Selected Papers on Multilevel Graphical Models

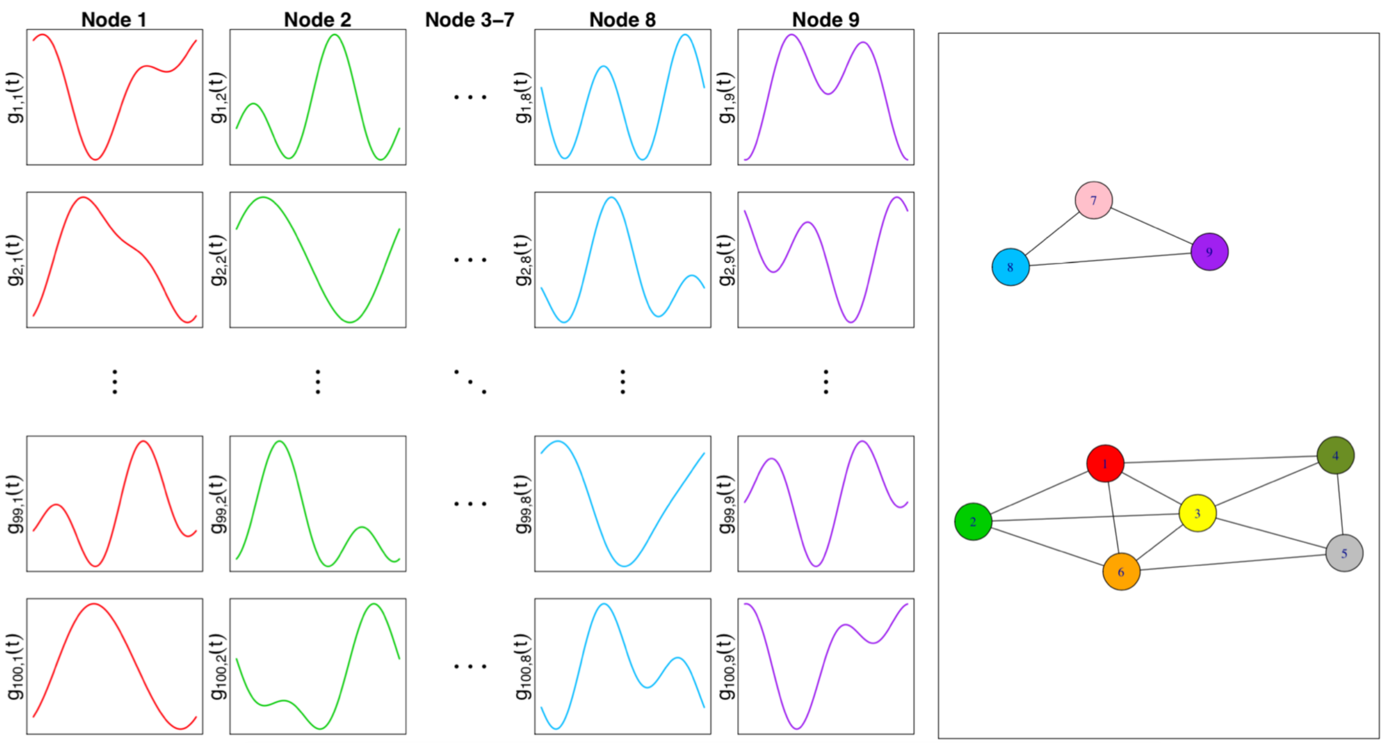
Javier Zapata

Today’s Plan

- Fast review on my research (< 7 mins)
- Discuss two different multilevel graphical models:
    - Multiple individual-level graphical models generated from a common group-level graphical model.
    - Multiple graphical models with different levels of resolution.
- The papers show application for neuroscience. But they can be applied in finance as well.

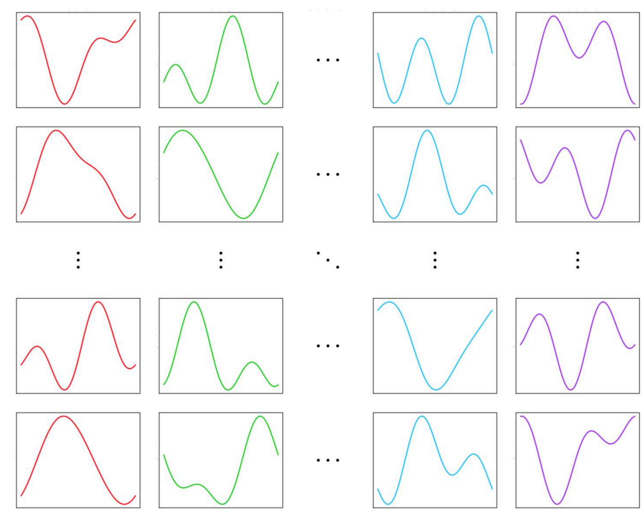
Functional Graphical Models

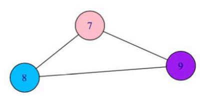
Functional al Brain Connectivity from fMRI Data



