estimation method as in our previous study (Zhang, Cai, et al., 2020). The similarities are highlighted through proper Bayesian priors and a meta-analysis procedure. Finally, we use the union of the distinct estimated graph and joint estimated graph as the prior for each group.

**BiLiNGAM**

We now illustrate the proposed BiLiNGAM to jointly estimate multiple DAGs for non-Gaussian data. First, we apply the joint Bayesian-incorporated \( \psi \)-learning to acquire FC networks as the prior information. For each group \( k, k = 1, \ldots, K \), we obtain its distinctly estimated edge structure \( E_{\text{d}, k} \) through \( \psi \)-learning and its jointly estimated edge structure with strengthened group similarities \( E_{\text{c}, k} \) through the Bayesian incorporating joint estimation from the observations \( \mathbf{X}^k \in \mathbb{R}^{n_k \times p} \), where \( n_k \) represents the number of samples in group \( k \) and \( p \) represents the number of variables. We then introduce the prior matrix \( A_{\text{prior}} \) to represent the potential directed edge space, where a value of 0 indicates no directed edge, and \(-1\) indicates uncertainty about the edge status. The prior matrix \( A_{\text{prior}, k} = \{d_{ij}^{\text{prior}, k}\} \) of group \( k \) combines the edge structures from \( E_{\text{d}, k} \) and \( E_{\text{c}, k} \) which is defined as

\[
d_{ij}^{\text{prior}, k} = \begin{cases} 0, & (i, j) \notin E_{\text{d}, k} \cup E_{\text{c}, k} \\ -1, & \text{otherwise.} \end{cases}
\]

Next we leverage the prior matrix \( A_{\text{prior}, k} \) with the direct LiNGAM to estimate the weighted adjacency matrix of the DAG for group \( k \), that is, \( \mathbf{B}^k \). The detailed procedure is summarized in Algorithm 1 with code available at https://github.com/Aiying0512/BiLiNGAM.

**Simulation Data Generation**

We evaluated the performance of the joint estimation of \( K \) different DAGs where we varied \( K \in \{3, 5\} \). For all experiments, we set the number of nodes \( p = 200 \) and the total number of observations \( N = 750 \). For each group, we set the number of samples equally, that is, \( n_1 = n_2 = \ldots = n_K = n = N/K \). The random DAG \( G \) is simulated through the R package pcalg, and density of the graph is controlled by the edge probability \( d/(p - 1) \), where \( d \) is the mean edge degree parameter with values \( \{1, 2, 5\} \). The true DAGs generation procedure is illustrated as follows. We first used the pcalg, to generate \( G^1 \). Given \( G^1 \), we assigned uniformly random weights to the edges to obtain the weighted adjacency matrix \( \mathbf{B}^1 = (b_{ij}^1) \): \( b_{ij}^1 \sim \text{Unif}(-0.8, -0.3) \cup (0.3, 0.8) \), if there is an edge \( i \rightarrow j \), otherwise \( b_{ij}^1 = 0 \). For \( G^k, k = 2, 3, \ldots, K \), we followed the same random edge deleting-adding procedure in a sequential manner. We randomly removed 5% edges in \( G^{k-1}, k = 2, \ldots, 5 \), by setting the corresponding nonzero elements in \( \mathbf{B}^k \) to be 0, and then added 5% edges at random by giving them values drawn from the uniform distribution \( U[0.3, 0.5] \) to obtain \( \mathbf{B}^k \). Given \( \mathbf{B}^k \)'s, we generated \( \mathbf{X}^k = (\mathbf{B}^k)^T \mathbf{X} + \mathbf{e}^k \in \mathbb{R}^p \) with \( \mathbf{e}^k \) from chi-squared (Chisq) noise with degree of freedom 1 and zero mean, that is, \( e_i^k \sim \chi_1^2, i = 1, 2, \ldots, p, k = 1, 2, \ldots, K \).

We considered four methods for comparison, which are the PC algorithm (Spirtes et al., 2000), the IMaGES (Ramsey et al., 2010), the ICA-LiNGAM (Shimizu et al., 2006), and the \( \psi \)-LiNGAM that we proposed previously (Zhang, Zhang, et al., 2020). The PC and ICA-LiNGAM were implemented through the R package pcalg, and the IMaGES was implemented through the Tetrad toolbox. The codes for \( \psi \)-LiNGAM and BiLiNGAM are available at https://github.com/Aiying0512. We set the significance level \( \alpha = 0.05 \) with FDR correction for the PC, \( \psi \)-LiNGAM and BiLiNGAM.
For each scenario, 10 datasets were simulated independently. We assessed the performances of the four methods through the true positive rate (TPR), false discovery rate (FDR), and structural hamming distance (SHD) (Tsamardinos, Brown, & Aliferis, 2006). TPR and FDR are two common measures of binary classification. Let us define an experiment from $P$ positive instances and $N$ negative instances for some conditions. In our case, the positive instance represents a directed edge from one node to the other. The four outcomes are summarized in Table 1. The definitions of TPR and FDR are given as follows:

$$TPR = \frac{TP}{TP + FN}, \quad FDR = \frac{FP}{FP + TP}.$$  

SHD is a frequently used metric based on the number of operations needed to transform the estimated DAG into the true graph (Kalisch & Bühlmann, 2007). In simple terms, SHD counts the total number of edge insertions, deletions, or flips during the transformation.

