A Bayesian Incorporated Linear Non-Gaussian Acyclic Model (BiLiNGAM) for Multiple Directed Acyclic Graph Estimation with application to causal brain connectivity using fMRI

Aiying Zhang, Gemeng Zhang, Yu-Ping Wang
Tulane University

INTRODUCTION

- Directed acyclic graph (DAG) models, also known as Bayesian networks, are commonly used to model causal relationships in complex systems.
- Existing methods have focused on estimating a single directed graphical model. However, in many brain studies, we have data from related classes, such as different developmental stages and different disease states.
- Regarding statistical models, this corresponds to jointly estimating multiple DAGs under distinct but related conditions.
- We propose a Bayesian incorporated linear Non-Gaussian Acyclic Model (BiLiNGAM)
- We apply it to the fMRI images from the Philadelphia Neurodevelopmental Cohort (PNC), which include 855 individuals aged 8–22 years who were divided into five adolescence-related stages.

METHODS

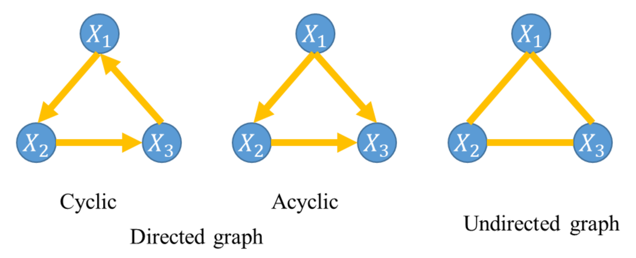
Directed acyclic graph (DAG)

- Notation:

A graph $G = (V, E)$

$V = \{1, 2, \ldots, p\}$, the node set

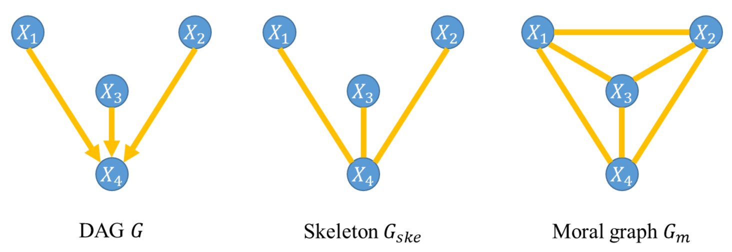
$E \subset V \times V$, the edge set



- Concepts:

Skeleton G_{ske}: a DAG G without the directions

Moral graph G_m: the undirected graph converted from a DAG G



Linear non-Gaussian acyclic model (LiNGAM) [1]

- The observed random vector $X = (X_1, \ldots, X_p) \in R_p$

$$X = B^T X + \epsilon$$

B: weight matrix that can be permuted to be strictly lower triangular

ϵ: continuous r. v., Independent, non-Gaussian, zero means and non-zero variances

BiLiNGAM:

- Joint DAG estimation for multiple groups
  - Consider the related but distinct information across groups
  - Effectively make use of the available data
- Main idea:

Step 1. Undirected graph estimations as prior using the Fast Bayesian integrative analysis (FBIA) [2]

- Principle: $G_{ske} \subset G_{und}$

Step 2. Apply LiNGAM with the prior for DAG estimation

ALGORITHM

Algorithm 1 BiLiNGAM algorithm

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

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

1. Prior estimation: joint Bayesian-incorporating ψ-learning.

Start:

a. For $k = 1, 2, \ldots, K$, use the nonparanormal transformation to render $\mathbf{X}^k$ normal (Gaussian).

b. Apply the ψ-learning method to each group k, $k = 1, 2, \ldots, K$ separately for distinct estimation and acquire the adjacency matrix $\mathbf{E}^{d,k}$.

c. Apply the Bayesian incorporating joint estimation to strengthen the similarities among the groups and acquire the $\mathbf{E}^{c,k}$, $\forall k$.

d. Extract the prior matrix $\mathbf{A}^{prior,k}$ from $\mathbf{E}^{prior,k} = \mathbf{E}^{c,k} \cup \mathbf{E}^{d,k}$, where $a_{ij}^{prior,k} = -1$, if $e_{ij}^{prior,k} = 1$ and otherwise $a_{ij}^{prior,k} = 0$.

End

2. Obtain the estimated weighted DAG adjacency matrices $\hat{\mathbf{B}}^k$: LiNGAM.

Start: For each k

a. Identify the casual order π^k using the direct LiNGAM with the prior matrix $\mathbf{A}^{prior,k}$.

b. Construct a strictly lower triangular matrix $\tilde{\mathbf{B}}^k$ by following the causal order π^k, and the corresponding $\tilde{\mathbf{A}}^{prior,k}$ with the same order.

c. Estimate the connection strengths $(\tilde{\mathbf{B}}_j^k)^T = (\tilde{b}_{1j}^k, \tilde{b}_{2j}^k, \ldots, \tilde{b}_{pj}^k)$ consistent with $\tilde{\mathbf{A}}^{prior,k}$ by solving sparse regressions of the form

$$\hat{\tilde{\mathbf{B}}}_j^k = \arg \min_{\tilde{\mathbf{B}}_j^k \subset supp(\tilde{a}_j^{prior,k})} ||\mathbf{X}_j^k - \mathbf{X}^k \tilde{\mathbf{B}}_j^k||_2^2$$

d. Obtain $\hat{\mathbf{B}}^k$ by converting $\tilde{\mathbf{B}}^k$ to the original order.