Step 1: Multivariate Karhunen-Loeve Decomposition (using partial separability)

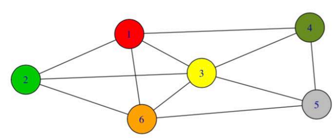
$$\underbrace{\begin{pmatrix} X_1(t) \\ \vdots \\ X_p(t) \end{pmatrix}}_{\text{Multivariate stochastic process}} = \sum_{l=1}^{\infty} \underbrace{\begin{pmatrix} \theta_{l1} \\ \vdots \\ \theta_{lp} \end{pmatrix}}_{\substack{\text{Multivariate random variable} \\ \text{(functional principal scores)}}} \overbrace{\phi_l(t)}^{\text{eigenfunctions}}$$





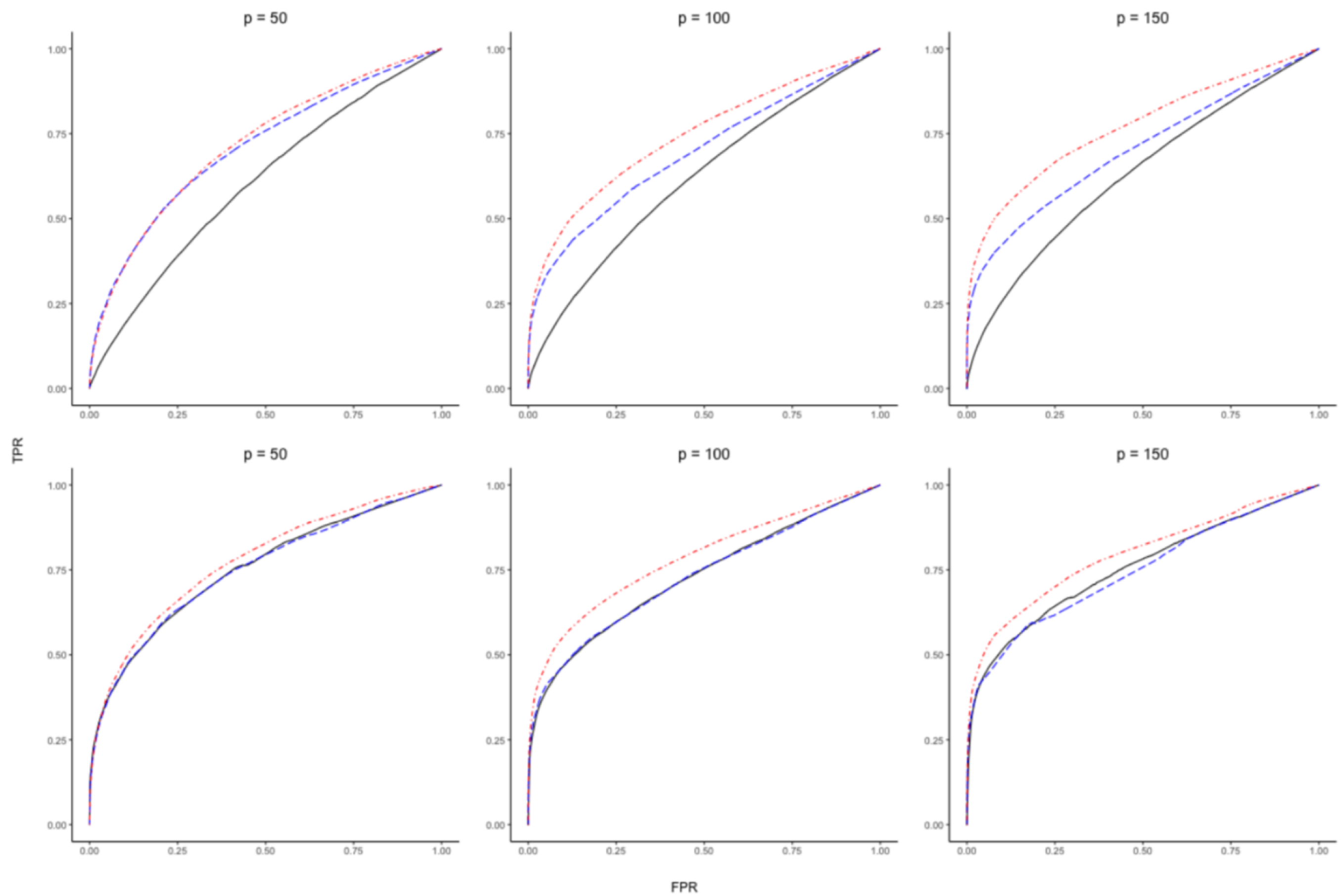
Step 2: Graphical Models

- High Dimensional Data ($p \gg n$)
- Regularization (Joint Graphical Lasso)



$$\max_{\Omega_1,\ldots,\Omega_L} \quad \sum_{l=1}^{L} \left[\log\left(\det \Omega_l\right) - \operatorname{trace}\left(S_l \Omega_l\right)\right] - P(\Omega_1, \ldots, \Omega_L) \qquad \hat{E} = \bigcup_{l=1}^{L} \hat{E}_l.$$

Results from Simulation

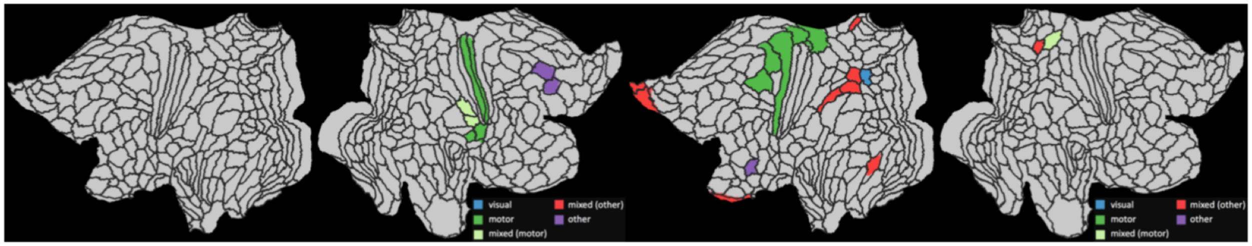


$FGGM$ (——)

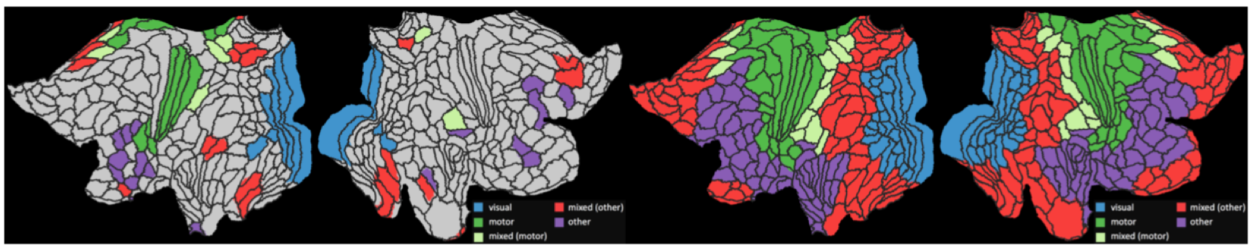
$psFGGM$ (- - - -)

$psFGGM$ (—·—) *at* 95% *of variance explained.*

Results from fMRI Data



(a) Activated ROIs unique to left-hand task (b) Activated ROIs unique to right-hand task



(c) Activated ROIs common to both tasks (d) ROI task activation map [11]

Figure 3: *psFGGM estimated functionally connected cortical ROIs for the left- and right-hand motor tasks. Each sub-figure shows a flat brain map of the left and right hemispheres (in that order). ROIs having a positive degree of connectivity in each estimated graph are colored based on their functionality [11]: visual (**blue**), motor (**green**), mixed motor (**light green**), mixed other (**red**) and other (**purple**).*

