Identifying Temporal Pathways Using Biomarkers in the Presence of Latent Non-Gaussian Components [1]

Shanghong Xie

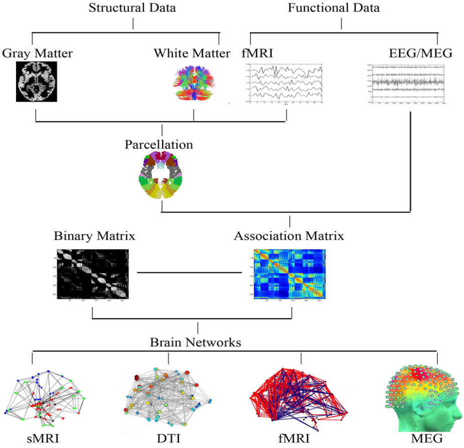
Department of Statistics

UNIVERSITY OF South Carolina

[1] Xie et al. (2024). Identifying Temporal Pathways Using Biomarkers in the Presence of Latent Non-Gaussian Components. *Biometrics* 80 (2), ujae033.

Background: Brain Networks

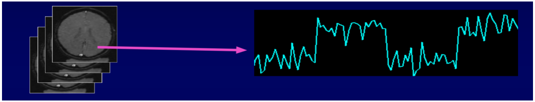
Network analysis: investigate the interrelationships between elements (e.g., brain regions, symptoms, genes) as a system. Nodes: brain regions; Edges: relations between regions



Bassett & Bullmore (2010) Curr Op Neurol

Background: Brain Effective Connectivity

Based on functional data (fMRI, EEG, MEG): time series recorded at various brain regions in a group of individuals



- **Brain functional connectivity**
  - Associations between time series of regions, does not infer directed temporal nature of relations between regions
- **Brain effective connectivity**
  - Directed temporal relations between brain regions based on time-series data (Bullmore and Bassett, 2011)

Existing Network Analysis Approaches

For time series data:

- **Granger causality analysis** $X(t) = \sum_k A_k X(t-k) + \varepsilon(t)$: build on the vector autoregression (VAR) framework with Gaussian noise $\varepsilon(t)$ (Granger, 1969; Bressler et al. 2011)
- **Dynamic causal modeling (DCM):**

$$\dot{X}(t) = AX(t) + Cu(t), \quad Y(t) = f(X(t)) + \varepsilon(t)$$

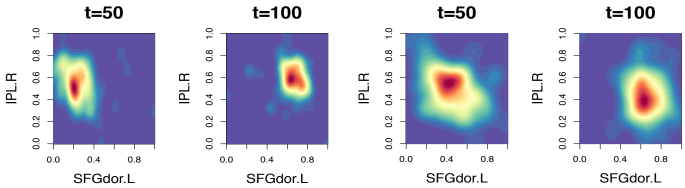
  state-space model, using latent state variables to describe the complex system by first-order ODE (Friston et al., 2003, 2011).

Assume Gaussian distribution for noises, do not accommodate the contemporaneous relations (i.e., associations between elements measured at the same time).

Challenges

In brain functional data

- Neuronal signal contaminated by artifacts and structured noises (Konrad and Eickhoff, 2010)
  - e.g physiological noise, motion-related artifacts, eye movement artifacts, or scanner-induced noise
- Recorded signals may have non-Gaussian properties (Wink and Roerdink, 2006)



(a) Case (b) Control

Figure: Kernel density of two brain regions at different time

Existing Network Analysis Approaches

Methods addressing non-Gaussian components:

- Linear Non-Gaussian Acyclic Model (**LiNGAM**; $X = BX + e$): estimate directed acyclic graph (DAG) B, in the presence of non-Gaussian noise e (Shimizu et al., 2006)
- **VARLiNGAM** ($X(t) = \sum_k B_k X(t-k) + e(t)$): extension to LiNGAM (Hyvärinen et al., 2010)
- Structural Independent Component Analysis (ICA) removal: requires expert knowledge and judgments (Griffanti et al., 2014)

Challenges:

- Directly model the temporal relations at the level of observed measurements, but neuronal signals are often not directly observed by non-invasive imaging techniques.
- VARLiNGAM designed for a single subject's data.