---

**Algorithm 1. BiLiNGAM algorithm**

**Input:** Collection of observation $X^k = (X_i^k) \in \mathbb{R}^{n_k \times p}$, where $k = 1, 2, \ldots, K$, $i = 1, 2, \ldots, p$ and $X_i^k$'s are non-Gaussian continuous.

**Output:** Collection of estimated weighted adjacency matrices $\hat{B}^k$.

1. Prior estimation: joint Bayesian-incorporating $\psi$-learning.
   **Start:**
   
   a. For $k = 1, 2, \ldots, K$, use the nonparanormal transformation (Liu et al., 2009) to render $X^k$ normal (Gaussian).
   b. Apply the $\psi$-learning method (Liang et al., 2015) to each group $k$, $k = 1, 2, \ldots, K$ separately for distinct estimation and acquire the adjacency matrix $E^d_k$.
   c. Apply the Bayesian incorporating joint estimation (Zhang, Padmanabhan, et al., 2019) to strengthen the similarities among the groups and acquire the $E^c_k$, $\forall k$.
   d. Extract the prior matrix $A_{prior}^k$ from $E_{prior}^k = E^c_k \cup E^d_k$ where $a_{ij}^{prior,k} = -1$ if $e_{ij}^{prior,k} = 1$ otherwise $a_{ij}^{prior,k} = 0$.
   **End**

2. Obtain the estimated weighted DAG adjacency matrices $\hat{B}^k$: LiNGAM.
   **Start:** For each $k$
   
   a. Identify the casual order $x^k$ using the direct LiNGAM with the prior matrix $A_{prior}^k$ (Shimizu et al., 2011).
   b. Construct a strictly lower triangular matrix $\tilde{B}^k$ by following the order $x^k$, and the corresponding $\tilde{A}_{prior}^k$ with the same order.
   c. Estimate the connection strengths $\tilde{B}_{ij}^k = (b_{1j}^k, b_{2j}^k, ..., b_{pj}^k)$ consistent with $\tilde{A}_{prior}^k$ by solving sparse regressions of the form
   $$\tilde{B}_{ij}^k = \arg \min_{\tilde{B}_{ij}^k \in \text{supp}(\tilde{A}_{prior}^k)} \|X^k - X^k \tilde{B}_{ij}^k\|_2$$
   d. Obtain $\hat{B}^k$ by converting $\tilde{B}^k$ to the original order.
   **End**

---

For each scenario, 10 datasets were simulated independently. We assessed the performances of the four methods through the true positive rate (TPR), false discovery rate (FDR), and structural hamming distance (SHD) (Tsamardinos, Brown, & Aliferis, 2006). TPR and FDR are two common measures of binary classification. Let us define an experiment from $P$ positive instances and $N$ negative instances for some conditions. In our case, the positive instance represents a directed edge from one node to the other. The four outcomes are summarized in Table 1. The definitions of TPR and FDR are given as follows:

$$TPR = \frac{TP}{TP + FN}, \quad FDR = \frac{FP}{FP + TP}.$$  

SHD is a frequently used metric based on the number of operations needed to transform the estimated DAG into the true graph (Kalisch & Bühlmann, 2007). In simple terms, SHD counts the total number of edge insertions, deletions, or flips during the transformation.
The Philadelphia Neurodevelopmental Cohort (PNC) dataset consists of fMRI images under emotion identification task from 855 individuals. The age range of the participating subjects was between 8 and 22 years. Due to physical and cognitive changes (Zhang, Cai, et al., 2020), we divided them into five groups, each representing a period related to adolescence (see Table 2).

**Imaging acquisition and preprocessing.** All MRI scans were acquired on a single 3T Siemens TIM Trio whole-body scanner. During the task, each subject was asked to label emotions displayed, which include happy, angry, sad, fearful, and neutral faces. The total scan duration was 10.5 min. Blood oxygenation level–dependent (BOLD) fMRI was acquired using a whole-brain, single-shot, multislice, gradient-echo (GE) echoplanar (EPI) sequence of 124 volumes (372 s) with the following parameters TR/TE = 3,000/32 ms, flip = 90°, FOV = 192 × 192 mm, matrix = 64 × 64, slice thickness/gap = 3 mm/0 mm. The resulting nominal voxel size was 3.0 × 3.0 × 3.0 mm (Satterthwaite et al., 2014). Standard preprocessing steps were applied using SPM12, including motion correction, spatial normalization to standard MNI space, and spatial smoothing with a 3-mm full width at half max (FWHM) Gaussian kernel. Then multiple regression considering the influence of motion was performed and the stimulus on-off contrast maps for each subject were obtained. Finally, 264 functionally defined regions of interest (ROIs) were extracted based on the power template (Power, Fair, Schlaggar, & Petersen, 2010). The 264 ROIs (nodes) can be divided into 12 functional network (FN) modules as provided by Power et al. (2010) (see Figure 1), including sensory/somatomotor network (SSN), cingulo-opercular task control network (CON), auditory network (AUD), default mode network (DMN), memory retrieval network (MRN), visual network (VN), fronto-parietal task control network (FPN), salience network (SN), subcortical network (SCN), ventral attention network (VAN), dorsal attention network (DAN), and cerebellum network (CERE).