For more details

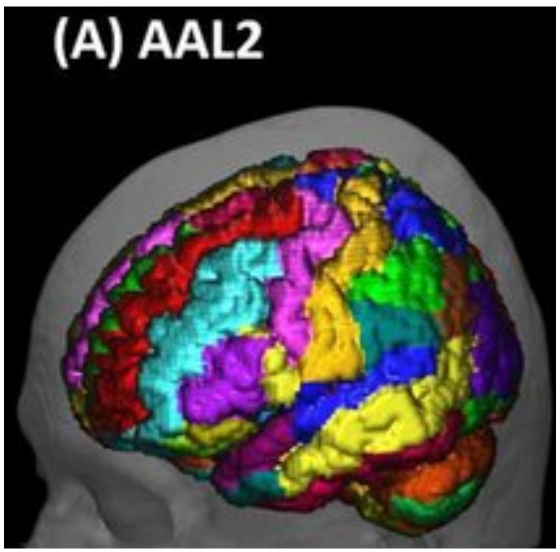
- Zapata J., Oh S. and Petersen A. - Partial Separability and Functional Gaussian Graphical Models
  - Currently under review (…)
  - url: https://arxiv.org/abs/1910.03134

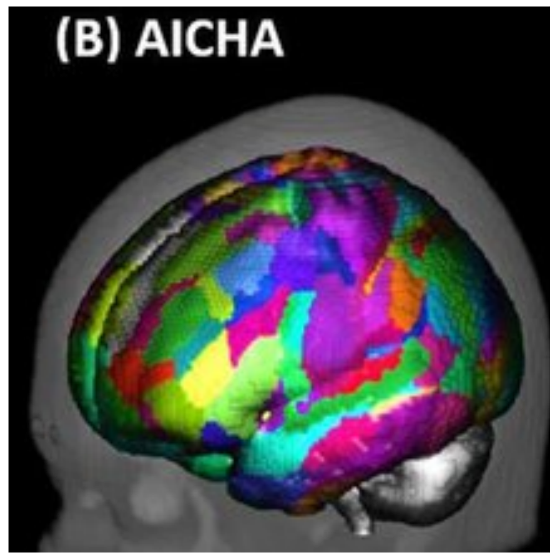
- R-package: fgm (available in the CRAN repository … soon …)

Multilevel Graphical Models

Motivation

- **Raw data is aggregated to obtain meaningful features**. For example:
  - fMRI data has 90k voxels (vertices representing local measurement)
  - Neuroscientists aggregate voxels into regions of interest (ROIs) using <1k parcellations. And then they proceed to analyze the data.





- **Zhang et al (2019)**

A random covariance model for bi-level graphical modeling with application to resting-state fMRI data

Goal:

Multiple individual graphical models Ω_l can be considered as variants of a common group-level graphical model Ω_0.

Estimate the group- and individual-level graphical models simultaneously.

- There are K patients and p brain regions of interest.
  - K datasets $Y^{(1)}, \dots, Y^{(K)}$
  - Each $Y^{(k)}$ has n_k observations of a p-dimensional vector $\boldsymbol{y}_i^{(k)} = (\boldsymbol{y}_{i1}^{(k)}, .., \boldsymbol{y}_{ip}^{(k)})$ for $k = 1, \dots, K$ and $i = 1, \dots, n_k$
  - $\boldsymbol{y}_i^{(k)} \sim N_p(\mu_k, \Sigma_k)$ $\quad\quad \Omega_k = \Sigma_k^{-1}$ $\quad\quad G_k \leftrightarrow \Omega_k$
  - $G_1, \dots, G_K$ are individual-level graphical models

- These K graphical models are variants of a common graphical model
  - G_0 is a group-level graphical model
  - $G_0 \leftrightarrow \Omega_0$

- Minimize a penalized negative log-likelihood

$$\sum_{k=1}^{K}\{-\mathrm{logdet}(\Omega_k) + tr(S_k\Omega_k)\} + P(\Omega_0, \{\Omega_K\})$$

$$P(\{\Omega_k\}, \Omega_0) = \underbrace{\lambda_1 \sum_{k=1}^{K} |\Omega_k|_1}_{P1} + \underbrace{\lambda_2 \sum_{k=1}^{K} \{-\log\det(\Omega_k\Omega_0^{-1}) + \mathtt{tr}(\Omega_k\Omega_0^{-1}) - p)\}}_{P2} + \underbrace{\lambda_3|\Omega_0|_1}_{P3}$$

  - P1 & P3 : L_1 penalties on $\{\Omega_k\}, \Omega_0$ to induce sparsity in each individual graphs $\{G_k\}$ and the group-level graph G_0
  - P2: penalty on the KL-divergence between each individual Ω_k and Ω_0 associated with the underlying group-level graph G_0

P2: penalty on the KL-divergence between each individual Ω_k and Ω_0 associated with the underlying group-level graph G_0

- Consider $\boldsymbol{y}_i^{(k)} | \Omega_k \sim N_p\left(\mu_k, \Omega_k^{-1}\right)$
- Wishart distribution is the conjugate prior for the precision matrix Ω_k

$$\Omega_k \sim W_p(\Omega_0, df = \lambda_2)$$

- The model is a random covariance model which each subject-level covariance is a random level from an inverse Wishart distribution centered at the group mean.