Our Contributions

Using data collected from a group of subjects to identify the temporal relationships between Gaussian components

- Decompose **observed measurements**
  - **Latent Gaussian process**: temporal relations between elements of interest
  - **Non-Gaussian components**: e.g., artifacts, structured noise, other unobserved non-intervenable factors
- ICA to address structured noise
- Moment estimations to obtain the temporal and contemporaneous networks
  - No distributional assumptions on non-Gaussian components

Methods

Model

- $\boldsymbol{Y}_i(t) = (Y_{i1}(t), ..., Y_{iK}(t))'$: observed biomarkers (e.g., BOLD signals) measured at time t:

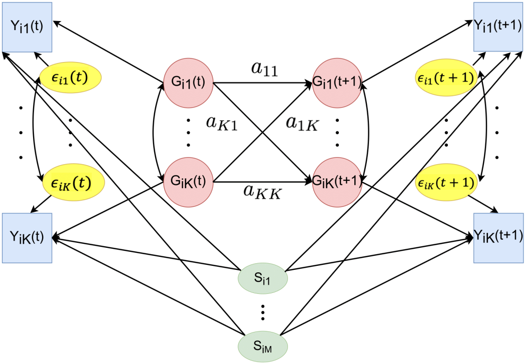
$$\boldsymbol{Y}_i(t) = \boldsymbol{U}_i(t) + \boldsymbol{G}_i(t) + \boldsymbol{\epsilon}_i(t)$$

- $\boldsymbol{U}_i(t) = \boldsymbol{w}(t)\boldsymbol{S}_i$: latent non-Gaussian processes
  - $\boldsymbol{S}_i = (S_{i1}, \ldots, S_{iM})'$: latent non-Gaussian sources, mutually independent with $E(\boldsymbol{S}_i) = 0$, $E(\boldsymbol{S}_i\boldsymbol{S}_i') = \boldsymbol{I}_{M\times M}$
  - Static brain activity, artifacts, and structured noise
- $\boldsymbol{G}_i(t) \sim N(0, \boldsymbol{\Sigma}_t)$: independent latent Gaussian processes that represent the signals of interest
- $\boldsymbol{\epsilon}_i(t) \sim N(0, \frac{1}{T}\boldsymbol{\Omega})$: residual errors that represent contemporaneous information

Model

For the Gaussian processes of interest: $\boldsymbol{G}_i(t+1) = \boldsymbol{A}\boldsymbol{G}_i(t)$

- $\boldsymbol{A} = (a_{kj})$: temporal network. a_{kj}: how jth component of $\boldsymbol{G}_i(t)$ at time t influences kth component at time $t+1$.
- $\boldsymbol{\Gamma} = \boldsymbol{\Omega}^{-1} = (\gamma_{kj})$: contemporaneous network. γ_{kj}: association between $\epsilon_{ik}(t)$ and $\epsilon_{ij}(t)$ conditioning on other $\epsilon_{il}(t)$



Model

Our goal is to infer two networks

- Temporal network $\boldsymbol{A}$
  - Temporal pathways among Gaussian components of interest (i.e., $\boldsymbol{G}_i(t)$)
  - In dynamic causal model (DCM), $\boldsymbol{A}$ is often used to infer temporal effects and effective connectivities
- Cross-sectional contemporaneous network $\boldsymbol{\Gamma}$
  - Undirected network obtained after accounting for the temporal effects.

Estimation

- Model

$$\boldsymbol{Y}_i(t) = \boldsymbol{U}_i(t) + \boldsymbol{G}_i(t) + \boldsymbol{\epsilon}_i(t), \quad \boldsymbol{G}_i(t+1) = \boldsymbol{A}\boldsymbol{G}_i(t)$$

- $\boldsymbol{G}_i(t+1) = \boldsymbol{A}\boldsymbol{G}_i(t)$ implies $\boldsymbol{G}_i(t) = \boldsymbol{A}^t\boldsymbol{G}_i(0)$, require $|\lambda(\boldsymbol{A})| < 1$, where $\lambda(\boldsymbol{A})$ denotes the eigenvalues of $\boldsymbol{A}$
  - When t is small, $\boldsymbol{A}^t$ is large and contains significant variability.
  - When t is large, $\boldsymbol{A}^t$ is small, and the variability is primarily in the residuals $\boldsymbol{\epsilon}_i(t)$.

  Thus, accurate estimation of $\boldsymbol{A}$ requires the first few time points, while accurate estimation of $\boldsymbol{\Gamma}$ requires later time points.