End

SIMULATIONS

- Total number of subjects N = 750
- Number of groups K = 3,5
- Number of subjects for each group n = N/K
- Variable size (node): p = 200
- Simulated the random DAG G through the R package pcalg
- Average edge per node d =1,2,5
- Noise distribution: exponential

Compared 4 estimation methods

- PC [3]
- LiNGAM [1]
- ψ-LiNGAM [4]
- BiLiNGAM

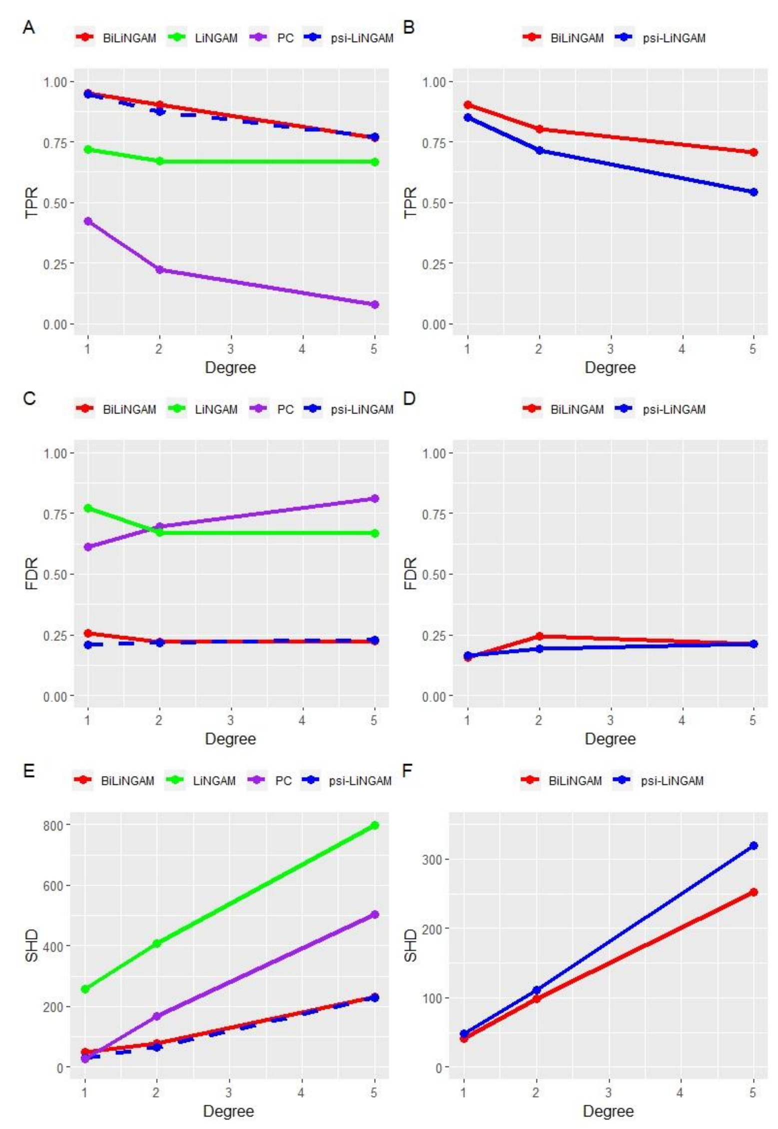


Fig. 1: Mean TPR, FDR and SHD over various graph densities. The results for K = 3 are on the left. Comparisons between BiLiNGAM and ψ-LiNGAM for K = 5 are on the right.

MATERIALS

- Dataset: the Philadelphia Neurodevelopmental Cohort (PNC)
- Subjects

Number of subjects: 855
Age range: 8-22

	Group	Age	# of subjects
1	Pre-adolescence	8-12	194
2	Early adolescence	12-14	150
3	Middle adolescence	14-16	158
4	Late adolescence	16-18	166
5	Post-adolescence	18-22	187

- Brain image: fMRI

Task fMRI

Emotion identification task
- 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 mins.
- Single 3T Siemens TIM Trio whole-body scanner

Standard preprocessing through SPM12

Preprocessing steps:
- Motion correction
- Co-registration
- Normalization
- Slice time correction
- Spatial smoothing
- Bandpass filtering: 0.01-0.1Hz

Beta-maps for 264 ROIs

- General linear regression
- Power atlas with a 5 mm sphere radius parameter
- 264 ROIs
- 12 functional network modules