**Analytical pipeline to understand emotion circuit development.** As shown in Figure 2, we investigate the development of emotion-related dFC from two aspects: (1) comparing the common and distinctive dFC patterns; (2) extracting significant network features in terms of functional

| Category index | Group name        | Age range | Number of subjects |
|----------------|-------------------|-----------|--------------------|
| 1              | Preadolescence    | 8–12      | 194                |
| 2              | Early adolescence | 12–14     | 150                |
| 3              | Middle adolescence| 14–16     | 158                |
| 4              | Late adolescence  | 16–18     | 166                |
| 5              | Postadolescence   | 18–22     | 187                |

---

**Table 1.** Outcomes of a binary decision

| Predicted positive | Actual positive (P) | Actual negative (N) |
|--------------------|---------------------|---------------------|
| True positive (TP) | False positive (FP) |
| False positive (FN)| True negative (TN)  |

**Analyses on the PNC Dataset**

The Philadelphia Neurodevelopmental Cohort (PNC) dataset consists of fMRI images under emotion identification task from 855 individuals. The age range of the participating subjects was between 8 and 22 years. Due to physical and cognitive changes (Zhang, Cai, et al., 2020), we divided them into five groups, each representing a period related to adolescence (see Table 2).
network modules and hub ROIs. The intra- and intermodular connectivity is calculated by counting the number of edges module-wise. A hub ROI refers to the region with a large number of edges, or a high degree, in the brain network. Here we define hubs as the nodes with degrees at least two standard deviations higher than the average degrees within the DAG (Fang et al., 2017). For a DAG, the degree of a node is composed of two elements: in-degree and out-degree, which are named according to the direction of the edges. Based on the types of degrees, we identified the hub structures: in-hub as the center to receive information, and out-hub as the center to send information.

![Figure 1](image1.png)  
**Figure 1.** Twelve functional network modules based on the 264 nodes from the template defined by Power et al. (2010).

![Figure 2](image2.png)  
**Figure 2.** Schematic illustration of the analysis pipeline with the proposed BiLiNGAM model.
RESULTS

Simulation Results

Figure 3A gives the results for $K = 3$, where each group has $n = 250$ samples and $p = 200$ nodes. We compared the results of all five methods with various graph densities controlled by the mean edge degree parameter $d$. The simulation results reveal the limited edge orientation ability of the PC and IMaGES. While the TPR curve of ICA-LiNGAM demonstrates reasonable performance, the high FDR and SHD values indicate that ICA-LiNGAM requires a larger number of observations to converge to the true graph. Both methods incorporating association networks as prior information ($\psi$-LiNGAM and BiLiNGAM) exhibit similarly strong performance. This suggests that when the sample size is sufficiently large ($n > p$), incorporating associations into LiNGAM models can significantly improve causal inferences.

Further, we compared the two methods under high-dimensional cases (i.e., $n < p$). Figure 3B shows the results for $K = 5$, where each group has $n = 150$ samples. The FDR curves of the two methods maintain at the same low level. The TPR of BiLiNGAM is always higher than that of $\psi$-LiNGAM, while the SHD performs the opposite. In addition, as the mean edge degree parameter $d$ increases, the differences in TPR and SHD also increase. Therefore, under high dimensional settings, both methods remain at a low FDR level, but BiLiNGAM outperforms $\psi$-LiNGAM in terms of TPR and SHD.

![Simulation Results](image)

Figure 3. Simulation results using various estimation methods with different network densities by varying the mean edge degree parameter $d \in \{1, 2, 5\}$. The average performance in terms of TPR (left), FDR (middle) and SHD (right) is presented for the setting of $N = 750$ total observations for $K$ groups with $p = 200$ variables, where the number of groups $K \in \{3, 5\}$ with equal observations in each group $n = N/K$. 

Network Neuroscience
Overall, BiLiNGAM has maintained a stable and accurate performance over various settings. Specifically, under high-dimensional cases, the performance of BiLiNGAM is superior. When the sample size is adequate (i.e., \( n > p \)), BiLiNGAM performs at least as good as \( \psi \)-LiNGAM. More comprehensive simulation studies are presented in the Supporting Information section A.

fMRI Results to Reveal the Emotion Circuit Development

We first conducted the Darling-Anderson test for non-Gaussianity and then applied BiLiNGAM. The parameter settings remain the same as in the simulation studies.

**Common and distinctive dFC patterns across the groups.** We extracted the common and distinctive directed connectivity patterns over the five groups as shown in Figure 4. As the brain develops, there is an increase in connections between the limbic system, responsible for emotion processing, and the prefrontal cortex, involved in higher order cognitive functions. Additionally, bilateral connections in the parietal and occipital lobes become more pronounced, supporting enhanced sensory integration and spatial awareness. These changes in connectivity facilitate the identification and regulation of emotions as part of the brain’s maturation process.

**Development of emotion-related intra- and intermodular connectivity.** We examined intramodular and intermodular connectivity over the five adolescent groups. In the Supporting Information section B, we visualized the average directed edge degrees within and across modules, where the rows indicate the beginning of the arrows and the columns indicate the end of the arrows.

