# Community based grouping for undirected graphical models

**Eugen Pircalabelu**

Institute of Statistics, Biostatistics and Actuarial Sciences - UCLouvain

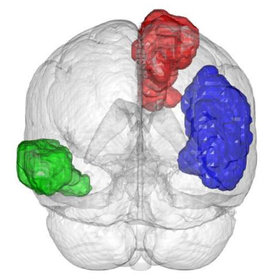
UCLouvain

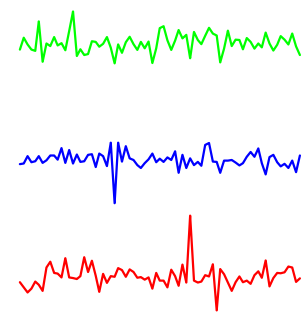
Joint work with **Gerda Claeskens**

26th Annual Meeting of the Royal Statistical Society of Belgium
Ovifat, 17-19 October, 2018

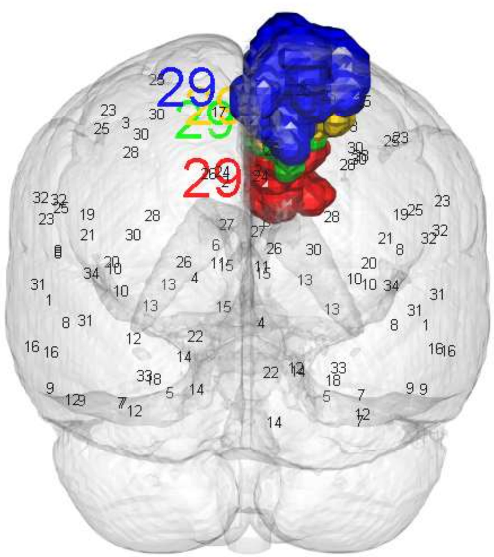
## Motivation: general fMRI setup

https://www.cedars-sinai.edu/Patients/Programs-and-Services/Imaging-Center/For-Patients/Exams-by-Procedure/MRI/





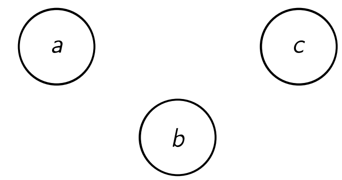
## Motivation: specific fMRI design





## Graph theoretical framework

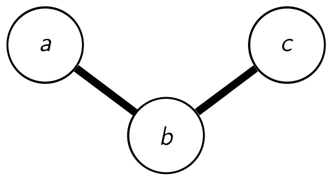
- $G = (E, V)$
- To each rv $Y_1 \dots Y_p$ a node in $V = \{1 \dots p\}$ is associated



## Graph theoretical framework

- $G = (E, V)$
- To each rv $Y_1 \dots Y_p$ a node in $V = \{1 \dots p\}$ is associated.

$(a, b) \cup (b, a) \in E \quad a - b \quad a$ neighbor of $b$

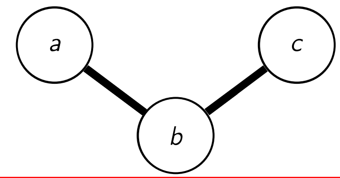





## Graph theoretical framework

- $G = (E, V)$
- To each rv $Y_1 \dots Y_p$ a node in $V = \{1 \dots p\}$ is associated.

$(a, b) \cup (b, a) \in E \quad a - b \quad a$ neighbor of $b$




- Markov property:

$$a \perp c \mid V \setminus \{a, c\} \text{ iff } (a, c) \cup (c, a) \notin E \text{ (e.g. } a \perp c | b)$$

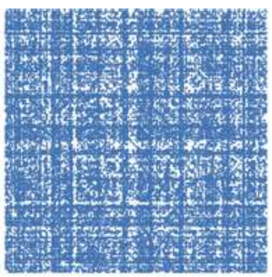
- If $\boldsymbol{Y} \sim N(\mathbf{0}, \mathbf{\Theta}^{-1})$ then $(a, c) \cup (c, a) \notin E$ iff $\mathbf{\Theta}_{a,c} = 0$
- (Unknown) Structure of $G \equiv$ Non-zero entries of $\mathbf{\Theta}$
- Enforce sparsity on $\mathbf{\Theta}$ (by shrinking ‘small’ coeff. to 0) crucial when $p \gg n$