RESULTS

- Development of emotion-related intra- and inter- module Connectivity

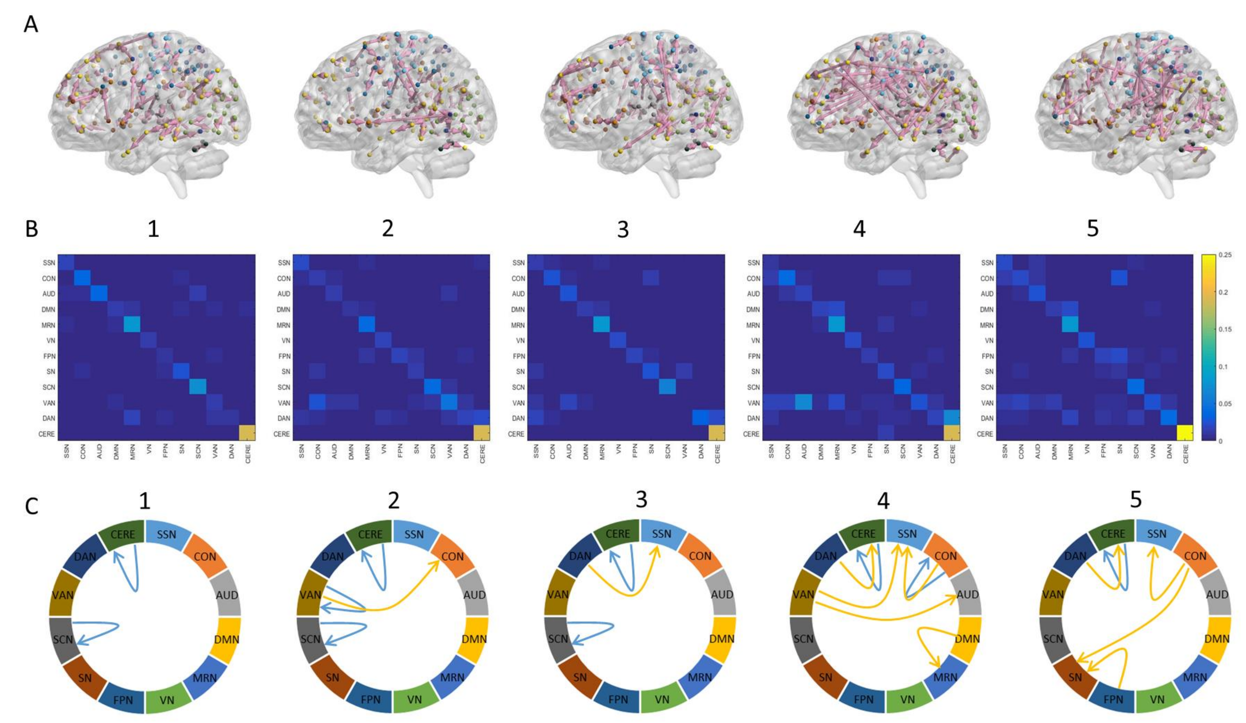


Fig. 2: Causal brain connectivity development from pre-adolescence to post-adolescence. The number index (1 to 5) corresponds to the age category. A, Sagittal views of emotion-related node-level causal networks, where the arrows indicate the causal flow. B, Heatmaps of the mean edge degrees, module-wise. C, Identified intra- (blue arrows) and inter- (yellow arrows) module causal flows.

- Development of emotion-related hubs:
  - Definition: Nodes with degrees at least two standard deviation higher than the mean degrees
  - Two types of hubs:

  8 In-hubs: based on in-degrees, centers to receive information

  25 Out-hubs: based on out-degrees, centers to convey out information

CONCLUSIONS

- We proposed the BiLiNGAM to jointly estimate multiple DAGs in the high dimensional setting for non-Gaussian data.
- The method accomplished the integration of the undirected graph and the directed acyclic graph.
- The analysis of brain's emotion circuit development revealed the trajectory of directed brain circuitry during emotion identification tasks over various adolescent groups.
- Our findings provide a causation template of emotion processing in the developing brain.

ACKNOWLEDGEMENT

The work has been funded by NIH (R01GM109068, R01MH104680, R01MH107354, P20GM103472, 2R01EB005846, 1R01EB006841), and NSF (#1539067).

REFERENCES

1. S. Shimizu, et al., “Direct lingam: A direct method for learning a linear non-gaussian structural equation model,” Journal of Machine Learning Research, vol. 12, pp. 1225–1248, 2011.
2. B. Jia, F. Liang, and the TEDDY study group, “Learning multiple gene regulatory networks in type 1 diabetes through a fast bayesian integrative method,” arXiv preprint, vol. arXiv:1805.02620v3, 2018.
3. P. Spirtes, C. Glymour, R. Scheines, D. Heckerman, C. Meek, G. Cooper, and T. Richardson, Causation, prediction, and search. MIT press, 2000.
4. Zhang, Aiying, et al. "Causal inference of brain connectivity from fMRI with ψ-Learning Incorporated Linear non-Gaussian Acyclic Model ψ-LiNGAM)." arXiv preprint arXiv:2006.09536 (2020).

Contact: Aiying Zhang
Email: azhang4@tulane.edu