Estimation

The parameters of components can be estimated in two steps

- Estimate the non-Gaussian process $\boldsymbol{U}_i(t) = \boldsymbol{w}(t)\boldsymbol{S}_i$ which involves the independent sources $\boldsymbol{S}_i$ and weight matrix $\boldsymbol{w}(t)$
- After removing $\boldsymbol{U}_i(t)$ from $\boldsymbol{Y}_i(t)$, estimate the temporal network $\boldsymbol{A}$ and the contemporaneous network $\boldsymbol{\Gamma}$ through moment estimations

Estimation: non-Gaussian Components

Latent non-Gaussian process

$$\boldsymbol{U}_i(t) = \boldsymbol{w}(t)\boldsymbol{S}_i.$$

- Contributions of Gaussian processes $\boldsymbol{G}_i(t)$ and $\boldsymbol{\epsilon}_i(t)$ become negligible with large T; only non-Gaussian components remain.
- Perform ICA on $\overline{\boldsymbol{Y}}_{iT} = T^{-1}\sum_{t=0}^{T-1}\left(\widetilde{Y}_{ij}(t)\right)_{j=1}^{K}$

$$\widehat{\boldsymbol{S}}_i = \widehat{\boldsymbol{C}}^{-1}\overline{\boldsymbol{Y}}_{iT}, \quad \widehat{\boldsymbol{C}}^{-1}\widehat{\boldsymbol{C}} = \boldsymbol{I}_{M\times M}.$$

  - FastICA (Hyvarinen, 1999)
  - Number of ICs determined by minimum description length (MDL) criteria
- Given $\widehat{\boldsymbol{S}}_i$, least squares to obtain $\boldsymbol{w}_j(t)$: $\sum_{i=1}^{n}(Y_{ij}(t) - \boldsymbol{S}_i^T\boldsymbol{w}_j(t))^2$

Estimation: Temporal Network $\boldsymbol{A}$

Let $\boldsymbol{R}_i(t) = \boldsymbol{Y}_i(t) - \boldsymbol{U}_i(t)$. Note:

$$\boldsymbol{R}_i(t) = \boldsymbol{A}^t \boldsymbol{G}_i(0) + \boldsymbol{\epsilon}_i(t)$$

$$\boldsymbol{\Theta} = \mathrm{Cov}(\boldsymbol{G}_i(0)), \Psi_{t,s} = \mathrm{Cov}(\boldsymbol{R}_i(t), \boldsymbol{R}_i(s)).$$

- For any pair of time (t, s), $s \neq t$, $\Psi_{t,s} = \boldsymbol{A}^t \boldsymbol{\Theta} (\boldsymbol{A}^s)'$
- For any three time points (t, s, l),

$$\Psi_{t,s}(\Psi_{l,s})^{-1} = \boldsymbol{A}^t \boldsymbol{\Theta} (\boldsymbol{A}^s)' (\boldsymbol{A}^s)'^{-1} \boldsymbol{\Theta}^{-1} \boldsymbol{A}^{-l} = \boldsymbol{A}^{t-l}.$$

  Thus, $\boldsymbol{A} = \Psi_{t+2,t}(\Psi_{t+1,t})^{-1}$ for any $t = 0, \ldots, T-3$.

- To stabilize estimation, use a fixed number of time points T_a,

$$\widehat{\boldsymbol{A}} = \Big(\frac{1}{T_a} \sum_{t=0}^{T_a - 1} \widehat{\Psi}_{t+2,t} \Big) \Big(\frac{1}{T_a} \sum_{t=0}^{T_a - 1} \widehat{\Psi}_{t+1,t} \Big)^{-1}$$

Estimation: Contemporaneous Network Γ

For $\boldsymbol{\Omega}$, consider covariance at the same time point:

$$\Psi_{t,t} = \mathrm{Cov}(\boldsymbol{R}_i(t), \boldsymbol{R}_i(t)) = \boldsymbol{\Sigma}_t + \frac{1}{T}\boldsymbol{\Omega},$$
$$\Psi_{t+1,t+1} = \mathrm{Cov}(\boldsymbol{R}_i(t+1), \boldsymbol{R}_i(t+1)) = \boldsymbol{A}\boldsymbol{\Sigma}_t\boldsymbol{A}' + \frac{1}{T}\boldsymbol{\Omega}.$$

Using the vectorization operator,

$$\mathrm{vec}(\boldsymbol{\Omega}) = (\boldsymbol{A} \otimes \boldsymbol{A} - \boldsymbol{I})^{-1}\,\mathrm{vec}(T\boldsymbol{Q}_t),$$

where $\boldsymbol{Q}_t = \boldsymbol{A}\Psi_{t,t}\boldsymbol{A}' - \Psi_{t+1,t+1}$.