![Figure 4](image-url)  
**Figure 4.** The common (A) and distinctive (B) dFC patterns of the five adolescent stages, where the arrows indicate the direction of the FC’s. (A) Left: the common directed FC in the axial view of the brain; Right: the common directed FC in network view, where the nodes are sorted by the lobes and separated into left (left half-circle) and right (right half-circle) hemispheres. (B) The distinctive dFC patterns of each adolescent groups in network view.
From the heat maps, the intramodular connectivity of DMN, SCN, and CERE is strongly activated for all five groups. As age increases, there is increasing intermodular connectivity. We then conducted hypergeometric tests based on the number of edges module-wise and the significant intra- and interdirected connections are shown in Figure 5 at significance level $\alpha = 0.05$ with FDR correction. The intramodular connectivity of the CERE was significantly activated of all adolescent groups for the emotion identification task, while the role of the SCN was only significant until the middle adolescent period. In addition, we found substantial intramodular connectivity of SCN in the early adolescent group and CON in the late adolescent group. From the aspect of intermodular connectivity, no significant directed flows were found in the preadolescent group, one each was identified for the early (VAN $\rightarrow$ CON) and middle (DAN $\rightarrow$ SSN) adolescent groups, five were identified for the late adolescent group (CON $\rightarrow$ SSN, VAN $\rightarrow$ SSN, VAN $\rightarrow$ AUD, DMN $\rightarrow$ MRN, DAN $\rightarrow$ CERE), and four directed flows were discovered in the postadolescent group (CON $\rightarrow$ SSN, DMN $\rightarrow$ MRN, CON $\rightarrow$ SN, FPN $\rightarrow$ SN).

Development of emotion-related hubs. To gain more insights into the affective emotion circuits and their development with age, we analyzed hub nodes for each group. Figure 6A gives the in-hub development and the detailed in-hub information. The ROI at SMA.L has a preadolescence specific pattern. The ROIs at PQ.R and ACG.R have increased activities of receiving messages, especially, the ROI at ACG.R only starts to develop from middle adolescence. The remaining in-hubs have fluctuating trajectories. From Figure 6B, several out-hubs start to develop in a fluctuating manner after preadolescence, whose anatomical locations are at AMY.L, SFG.L (ROI 9, 10), PG.R, PQ.R, LG.R, Q.R, MCG.R, IFGT.R, VA.R, STG.R, and MTG (ROI 23, 24). Some group-specific patterns have also been detected: the ROIs at SFGM.R for preadolescent group; the ROIs at IFGT.R and MTG.L for early adolescent group; the ROIs at SFG.L, MCG.R, and VA.R for middle adolescent group; the ROIs at MCG.L (ROI: 5, 8), and STG.R for late adolescent group; and the ROIs at AMY.L and SFG.R for postadolescent group.

Robustness of the findings regarding sample size. We examined the robustness of our proposed method and assessed the effect of sample size on the robustness. We randomly drew $m$ percent of $N$ participants ($m \in \{20\%, 50\%, 80\%, 90\%\}$, $N = 855$) from the PNC dataset sample without replacement and with the proportion of the five adolescent groups. We then applied the BiLiNGAM method to the subsamples. For each sample size, we repeated the procedure 10 times. In Figure 7, we showed the mean number of edges detected for each adolescent group under various sample sizes. From preadolescence to postadolescence, the trajectory of each sample size remains similar. However, we found that limited edges were identified with small sample size ($m = 20\%$). As the sample size increases, the identified number of edges for each group remains steady and the variance decreases significantly.
Comparison with other dFC methods. To prove the effectiveness and reliability of the proposed method, we compared the dFC estimated by BiLiNGAM, $\psi$-LiNGAM (Zhang, Zhang, et al., 2020) and the PC-algorithm (Spirtes et al., 2000) using the fMRI data from the PNC study. We calculated the Jaccard index as in Zhang, Padmanabhan, et al. (2019) to quantify the graph robustness of BiLiNGAM with various percentages (20%, 50%, 80%, 90%) of samples ($N = 855$) from the PNC dataset. Data were mean number of edges ± standard deviation.

Figure 6. The developing trajectories of the identified in-hubs (A) and out-hubs (B), where the locations of the ROIs in the brain are visualized in the left. The color map of the ROIs represent the functional networks. The abbreviated name of the ROIs are provided and the detailed information can be found in the Supporting Information section B. The index of developing trajectories on the right correspond to the label on the left.

Figure 7. The robustness of BiLiNGAM with various percentages (20%, 50%, 80%, 90%) of samples ($N = 855$) from the PNC dataset. Data were mean number of edges ± standard deviation.
similarities among different adolescent groups. The Jaccard index is defined as the ratio of the intersection between two sets of edges divided by their union, ranging from 0 (no overlap) to 1 (full overlap). As shown in Table 3 (the columns of among groups), the BiLiGAM method has significantly reinforced the group similarities (specially when compared with $\psi$-LiNGAM).

We then compared the method-wise similarities (see Table 3, columns of among methods). Fisher’s exact test was applied to examine the independence of the dFC’s estimated by the three methods. The $p$ values were all lower than $10^{-25}$, indicating that the estimated dFC from the three methods are significantly correlated. The comparison with PC algorithm shows lower Jaccard index. This discrepancy can be attributed to the limitations of the PC algorithm in accurately identifying the orientation of edges. As a result, the PC algorithm often produces graphs with numerous undirected edges, which may lead to differences in the estimated graphs compared to our proposed method. The $\psi$-LiNGAM method exhibits higher similarities with BiLiNGAM. However, these similarities are less consistent across groups as indicated by the variance of Jaccard index. This inconsistency is due to sample size limitations, as demonstrated in the simulation studies.