## Network theoretical framework

- Network data: represented by a graph with n nodes via Adjacency matrix

$$\boldsymbol{A}_{n\times n},\ \text{with } A_{a,b} = \begin{cases} 1 & \text{if there is an edge between nodes } a \text{ and } b \\ 0 & \text{otherwise} \end{cases}$$

- For each node $a$ let $\boldsymbol{Z}_a \in \{0,1\}^K$ be a (unobserved) labeling vector
- Let $\boldsymbol{B}_{K\times K}$: specify prob. of edges within/between the communities
- The rv. $A_{a,b}$ iid Bernoulli with $\mathrm{E}(A_{a,b}|\boldsymbol{Z}_a, \boldsymbol{Z}_b) = \boldsymbol{Z}_a^\top \boldsymbol{B} \boldsymbol{Z}_b$
  *Membership* matrix: $\boldsymbol{Z}_{n\times K} = [\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n]^\top$ then $\mathrm{E}(A|\boldsymbol{Z}) = \boldsymbol{Z}\boldsymbol{B}\boldsymbol{Z}^\top$



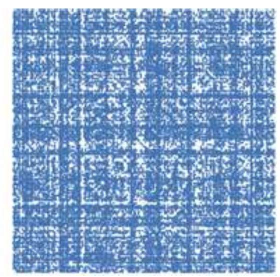


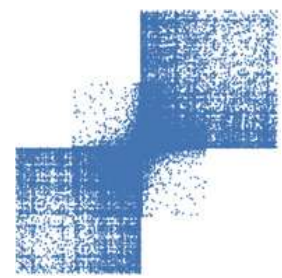
## Network theoretical framework

- Network data: represented by a graph with n nodes via Adjacency matrix

$$\boldsymbol{A}_{n\times n},\ \text{with } A_{a,b} = \begin{cases} 1 & \text{if there is an edge between nodes } a \text{ and } b \\ 0 & \text{otherwise} \end{cases}$$

- For each node $a$ let $\boldsymbol{Z}_a \in \{0,1\}^K$ be a (unobserved) labeling vector
- Let $\boldsymbol{B}_{K\times K}$: specify prob. of edges within/between the communities
- The rv. $A_{a,b}$ iid Bernoulli with $\mathrm{E}(A_{a,b}|\boldsymbol{Z}_a, \boldsymbol{Z}_b) = \boldsymbol{Z}_a^\top \boldsymbol{B} \boldsymbol{Z}_b$
  *Membership* matrix: $\boldsymbol{Z}_{n\times K} = [\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n]^\top$ then $\mathrm{E}(A|\boldsymbol{Z}) = \boldsymbol{Z}\boldsymbol{B}\boldsymbol{Z}^\top$





Olhede, S. C. and Wolfe, P. J. (2014). Network histograms and universality of blockmodel approximation. PNAS, 11(41), 14722-14727.

5 / 13

## Community-based group graphical lasso

What:

- an 'all-at-once' procedure that bridges PGMs and networks
- joint estimation of (i) edges and (ii) communities of similar nodes

Why:

- functional connectivity within the brain
- evaluate homogeneity/similarity of nodes in the graph

How:

- group $\ell_1$ penalized estimation of the graph
- unknown grouping, but estimable from the data
- Network-like approach to estimate the communities



## Proposed model

- Suppose conditional on $\boldsymbol{Z}_{p\times K}$

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\epsilon}$$

  ▷ $\boldsymbol{u}$ is a random effects vector of length $K$, $\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{I}_{K\times K})$
  ▷ $\boldsymbol{\epsilon}$ is a vector of random errors of length $p$, $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_{p\times p})$
  ▷ $\boldsymbol{\epsilon} \perp \boldsymbol{u}$

$$\mathrm{Var}(\boldsymbol{Y}|\boldsymbol{Z}) = \boldsymbol{Z}\mathrm{Var}(\boldsymbol{u})\boldsymbol{Z}^{\top} + \mathrm{Var}(\boldsymbol{\epsilon}) = \boldsymbol{Z}\boldsymbol{Z}^{\top} + \boldsymbol{\Sigma}$$

$$\boldsymbol{Z}\mathrm{Var}(\boldsymbol{u})\boldsymbol{Z}^{\top} \approxeq \boldsymbol{Z}\boldsymbol{B}\boldsymbol{Z}^{\top} \text{ (expect. of adjacency matrix in SBM)}$$