- $\boldsymbol{\Sigma}_t = \boldsymbol{A}^t\boldsymbol{\Theta}(\boldsymbol{A}')^t$ becomes small when t is large enough (i.e., on the scale of $T^{1/2}$), use the time points $t \geq T_c$ to estimate $\boldsymbol{\Omega}$.
- Contemporaneous network $\boldsymbol{\Gamma}$ is estimated as $\widehat{\boldsymbol{\Gamma}} = \widehat{\boldsymbol{\Omega}}^{-1}$.

Identifiability

Lemma 1

Suppose that model (1) holds for another set of latent variables $\widetilde{\mathbf{S}}_i, \widetilde{\boldsymbol{G}}_i(t), \widetilde{\boldsymbol{\epsilon}}_i(t)$ *but with different parameters* $\boldsymbol{w}(t), \boldsymbol{A}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}$*, and the distribution of* $\widetilde{\mathbf{S}}_i$*,* f_s*. Under technical conditions,* $\boldsymbol{w}(t) = \boldsymbol{w}_0(t)$*,* $\boldsymbol{A} = \boldsymbol{A}_0$*,* $\boldsymbol{\Omega} = \boldsymbol{\Omega}_0$*,* $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_0$*, and* $f_s = f_0$*.*

Asymptotic Properties

Theorem 1

Under technical conditions and T_a is a fixed number of time points, $\sqrt{n}\{\mathrm{vec}(\widehat{\boldsymbol{A}}) - \mathrm{vec}(\boldsymbol{A}_0)\}$ *converges in distribution to a mean-zero normal distribution.*

Theorem 2

Under technical conditions and $T_c = O(\sqrt{T})$, $\sqrt{n}\{\mathrm{vec}(\widehat{\boldsymbol{\Gamma}}) - \mathrm{vec}(\boldsymbol{\Gamma}_0)\}$ *converges in distribution to a mean-zero normal distribution.*

Require earlier time points in a time-series (i.e., $t < T_a$) to estimate $\boldsymbol{A}$ and later time points (i.e., $t \geq T_c$) to estimate $\boldsymbol{\Gamma}$.

Asymptotic covariances of $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{\Gamma}}$ are of complex form, use bootstraps to estimate the asymptotic covariance matrix in the simulation studies.

Simulation Studies

Simulation Settings

- Number of nodes $K = 5, 10, 20$, $n = 100$, $T = 200, 400, 1000, 2000$
- Scenario 1: generate data from our model (1)
  - Temporal network $\boldsymbol{A}$: $a_{jj} = 0.8$, non-null $a_{jk} = 0.2$
- Scenario 2: generate data from a dynamic system based on a stochastic differential equation

$$\dot{\boldsymbol{G}}_i = \boldsymbol{B}\boldsymbol{G}_i + diag(0.1, \ldots, 0.1)\dot{\boldsymbol{\epsilon}}_i, \quad \boldsymbol{Y}_i(t) = \boldsymbol{U}_i(t) + \boldsymbol{G}_i(t),$$

- $\boldsymbol{S}_i$: three independent $Unif(1, 3)$
- $w_{11}(t) = w_{11}(0) + 5t/T$, $w_{22}(t) = w_{22}(0) + 5(t/T)^2$, $w_{33}(t) = w_{33}(0) + 5\sin(2t/T)$, $w_{43}(t) = w_{43}(0) + 5\cos(3t/T)$, and $w_{jm}(t) = w_{jm}(0)$ for all the remaining elements, where $w_{jm}(0) \sim N(5, 1)$

Compared to: no IC approach, LiNGAM, VARLiNGAM

Simulation Results

Table: Simulation performance of estimated $\boldsymbol{A}$ and $\boldsymbol{\Gamma}$ in Scenario 1.