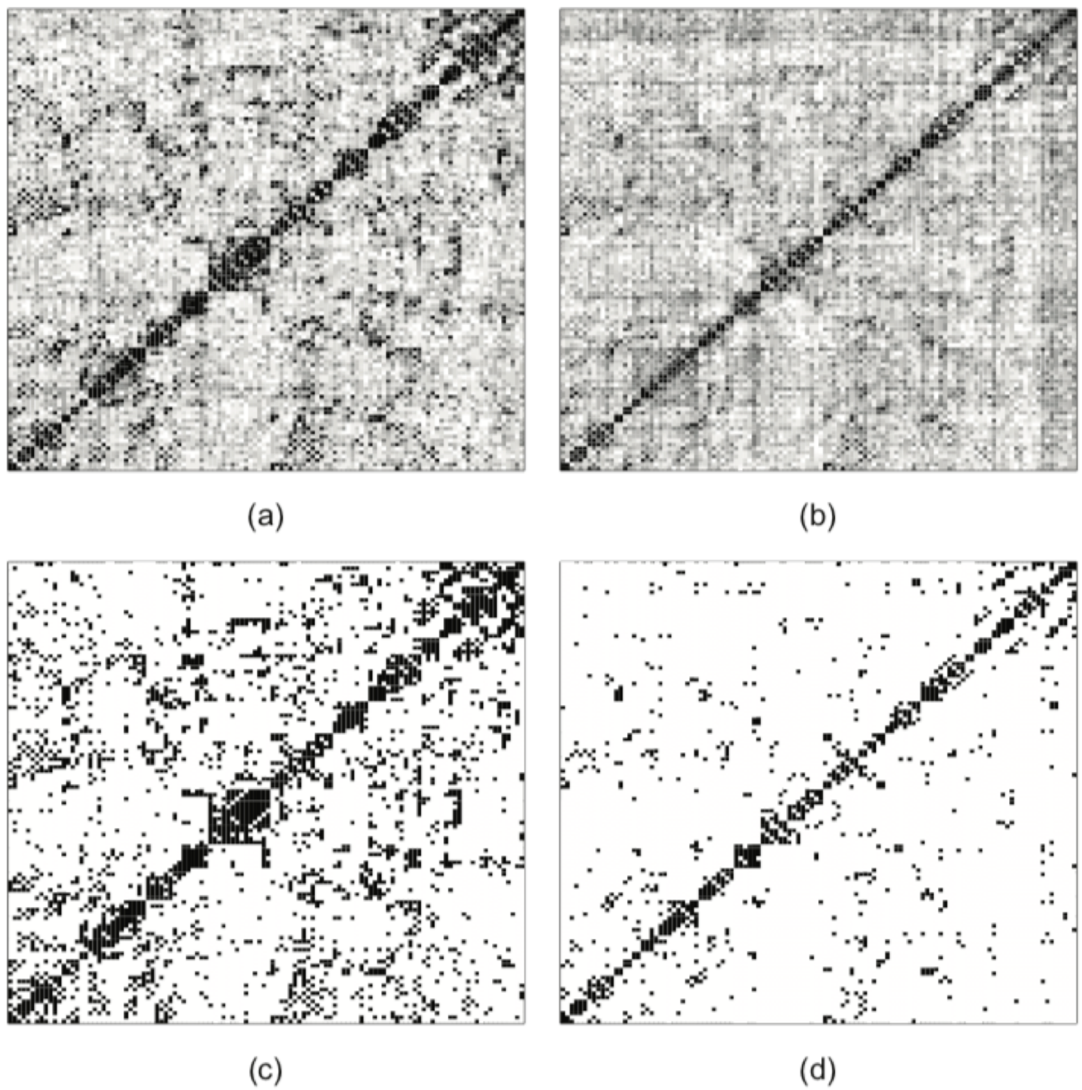
Computational Algorithm

1. Initialize $\widehat{\Omega}_k = (1-\rho)S_k + \rho I_k$ for $k = 1, \ldots, K$, and $\widehat{\Omega}_0 = \sum_k \Omega_k / K$, where ρ is a small value.