**DISCUSSION**

In this paper, we propose a multiple DAG estimation method for non-Gaussian data, BiLiNGAM, and further apply it to the study of brain development. The main contributions of our work can be summarized in three primary points. First, from a mathematical perspective, we relaxed the assumption of common ordering from the joint LiNGAM (Shimizu, 2012) and made it possible to estimate larger graphs. Second, our proposed method integrated undirected and directed graphs, by incorporating the undirected graph estimation as prior information into the direct LiNGAM model for better DAG construction. We use the undirected graphs to mitigate the irrelevant information for better casual inferences, in addition to optimizing convergence and computation. Third, the simulation results show that BiLiNGAM maintains a stable and accurate performance over various settings. In particular, the proposed BiLiNGAM is superior for high-dimensional cases. Finally, the analysis of brain’s emotion circuit development revealed the trajectory of directed brain circuitry during emotion identification tasks over various adolescent groups. We identified several significant intra- and intermodular networks that change over developmental stage, and pinpointed emotion-related hubs as well as various group-specific patterns.

**Intramodular Development**

We found a developmentally stable intramodular activation anchored in the default mode (DMN), subcortical (SCN), and cerebellum (CERE) networks. The default mode network is important for mentalizing and inferring emotional states of others (Blakemore, 2008); subcortical regions have a pivotal role in cognitive, affective, and social functions in humans.
(Koshiyama et al., 2018); the cerebellum contributes prominently in processing emotional facial expression (Ferrucci et al., 2012). The intramodular activities of CERE increased significantly in the postadolescent group, which may emerge in late puberty (Tiemeier et al., 2010). The significance of intramodular activities of the SCN appeared from preadolescence to middle adolescence, which has been proven to be an important developmental period for subcortical brain maturation (Dennison et al., 2013).

**Intermodular Development**

As age increases, more intermodular connectivity emerges during emotion-related processing. Starting from early adolescence, interconnections start to build among VAN, CON, SN, and SSN. Specifically, SSN emerges with significant interconnections that receives information from other functional networks after middle adolescence. VAN plays an important role in conveying information in late adolescent group, and SN is crucial for receiving information in postadolescent group. Two stable directed influences from CON to SSN, and from DMN to MRN become established after the late adolescent period. CON facilitates the maintenance of task-relevant goals and the incorporation of error information to adjust behaviors (Cocchi, Zalesky, Fornito, & Mattingley, 2013) and SSN (including somatosensory cortex, motor regions, and supplementary motor areas) is involved in performing and coordinating motor-related tasks like finger tapping. Gehringer et al. (2019) proposed that maturation of the somatosensory system during adolescence contributes to improved motor control. They further discovered that altered attenuation of the somatosensory cortical oscillations might be central to the underdeveloped somatosensory processing and motor performance characteristics in adolescents. Our results agree with their conclusion and provide additional explanations. Besides, Sestieri, Corbetta, Romani, and Shulman (2011) confirmed that responses in DMN regions peaked sooner than non-DMN regions during memory retrieval, and the parietal regions of DMN directly supported memory retrieval.

**Emotion-Related Hubs**

Most identified hubs are located in the right hemisphere, which are dominant in the perception of facial expression and important for processing primary emotions (Alfano & Cimino, 2008). Particularly, the hubs at PG, MCG, PQ, and MTG play central roles in socioemotional processing. The precentral gyrus (PG) of the somatosensory cortex is related to the recognition of facial and vocal expressions of emotion and a main effect of emotional valence on brain activity has been found in the PG.R (Seo et al., 2014). A previous study (Shackman et al., 2011) verified that the mid-cingulate gyrus (MCG) is a hub linking incoming affective information with brain regions involved in goal-directed behavior, and we further discovered that it is also a hub for distributing affective information. Precuneus (PQ) activation has been implicated in emotional and memory-loaded processes. The current study suggests that the PQ may play a direct role in the regulation of amygdala reactivity to emotional stimuli (Ferri, Schmidt, Hajcak, & Canli, 2016), which explains its prominence as a out-hub location. Studies of emotional face recognition (Haxby, Hofman, & Gobbini, 2002; Sabatinelli et al., 2011) identified the middle temporal gyrus (MTG) as a primary neural substrate for suprathreshold processing of the emotional expression of faces, which is consistent with our result of MTG as a central node to pass out information.

**Developmental Hub Structures and Group-Specific Hub Patterns**

The majority of networks during development fluctuate, except for the steady increase of the in-hub activities at the PQ.R and ACG.R. Group-specific patterns have also been identified: the in-hub at SMA.L and out-hub at SFGM.R for preadolescence; the out-hubs at IFGT.R and
MTG.L for early adolescence; the out-hubs at SFG.L, MCG.R, and VAR for middle adolescence; the out-hubs at MCG.L and STG.R for late adolescence; the out-hubs at AMY.L and SFG.R for postadolescence. Some of our results have been previously supported in the literature. The role of developmental centers at PG, PQ, ACG, LING, and PHIG remains consistent with our previous study of brain connectivity development in adolescence (Zhang, Padmanabhan, et al., 2019). In this study, we further pinpoint their specific functions in the emotion circuit through directed graphical models. Another study of brain development from adolescence to adulthood (Kundu et al., 2018) also brought attention to age-related changes in the PQ. In Simmonds, Hallquist, and Luna (2017), fluctuating trajectories in the MCG during adolescence were discovered.