Number of Time Points	Number of Nodes	Network	Method	MSE	AUC	95% Coverage probability	95% Coverage length
$T = 200$	$K = 5$	$\boldsymbol{A}$	Our method	0.01	0.989	0.92	0.071
			No IC	0.148	0.523	–	–
		$\boldsymbol{\Gamma}$	Our method	< 0.001	1	0.92	0.003
			No IC	0.168	0.499	–	–
	$K = 10$	$\boldsymbol{A}$	Our method	0.019	0.994	0.95	0.05
			No IC	0.134	0.894	–	–
		$\boldsymbol{\Gamma}$	Our method	0.002	1	0.94	0.008
			No IC	1.601	0.326	–	–
	$K = 20$	$\boldsymbol{A}$	Our method	0.071	0.999	0.96	0.054
			No IC	0.318	0.827	–	–
		$\boldsymbol{\Gamma}$	Our method	0.081	0.956	0.98	0.164
			No IC	9.361	0.567	–	–

MSE: mean squared error; – indicates the average 95% coverage probability and coverage length over 100 simulations for no IC approach were not applicable.

Simulation Results

Table: Simulation performance of estimated $\boldsymbol{A}$ by LiNGAM and VARLiNGAM in Scenario 1.

Number of Time Points	Number of Nodes	Method	MSE	AUC
$T = 200$	$K = 5$	LiNGAM	22.998	0.657
		VARLiNGAM	1.728	0.509
	$K = 10$	LiNGAM	63.955	0.576
		VARLiNGAM	5.676	0.661
	$K = 20$	LiNGAM	37.483	0.444
		VARLiNGAM	16.127	0.512

MSE: mean squared error.

Simulation Results

Table: Simulation performance of estimated $\boldsymbol{A}$ and $\boldsymbol{B}$ in Scenario 2.

Number of Time Points	Number of Nodes	Method	MSE of $\boldsymbol{A}$	MSE of $\boldsymbol{B}$	AUC of $\boldsymbol{A}$	AUC of $\boldsymbol{B}$
$T = 200$	$K = 5$	Our method	0.069	0.113	0.952	0.964
		No IC	0.307	0.435	0.574	0.494
		LiNGAM	24.152	–	0.682	–
		VARLiNGAM	0.338	0.469	0.791	0.633
	$K = 10$	Our method	0.308	0.526	0.804	0.943
		No IC	0.535	0.797	0.686	0.779
		LiNGAM	61.707	–	0.613	–
		VARLiNGAM	0.400	0.546	0.811	0.957
	$K = 20$	Our method	1.154	2.151	0.813	0.975
		No IC	1.300	2.190	0.815	0.920
		LiNGAM	49.849	–	0.609	–
		VARLiNGAM	0.717	1.027	0.784	0.963

MSE: mean squared error; – indicates that LiNGAM was not able to estimate $\boldsymbol{B}$.

Real Data Application

Real Data Application: ADHD-200 Consortium Data

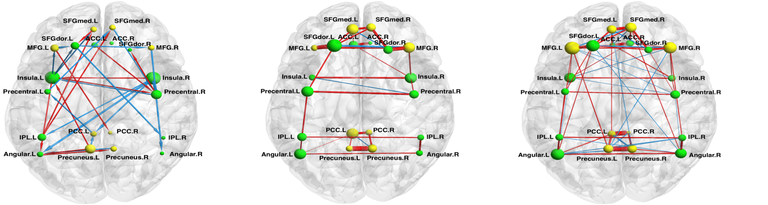
ADHD-200 consortium at NYU Child Study Center: 88 healthy controls, 117 ADHD individuals.

Resting-state fMRI data: time courses of regions of interest (ROIs), 172 time points. Extract 20 commonly studied ROIs:

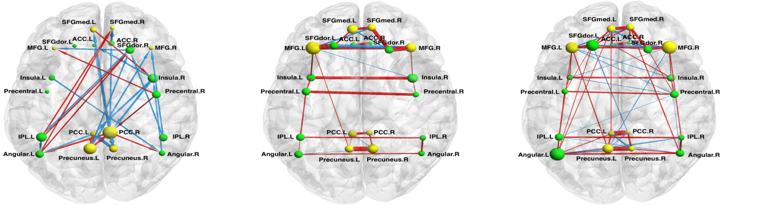
- default mode network: bilateral middle frontal gyrus (MFG), posterior cingulate cortex (PCC), medial superior frontal gyrus (SFGmed), and precuneus regions.
- cognitive control network: bilateral angular, insula, dorsolateral superior frontal gyrus (SFGdor), anterior cingulate cortex (ACC), precentral, and inferior parietal (IPL) regions