2. For $k = 1, \ldots, K$, update $\widehat{\Omega}_k$ by solving
$$\arg\min_{\Omega_k} \left\{ -\log\det\Omega_k + \mathtt{tr}\left(\frac{S_k + \lambda_2 \widehat{\Omega}_0^{-1}}{1+\lambda_2} \Omega_k \right) + \frac{\lambda_1}{1+\lambda_2} |\Omega_k|_1 \right\} \tag{3}$$

3. Update $\widehat{\Omega}_0$ by solving
$$\arg\min_{\Omega_0} \left\{ \log\det\Omega_0 + \mathtt{tr}\left(\frac{\sum_{k=1}^{K} \widehat{\Omega}_k}{K} \Omega_0^{-1} \right) + \frac{\lambda_3}{K\lambda_2} |\Omega_0|_1 \right\} \tag{4}$$

4. Repeat Steps 2 and 3 until convergence is achieved.



Resting-state fMRI

116 ROIs

16 schizophrenia patients

(a) The mean adjacency matrix inferred by the random covariance model

(b) The mean adjacency matrix inferred by the independent Graphical Lasso

(c) The group-level network inferred by the random covariance model.

(d) The common network by the independent graphical lasso (edges shared by a majority of individual estimates)

- **Pircalabelu et al (2020)**

Zoom-in–out joint graphical lasso for different coarseness scales

Goal: Estimate graphs a different scales having a hierarchical relationship:
finer graphs originate from partitioning coarser graphs

- There are (1,…., K) partition scales which are known.
  - If $k > k'$ then scale k is a **finer** scale than k' (i.e. k' is a **coarser** scale)
  - Each scale k consists of p^k **regions** (i.e. features)
  - Every scale has its own graph: $G^k(E_u^k, V^k)$ with $\left|V^k\right| = p^k$ (i.e. nodes=regions)
  - $E_u^k = \{(a,b) \in V^k \times V^k \middle| a \neq b \text{ and } \Theta_{a,b}^k \neq 0\}$

- Nodes at a coarser scale are partitioned into children nodes in finer graphs.
  - $V_s^{1,k}(a) \coloneqq \{v \in V^k: a \in V^1, v \in children(a)\} \cup \{a\}$ for $k > 1$
  - k is a **coarser** than k' if: $p^k < p^{k\prime}$ and $\left|V_s^{1,k}(a)\right| \leq \left|V_s^{1,k'}(a)\right|$ for all nodes a

- **Zoom-in**: If $a \sim b \in E_u^k$ then exists $(a', b') \in V_s^{k,k'}(a) \times V_s^{k,k'}(b)$

  with $a' \sim b' \in E_u^{k'}$ for some $k' > k$
- **Zoom-out**: Reverse idea: an edge at a finer graph induces an edge at some coarser graph

Multilevel Graph: An Illustration of ‘Zoom-in’ and ‘Zoom-out’