Our findings provide a directed functional connectivity template of emotion processing in the developing brain, thereby shedding light on the understanding of brain networks underlying emotion processing behaviors.

SUPPORTING INFORMATION

Supporting information for this article is available at https://doi.org/10.1162/netn_a_00384.

AUTHOR CONTRIBUTIONS

Aiying Zhang: Conceptualization; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing – original draft; Writing – review & editing. Gemeng Zhang: Conceptualization; Investigation; Methodology; Writing – review & editing. Biao Cai: Conceptualization; Visualization; Writing – review & editing. Tony W. Wilson: Funding acquisition; Supervision; Writing – review & editing. Vince D. Calhoun: Funding acquisition; Supervision; Writing – review & editing. Yu-Ping Wang: Data curation; Funding acquisition; Resources; Supervision; Writing – review & editing.

FUNDING INFORMATION

Tony W. Wilson, National Institute of Mental Health (https://dx.doi.org/10.13039/10.2044/0025), Award ID: R01MH121101. Yu-Ping Wang, National Institute of Mental Health (https://dx.doi.org/10.13039/10.2044/0025), Award ID: R01MH104680. Yu-Ping Wang, National Institute of Mental Health (https://dx.doi.org/10.13039/10.2044/0025), Award ID: R01MH107354. Vince D. Calhoun, National Institute of Mental Health (https://dx.doi.org/10.13039/10.2044/0025), Award ID: R01MH103220. Vince D. Calhoun, National Institute of General Medical Sciences (https://dx.doi.org/10.13039/10.2044/0057), Award ID: P20GM109068. Tony W. Wilson, National Institute of General Medical Sciences (https://dx.doi.org/10.13039/10.2044/0057), Award ID: P20GM144641. Vince D. Calhoun, National Institute of Biomedical Imaging and Bioengineering (https://dx.doi.org/10.13039/10.2044/0070), Award ID: R01EB020407. Yu-Ping Wang, National Institute of Mental Health (https://dx.doi.org/10.13039/10.2044/0025), Award ID: R56MH124925.

REFERENCES

Alfano, K. M., & Cimino, C. R. (2008). Alteration of expected hemispheric asymmetries: Valence and arousal effects in neuropsychological models of emotion. Brain and Cognition, 66(3), 213–220. https://doi.org/10.1016/j.bandc.2007.08.002, PubMed: 17928118
Blakemore, S.-J. (2008). The social brain in adolescence. Nature Reviews Neuroscience, 9(4), 267–277. https://doi.org/10.1038/nrn2353, PubMed: 18354399
Chickering, D. (2002). Optimal structure identification with greedy search. Journal of Machine Learning Research, 3, 507–554. https://doi.org/10.1162/153244303321897717
Cocchi, L., Zalesky, A., Fornito, A., & Mattingley, J. B. (2013). Dynamic cooperation and competition between brain systems during cognitive control. Trends in Cognitive Sciences, 17(10), 493–501. https://doi.org/10.1016/j.tics.2013.08.006, PubMed: 24021711
A BiLiNGAM method to study emotion circuit development

Dennison, M., Whittle, S., Yücel, M., Vijayakumar, N., Kline, A., Simmons, J., & Allen, N. B. (2013). Mapping subcortical brain maturation during adolescence: Evidence of hemisphere- and sex-specific longitudinal changes. *Developmental Science, 16*(5), 772–791. https://doi.org/10.1111/desc.12057, PubMed: 24033581

Fang, J., Xu, C., Zille, P., Lin, D., Deng, H.-W., Calhoun, V. D., & Wang, Y.-P. (2017). Fast and accurate detection of complex imaging genetics associations based on greedy projected distance correlation. *IEEE Transactions on Medical Imaging, 37*(4), 860–870. https://doi.org/10.1109/TMI.2017.2783244, PubMed: 29990017

Ferri, J., Schmidt, J., Hajcak, G., & Canli, T. (2016). Emotion regulation and amygdala-precuneus connectivity: Focusing on attentional deployment. *Cognitive, Affective, & Behavioral Neuroscience, 16*(6), 991–1002. https://doi.org/10.3758/s13415-016-0447-y, PubMed: 27444935

Ferrucci, R., Giannicola, G., Rosa, M., Fumagalli, M., Boggio, P. S., Hallett, M., … Priori, A. (2012). Cerebellum and processing of negative facial emotions: Cerebellar transcranial DC stimulation specifically enhances the emotional recognition of facial anger and sadness. *Cognition and Emotion, 26*(5), 786–799. https://doi.org/10.1080/02699931.2011.619520, PubMed: 22077643