Our goal: use the temporal dynamics in fMRI signals to analyze temporal relations (i.e., effective connectivity)

Temporal network | Contemporaneous network | Functional connectivity



(a) Cases



(b) Controls

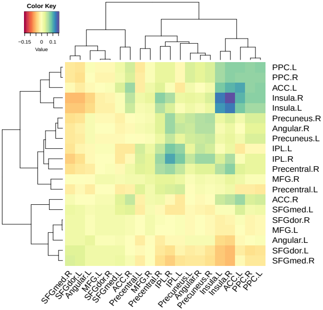
Yellow: default mode network; Green: cognitive control network. Blue: positive edge; Red edge: negative edge. Edge width is proportional to the edge strength. Top 30 (or 60) edges based on FDR adjusted p-values.

Real Data Results

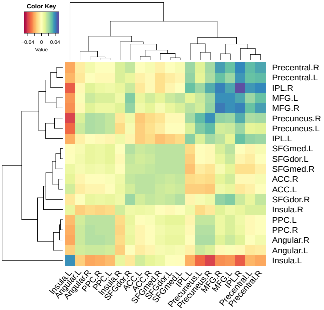
- Edges identified by the functional connectivity study were mostly contemporaneous edges instead of temporal edges
- Temporal networks:
  - Reduced effective connectivity within default mode network (DMN), especially between precuneus and other DMN regions, including MFG and PCC.
  - Increased connectivity within cognitive control (CC) network
  - Decreased connectivity between DMN and CC network
- Consistent with a meta-analysis of 20 studies (Sutcubasi et al. 2020)
- Support the hypothesis that one potential mechanism of ADHD is disconnection between regions within the default mode network (Konrad and Eickhoff, 2010; Castellanos et al., 2008; Uddin et al., 2008)

Figure: Heatmaps of spatial correlations from $\boldsymbol{U}(t)$ of the case versus the control group.

Case



Control

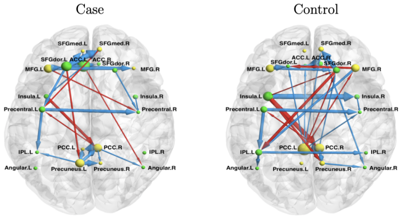


Non-Gaussian effects were spatially clustered. Same regions in the left and right hemispheres tend to form a cluster.

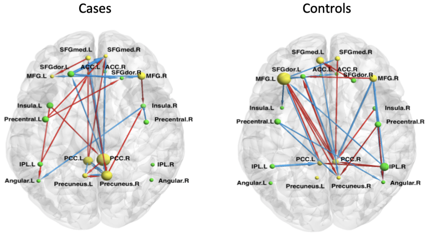
Case group: insula, ACC (CC network); Control: MFG, precentral, insula (CC and DMN)

Comparison with Alternative Methods

- LiNGAM: does not differentiate temporal/contemporaneous network, less consistent with functional connectivity network.



- Structural ICA denoising: failed to identify insula as the key region for the case group



Discussion

- Discover temporal network from time-series biomarkers
- Decompose observed biomarker measurements (contain multiple sources) into Gaussian+non-Gaussian components
- Separate temporal network from contemporaneous network
- Not accounting for non-Gaussian components may bias the temporal network between Gaussian signals
- Designed for a group of subjects to characterize the group-level networks

Discussion

Extensions

- AR(1) can be extended to higher orders
- A is time-invariant, but may depend on time
- Assume A is homogeneous across a group of similar subjects. In a heterogeneous population, model A as subject-specific
- Extension to other data modalities, high-dimensional applications

Reference: Xie et al. (2024). Identifying Temporal Pathways Using Biomarkers in the Presence of Latent Non-Gaussian Components. *Biometrics* 80 (2), ujae033.

R package: `https://github.com/shanghongxie/ICATemporalNetwork`

Collaborators

- Donglin Zeng, Department of Biostatistics, University of Michigan
- Yuanjia Wang, Department of Biostatistics, Columbia University

THANK YOU !