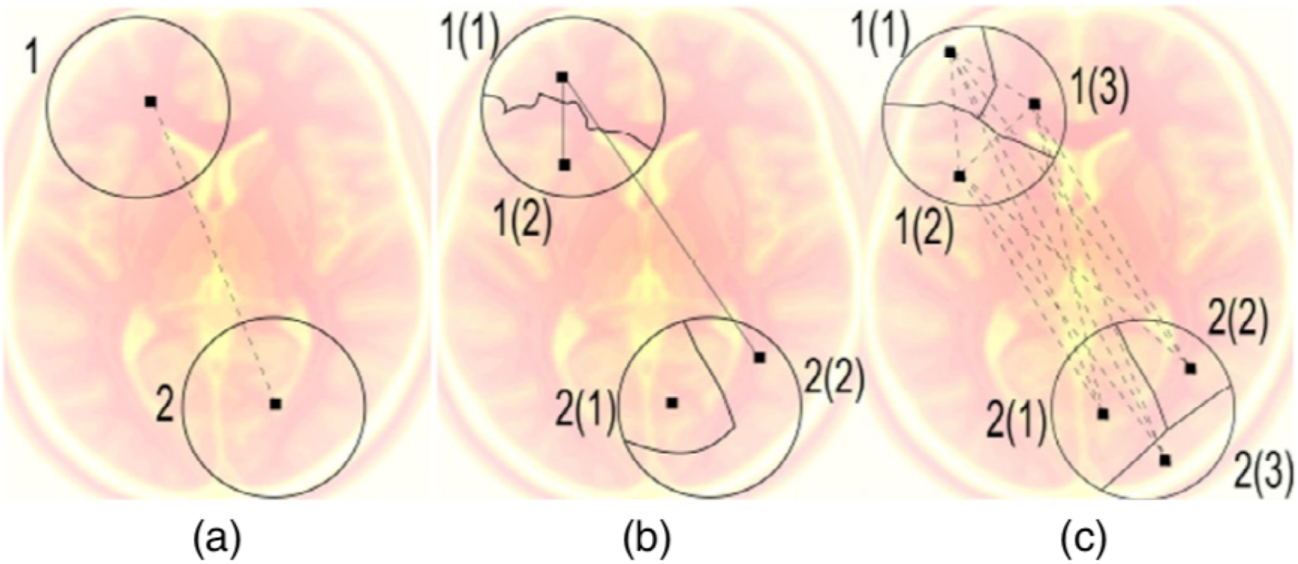


Fig. 2. Hypothetical example (panels (a)–(c) present regions 1 and 2 at $K = 3$ scales): (a) scale $k = 1$ corresponds to the unsplit regions and at scales (b) $k = 2$ and (c) $k = 3$ each region contains two (scale 2) and three (scale 3) partitions (———, edges at scale $k^* = 2$; - - - - -, set of allowed edges at the finer scale $k'' = 3$ (panel (c)) and at the coarser scale $k' = 1$ (panel (a))); moving from (b) to (c) illustrates the zoom-in property (the graph becomes larger), whereas moving from (b) to (a) illustrates the zoom-out property (the graph becomes smaller)

- For every subject we observe signals at every scale and every region

  $X_{j,t}^{k}$:= signal at scale k for region j at time t. Hence, $X_{\cdot,t}^{k} \in \mathbb{R}^{p^k}$

  $X_{\cdot,t}^{k} | X_{\cdot,t-1}^{k} \sim \mathcal{N}(\Gamma^k X_{\cdot,t-1}^{k}, \Sigma^k)$ (1st order Markov model)