Gee, D. G., Hanson, C., Caglar, L. R., Fareri, D. S., Gabard-Durnam, L. J., Mills-Finnerty, C., … Tottenham, N. (2022). Experimental evidence for a child-to-adolescent switch in human amygdala-prefrontal cortex communication: A cross-sectional pilot study. *Developmental Science, 25*(4), e13238. https://doi.org/10.1111/desc.13238, PubMed: 35080089

Gehringer, J. E., Arpin, D. J., VerMaas, J. R., Trevarrow, M. P., Wilson, T. W., & Kurz, M. J. (2019). The strength of the movement-related somatosensory cortical oscillations differ between adolescents and adults. *Scientific Reports, 9*(1), 18520. https://doi.org/10.1038/s41598-019-55004-1, PubMed: 31811232

Hanson, C., Hanson, S. J., Ramsey, J., & Glymour, C. (2013). Atypical effective connectivity of social brain networks in individuals with autism. *Brain Connectivity, 3*(6), 578–589. https://doi.org/10.1089/brain.2013.0161, PubMed: 24093627

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Journal of the American Statistical Association, 97*, 67–73. https://doi.org/10.1198/016214502753478551

Henry, T., & Gates, K. (2017). Causal search procedures for fMRI: Review and suggestions. *Behaviormetrika, 44*(1), 193–225. https://doi.org/10.1007/s10223-016-0010-8

Iyer, S. P., Shafran, I., Grayson, D., Gates, K., Nigg, J. T., & Fair, D. A. (2013). Inferring functional connectivity in MRI using Bayesian network structure learning with a modified PC algorithm. *NeuroImage, 75*, 165–175. https://doi.org/10.1016/j.neuroimage.2013.02.054, PubMed: 23501054

Kalisch, M., & Bühlmann, P. (2007). Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *Journal of Machine Learning Research, 8*, 613–636.

Koshiyama, D., Fukunaga, M., Okada, N., Yamashita, F., Yamamori, H., Yasuda, Y., … Hashimoto, R. (2018). Role of subcortical structures on cognitive and social function in schizophrenia. *Scientific Reports, 8*(1), 1183. https://doi.org/10.1038/s41598-017-18950-2, PubMed: 29352126

Kundu, P., Benson, B. E., Rosen, D., Frangou, S., Leibenuft, E., Luh, W.-M., … Ernst, M. (2018). The integration of functional brain activity from adolescence to adulthood. *Journal of Neuroscience, 38*(14), 3559–3570. https://doi.org/10.1523/JNEUROSCI.1864-17.2018, PubMed: 29487126

Liang, F., Song, Q., & Qiu, P. (2015). An equivalent measure of partial correlation coefficients for high-dimensional gaussian graphical models. *Journal of the American Statistical Association, 110*, 1248–1265. https://doi.org/10.1080/01621459.2015.1012391

Liu, H., Lafferty, J., & Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research, 10*, 2295–2328.

Liu, Y., Wu, X., Zhang, J., Guo, X., Long, Z., & Yao, L. (2015). Altered effective connectivity model in the default mode network between bipolar and unipolar depression based on resting-state fMRI. *Journal of Affective Disorders, 182*, 8–17. https://doi.org/10.1016/j.jad.2015.04.009, PubMed: 25942576

Oates, C. J., Smith, J. Q., Mukherjee, S., & Cussens, J. (2016). Exact estimation of multiple directed acyclic graphs. *Statistics and Computing, 26*(4), 797–811. https://doi.org/10.1007/s11222-015-9570-9

Power, J. D., Fair, D. A., Schlaggar, B. L., & Petersen, S. E. (2010). The development of human functional brain networks. *Neuron, 67*(5), 735–748. https://doi.org/10.1016/j.neuron.2010.08.017, PubMed: 20826306

Ramsey, J., Sanchez-Romero, R., & Glymour, C. (2014). Non-Gaussian methods and high-pass filters in the estimation of effective connections. *NeuroImage, 84*, 986–1006. https://doi.org/10.1016/j.neuroimage.2013.09.062, PubMed: 24099845

Ramsey, J. D., Hanson, S. J., & Glymour, C. (2011). Multi-subject search correctly identifies causal connections and most causal directions in the DCM models of the Smith et al. simulation study. *NeuroImage, 58*(3), 838–848. https://doi.org/10.1016/j.neuroimage.2011.06.068, PubMed: 21745580

Ramsey, J. D., Hanson, S. J., Hanson, C., Halchenko, Y. O., Poldrack, R. A., & Glymour, C. (2010). Six problems for causal inference from fMRI. *NeuroImage, 49*(2), 1545–1558. https://doi.org/10.1016/j.neuroimage.2009.08.065, PubMed: 19747552

Rasero, J., Jimenez-Marin, A., Diez, I., Toro, R., Hasan, M. T., & Cortes, J. M. (2023). The neurogenetics of functional connectivity alterations in autism: Insights from subtyping in 657 individuals. *Biological Psychiatry, 94*(10), 804–813. https://doi.org/10.1016/j.biopsych.2023.04.014, PubMed: 37088169

Reid, A. T., Headley, D. B., Mill, R. D., Sanchez-Romero, R., Uddin, L. Q., Marinazzo, D., … Cole, M. W. (2019). Advancing functional connectivity research from association to causation. *Nature Neuroscience, 22*(11), 1751–1760. https://doi.org/10.1038/s41593-019-0510-4, PubMed: 31617005