- Recovering communities using covariance info. from data at the nodes

$$\neq$$

recovering communities from random adjacency matrices

## Proposed model (cont'd)

- Given $n$ i.i.d copies of $\boldsymbol{Y}$, ie $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$, suppose we use
$$\boldsymbol{Z}\boldsymbol{Z}^\mathsf{T} + \boldsymbol{S}$$
as estimator for the conditional variance $\mathrm{Var}(\boldsymbol{Y}|\boldsymbol{Z})$
▷ $\boldsymbol{S} = (1/n)\sum_{i=1}^n \boldsymbol{Y}_i \boldsymbol{Y}_i^\mathsf{T}$ is the empirical covariance matrix

- Conditional on the membership matrix
$$\boldsymbol{Y}|\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{\Theta}^{-1})$$
▷ $\boldsymbol{\Theta}$ : inverse covariance matrix (or the concentration matrix)

- Negative log-likelihood proportional to
$$\mathcal{L}(\boldsymbol{\Theta}) = -\log\det\boldsymbol{\Theta} + \mathrm{tr}\{(\boldsymbol{S} + \boldsymbol{Z}\boldsymbol{Z}^\mathsf{T})\boldsymbol{\Theta}\} = -\log\det\boldsymbol{\Theta} + \mathrm{tr}(\boldsymbol{S}\boldsymbol{\Theta}) + \mathrm{tr}(\boldsymbol{Z}\boldsymbol{Z}^\mathsf{T}\boldsymbol{\Theta})$$

- In practice $\boldsymbol{Z}$ is unknown, but exact recovery is NP-hard $\Rightarrow$ relaxation



## Objective function for ComGGL

$$\min_{\boldsymbol{\Theta},\boldsymbol{X}} \Big( \underbrace{\mathrm{tr}(\boldsymbol{S}\boldsymbol{\Theta}) - \log\det\boldsymbol{\Theta}}_{\mathcal{L}_0} + \underbrace{\mathrm{tr}(\boldsymbol{X}\boldsymbol{\Theta})}_{\mathcal{L}_1} + \underbrace{\lambda_{n1}\sum_{a\neq b}|\boldsymbol{\Theta}_{a,b}| + \lambda_{n2}\sum_{k=1}^{K_n}\big(\sum_{a\neq b\in\mathcal{C}_k}(\boldsymbol{\Theta}_{a,b}^k)^2\big)^{1/2}}_{\mathcal{L}_2} \Big)$$

s.t.

(i) $\boldsymbol{\Theta} \succ 0$ (positive definite), $\boldsymbol{X} \succeq 0$ (positive semi-definite),
(ii) $0 \leq \boldsymbol{X}_{a,b} \leq 1,\ \boldsymbol{X}_{a,a} = 1$
(iii) the membership of nodes to the $k$th community (ie. $\mathcal{C}_k$) depends on $\boldsymbol{X}$
(iv) $\lambda_{n1}$ and $\lambda_{n2}$ are assumed to be positive and known
(v) $K_n$ is assumed to be a known (estimable) positive integer

## Objective function for ComGGL

$$\min_{\boldsymbol{\Theta},\boldsymbol{X}} \left( \underbrace{\mathrm{tr}(\boldsymbol{S}\boldsymbol{\Theta}) - \log\det\boldsymbol{\Theta}}_{\mathcal{L}_0} + \underbrace{\mathrm{tr}(\boldsymbol{X}\boldsymbol{\Theta})}_{\mathcal{L}_1} + \underbrace{\lambda_{n1}\sum_{a\neq b}|\Theta_{a,b}| + \lambda_{n2}\sum_{k=1}^{K_n}\Big(\sum_{a\neq b\in\mathcal{C}_k}(\Theta^k_{a,b})^2\Big)^{1/2}}_{\mathcal{L}_2} \right)$$

s.t.