- Log-likelihood $\mathcal{L}(\boldsymbol{\Gamma}^1, \ldots, \boldsymbol{\Theta}^K) = \Sigma_{k=1}^{K} [\log\{\det(\boldsymbol{\Theta}^k)\} - \mathrm{tr}(\mathbf{S}_{\boldsymbol{\Gamma}^k} \boldsymbol{\Theta}^k)]$,

where: $\mathbf{S}_{\boldsymbol{\Gamma}^k} = (nT)^{-1} \sum_{i=1}^{n} \sum_{t=2}^{T} (\mathbf{X}_{i,\cdot,t}^{k} - \boldsymbol{\Gamma}^k \mathbf{X}_{i,\cdot,t-1}^{k})(\mathbf{X}_{i,\cdot,t}^{k} - \boldsymbol{\Gamma}^k \mathbf{X}_{i,\cdot,t-1}^{k})^{\mathrm{T}}$

- They solve a penalized negative log-likelihood using regularization to produce a hierarchical structure across different scales.

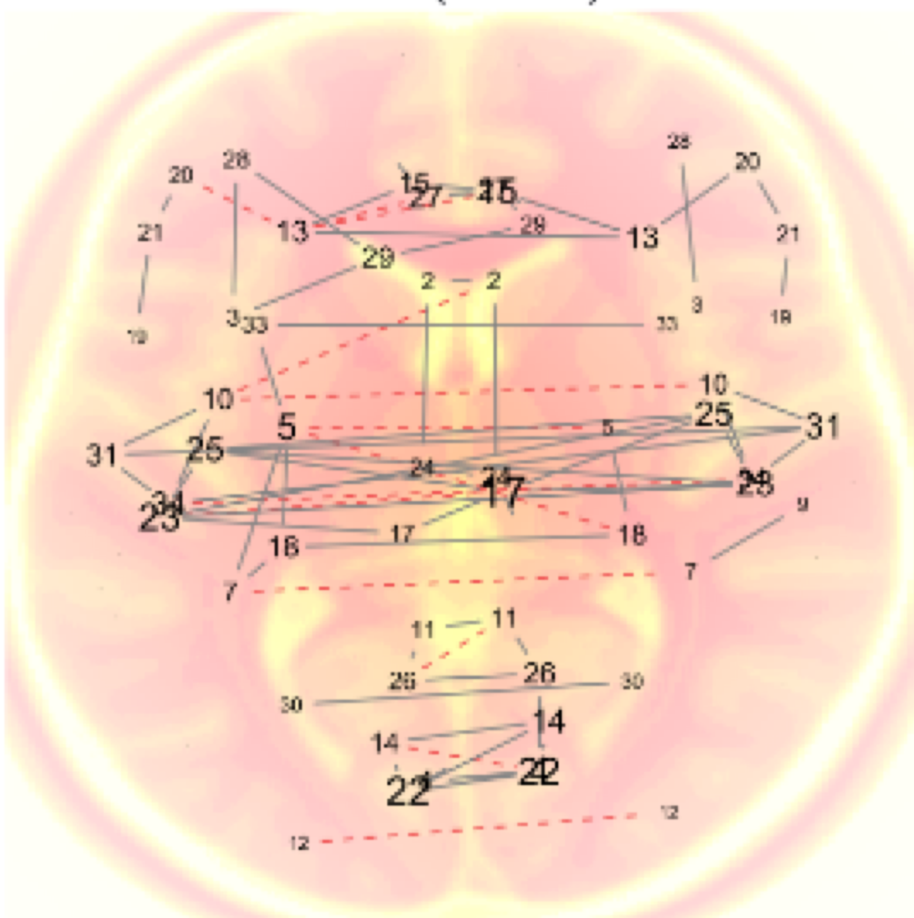
$$\min_{\mathcal{A}} \{-\mathcal{L}(\boldsymbol{\Gamma}^1, \ldots, \boldsymbol{\Theta}^K) + P_{\tilde{\boldsymbol{\lambda}}_1}^{\mathrm{u}}(\boldsymbol{\Theta}_{a,b}^{1}, \ldots, \boldsymbol{\Theta}_{a,b}^{K}) + P_{\tilde{\boldsymbol{\lambda}}_2}^{\mathrm{d}}(\boldsymbol{\Gamma}_{a,b}^{1}, \ldots, \boldsymbol{\Gamma}_{a,b}^{K})\}$$