Sabatinelli, D.,fortune, E. E., Li, Q., Siddiqui, A., Krafft, C., Oliver, W. T., … Jeffries, J. (2011). Emotional perception: Meta-analyses of face and natural scene processing. *NeuroImage, 54*(3), 2524–2533. https://doi.org/10.1016/j.neuroimage.2010.10.011, PubMed: 20951215

Satterthwaite, T. D., Elliott, M. A., Ruparel, K., Loughead, J., Prabhakaran, K., Calkins, M. E., … Gur, R. E. (2014). Neuroimaging of the Philadelphia Neurodevelopmental Cohort. *NeuroImage, 86*, 544–553. https://doi.org/10.1016/j.neuroimage.2013.07.064, PubMed: 23921101
A BiLiNGAM method to study emotion circuit development

Seo, D., Olman, C. A., Haut, K. M., Sinha, R., MacDonald, A. W., III, & Patrick, C. J. (2014). Neural correlates of preparatory and regulatory control over positive and negative emotion. *Social Cognitive and Affective Neuroscience, 9*(4), 494–504. https://doi.org/10.1093/scan/nst115, PubMed: 23887812

Sestieri, C., Corbetta, M., Romani, G. L., & Shulman, G. L. (2011). Episodic memory retrieval, parietal cortex, and the default mode network: Functional and topographic analyses. *Journal of Neuroscience, 31*(12), 4407–4420. https://doi.org/10.1523/JNEUROSCI.3335-10.2011, PubMed: 21430142

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., & Davidson, R. J. (2011). The integration of negative affect, and search & Giedd, J. N. (2010). Cerebellum development during childhood and adolescence: A longitudinal morphometric MRI study. *NeuroImage, 49*(1), 63–70. https://doi.org/10.1016/j.neuroimage.2009.08.016, PubMed: 19683586

Shimizu, S., Inazumi, T., Sogawa, Y., Hyvärinen, A., Kawahara, Y., Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, Tiemeier, H., Lenroot, R. K., Greenstein, D. K., Tran, L., Pierson, R., & Giedd, J. N. (2010). Cerebellum development during childhood and adolescence: A longitudinal morphometric MRI study. *NeuroImage, 49*(1), 63–70. https://doi.org/10.1016/j.neuroimage.2009.08.016, PubMed: 19683586

Zheng, X., Aragam, B., Ravikumar, P. K., & Xing, E. P. (2018). DAGs with no tears: Continuous optimization for structure learning. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (pp. 9492–9503).

Tsamardinos, I., Brown, L., & Aliferis, C. (2006). The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning, 65*, 31–78. https://doi.org/10.1007/s10994-006-6889-7

Wang, Y., Segarra, S., & Uhler, C. (2018). High-dimensional joint estimation of multiple directed Gaussian graphical models. *arXiv*. https://doi.org/10.48550/arXiv.1804.00778

Xiao, L., Cai, B., Qu, G., Zhang, G., Stephen, J. M., Wilson, T. W., Wang, Y.-P. (2022). Distance correlation-based brain functional connectivity estimation and non-convex multi-task learning for developmental fMRI studies. *IEEE Transactions on Biomedical Engineering, 69*(10), 3039–3050. https://doi.org/10.1109/TBME.2022.3160447, PubMed: 35316180

Zhang, A., Cai, B., Hu, W., Jia, B., Liang, F., Wilson, T. W., & Wang, Y.-P. (2020). Joint Bayesian-incorporating estimation of multiple Gaussian graphical models to study brain connectivity development in adolescence. *IEEE Transactions on Medical Imaging, 39*(2), 357–365. https://doi.org/10.1109/TMI.2019.2926667, PubMed: 31283500

Zhang, A., Fang, J., Liang, F., Calhoun, V. D., & Wang, Y.-P. (2019). Aberrant brain connectivity in schizophrenia detected via a fast Gaussian graphical model. *IEEE Journal of Biomedical and Health Informatics, 23*(4), 1479–1489. https://doi.org/10.1109/JBHI.2018.2854659, PubMed: 29994624

Zhang, A., Zhang, G., Cai, B., Hu, W., Xiao, L., Wilson, T. W., & Wang, Y.-P. (2020). Causal inference of brain connectivity from fMRI with μ-learning incorporated linear non-Gaussian acyclic model (μ-LiNGAM). *arXiv*. https://doi.org/10.48550/arXiv.2006.09536

Zhang, G., Cai, B., Zhang, A., Tu, Z., Xiao, L., Stephen, J. M., … Wang, Y.-P. (2022). Detecting abnormal connectivity in schizophrenia via a joint directed acyclic graph estimation model. *NeuroImage, 260*, 119451. https://doi.org/10.1016/j.neuroimage.2022.119451, PubMed: 35842099

Zhang, Y., Padmanabhan, A., Gross, J. J., & Menon, V. (2019). Development of human emotion circuits investigated using a big-data analytic approach: Stability, reliability, and robustness. *Journal of Neuroscience, 39*(36), 7155–7172. https://doi.org/10.1523/JNEUROSCI.0220-19.2019, PubMed: 31332001

Zhang, A., Aragam, B., Ravikumar, P. K., & Xing, E. P. (2018). DAGs with no tears: Continuous optimization for structure learning. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (pp. 9492–9503).