- The optimization iterates over the scales. For $k^* \in \{1, \ldots, K\}$:
  - Step 1: Estimate Θ^{k^*} and Γ^{k^*} to solve for graph $G^{k^*}\left(E_u^{k^*}, V^{k^*}\right)$
  - Step 2: Estimate the graph at all other scales $k \neq k^*$ following the 'zoom-in' and 'zoom-out' principles

$$P^{\mathrm{u}}_{\tilde{\boldsymbol{\lambda}}_1}(\boldsymbol{\Theta}^1_{a,b}, \ldots, \boldsymbol{\Theta}^K_{a,b}) = \tilde{\lambda}_{n1} \sum_{a^* \neq b^*} |\boldsymbol{\Theta}^{k^*}_{a^*,b^*}| + \underbrace{\sum_{k'=1}^{k^*-1} \tilde{\lambda}_{n2,k'} \sum_{(a',b') \in R^{\mathrm{u}}_{k',k^*}} |\boldsymbol{\Theta}^{k'}_{a',b'}|}_{\text{zoom out undirected on coarser scales}} + \underbrace{\sum_{k''=k^*+1}^{K} \tilde{\lambda}_{n3,k''} \sum_{(a'',b'') \in R^{\mathrm{u}}_{k'',k^*}} |\boldsymbol{\Theta}^{k''}_{a'',b''}|}_{\text{zoom in undirected on finer scales}},$$

$$P^{\mathrm{d}}_{\tilde{\boldsymbol{\lambda}}_2}(\boldsymbol{\Gamma}^1_{a,b}, \ldots, \boldsymbol{\Gamma}^K_{a,b}) = \tilde{\lambda}_{n4} \sum_{a^*, b^*} |\boldsymbol{\Gamma}^{k^*}_{a^*,b^*}| + \underbrace{\sum_{k'=1}^{k^*-1} \tilde{\lambda}_{n5,k'} \sum_{(a',b') \in R^{\mathrm{d}}_{k',k^*}} |\boldsymbol{\Gamma}^{k'}_{a',b'}|}_{\text{zoom out directed on coarser scales}} + \underbrace{\sum_{k''=k^*+1}^{K} \tilde{\lambda}_{n6,k''} \sum_{(a'',b'') \in R^{\mathrm{d}}_{k'',k^*}} |\boldsymbol{\Gamma}^{k''}_{a'',b''}|}_{\text{zoom in directed on finer scales}},$$

$k = 1 (93.4\%)$



rs-fMRI data

Model estimated with three starting scales: k*=1,3,5

Full-lines: edge in all three graphs regardless of

Dotted-lines: edges present in at most two of the three graphs

Percent on top: proportion of common (full line) edges

References

- Pircalabelu et al (2020)

  Zoom-in–out joint graphical lasso for different coarseness scales
- Zhang et al (2019)

  A random covariance model for bi-level graphical modeling with application to resting-state fMRI data