(i) $\boldsymbol{\Theta} \succ 0$ (positive definite), $\boldsymbol{X} \succeq 0$ (positive semi-definite),
(ii) $0 \leq \boldsymbol{X}_{a,b} \leq 1, \ \boldsymbol{X}_{a,a} = 1$
(iii) the membership of nodes to the $k$th community (ie. $\mathcal{C}_k$) depends on $\boldsymbol{X}$
(iv) $\lambda_{n1}$ and $\lambda_{n2}$ are assumed to be positive and known
(v) $K_n$ is assumed to be a known (estimable) positive integer

$\mathcal{L}_0$ log-likelihood if the nodes were unlabeled



## Objective function for ComGGL

$$\min_{\boldsymbol{\Theta},\boldsymbol{X}} \left( \underbrace{\mathrm{tr}(\boldsymbol{S}\boldsymbol{\Theta}) - \log\det\boldsymbol{\Theta}}_{\mathcal{L}_0} + \underbrace{\mathrm{tr}(\boldsymbol{X}\boldsymbol{\Theta})}_{\mathcal{L}_1} + \underbrace{\lambda_{n1}\sum_{a\neq b}|\Theta_{a,b}| + \lambda_{n2}\sum_{k=1}^{K_n}\Big(\sum_{a\neq b\in\mathcal{C}_k}(\Theta^k_{a,b})^2\Big)^{1/2}}_{\mathcal{L}_2} \right)$$

s.t.

(i) $\boldsymbol{\Theta} \succ 0$ (positive definite), $\boldsymbol{X} \succeq 0$ (positive semi-definite),
(ii) $0 \leq \boldsymbol{X}_{a,b} \leq 1, \ \boldsymbol{X}_{a,a} = 1$
(iii) the membership of nodes to the $k$th community (ie. $\mathcal{C}_k$) depends on $\boldsymbol{X}$
(iv) $\lambda_{n1}$ and $\lambda_{n2}$ are assumed to be positive and known
(v) $K_n$ is assumed to be a known (estimable) positive integer

$\mathcal{L}_0$ log-likelihood if the nodes were unlabeled

$\mathcal{L}_1$ relates the graph info. (through $\boldsymbol{\Theta}$) to the labeling info. (through $\boldsymbol{X}$)
- ▷ structures the graph similarly to an SBM
- ▷ membership will be estimated using spectral methods on $\boldsymbol{X}$

## Objective function for ComGGL

$$\min_{\Theta, X} \Big( \underbrace{\operatorname{tr}(S\Theta) - \log\det\Theta}_{\mathcal{L}_0} + \underbrace{\operatorname{tr}(X\Theta)}_{\mathcal{L}_1} + \underbrace{\lambda_{n1}\sum_{a\neq b}|\Theta_{a,b}| + \lambda_{n2}\sum_{k=1}^{K_n}\big(\sum_{a\neq b\in\mathcal{C}_k}(\Theta^k_{a,b})^2\big)^{1/2}}_{\mathcal{L}_2} \Big)$$

s.t.

(i) $\Theta \succ 0$ (positive definite), $X \succeq 0$ (positive semi-definite),

(ii) $0 \leq X_{a,b} \leq 1,\ X_{a,a} = 1$

(iii) the membership of nodes to the $k$th community (ie. $\mathcal{C}_k$) depends on $X$

(iv) $\lambda_{n1}$ and $\lambda_{n2}$ are assumed to be positive and known

(v) $K_n$ is assumed to be a known (estimable) positive integer

$\mathcal{L}_0$ log-likelihood if the nodes were unlabeled

$\mathcal{L}_1$ relates the graph info. (through $\Theta$) to the labeling info. (through $X$)
- ▷ structures the graph similarly to an SBM
- ▷ membership will be estimated using spectral methods on $X$

$\mathcal{L}_2$ effect of the grouping of the nodes on the estimation of the graph
- ▷ $\ell_1$-term that shrinks small entries of $\Theta$ to 0 (sparsity)
- ▷ grouping term to make entries in the community more similar



## Properties

Frobenius norm convergence:

(1) Under suitable regularity conditions there exist estimators $(\hat{\Theta}, \hat{X})$ obtained based on the objective function $\ell(\Theta, X)$ s.t.

$$\max(||\hat{\Theta}-\Theta_0||_F, ||\hat{X}-ZZ^{\mathsf{T}}||_F) = O_p\Big(\max\Big\{\sqrt{(p_n+s_n)\frac{\log p_n}{n}}, \sqrt{\frac{p_n^2}{nK_n}-\frac{p_n}{n}}\Big\}\Big).$$

Sparsistency:

(2) Under suitable regularity conditions, for estimators $(\hat{\Theta}, \hat{X})$ based on the objective function $\ell(\Theta, X)$ that satisfy (i) $||\hat{\Theta}-\Theta_0|| = O_p(\sqrt{\eta_{n1}})$ and (ii) $||\hat{X}-ZZ^{\mathsf{T}}|| = O_p(\sqrt{\eta_{n2}})$ for sequences $\eta_{n1}, \eta_{n2} \to 0$ if

$$\sqrt{\frac{\log p_n}{n}} + \sqrt{\eta_{n1}} + \sqrt{\eta_{n2}} + \lambda_{n2}\Theta^k_{a,b}\Big/\sqrt{\sum_{a\neq b\in\mathcal{C}_k}(\Theta^k_{a,b})^2} = O(\lambda_{n1}),$$

with probability tending to 1, $\hat{\Theta}_{a,b} = 0$ for all $(a,b) \in \mathcal{S}^c$ from the $k$-th community.

## Properties (cont'd)

Spectral clustering:

(3) Let $\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathsf{T}}$ be the eigen decomp. of $\boldsymbol{Z}\boldsymbol{Z}^{\mathsf{T}}$. There exists a matrix $\boldsymbol{W}_{K_n \times K_n}$ with real elements s.t. $\boldsymbol{Q} = \boldsymbol{Z}\boldsymbol{W}$ and

$$||\boldsymbol{W}_l - \boldsymbol{W}_m|| = \{(\#\mathcal{C}_l)^{-1} + (\#\mathcal{C}_m)^{-1}\}^{1/2} \qquad \forall 1 \leq l < m \leq K_n.$$

▷ Eigenvectors $\boldsymbol{Q}$ contain info. about the community membership matrix $\boldsymbol{Z}$



## Properties (cont'd)

Spectral clustering:

(3) Let $\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathsf{T}}$ be the eigen decomp. of $\boldsymbol{Z}\boldsymbol{Z}^{\mathsf{T}}$. There exists a matrix $\boldsymbol{W}_{K_n \times K_n}$ with real elements s.t. $\boldsymbol{Q} = \boldsymbol{Z}\boldsymbol{W}$ and

$$||\boldsymbol{W}_l - \boldsymbol{W}_m|| = \{(\#\mathcal{C}_l)^{-1} + (\#\mathcal{C}_m)^{-1}\}^{1/2} \qquad \forall 1 \leq l < m \leq K_n.$$

▷ Eigenvectors $\boldsymbol{Q}$ contain info. about the community membership matrix $\boldsymbol{Z}$

▷ SC with $k$-means:

$$(\hat{\boldsymbol{Z}}, \hat{\boldsymbol{W}}) = \arg \min_{\boldsymbol{Z} \in \mathbb{M}_{p_n \times K_n}, \boldsymbol{W} \in \mathbb{R}_{K_n \times K_n}} ||\boldsymbol{Z}\boldsymbol{W} - \hat{\boldsymbol{Q}}||_F^2$$

where $\hat{\boldsymbol{Q}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{Q}}^{\mathsf{T}}$ is the $K_n$-dimensional eigen decomp. of $\boldsymbol{X}$.

## Properties (cont'd)

Spectral clustering:

(3) Let $\boldsymbol{Q\Lambda Q}^{\top}$ be the eigen decomp. of $\boldsymbol{ZZ}^{\top}$. There exists a matrix $\boldsymbol{W}_{K_n \times K_n}$ with real elements s.t. $\boldsymbol{Q} = \boldsymbol{ZW}$ and

$$||\boldsymbol{W}_l - \boldsymbol{W}_m|| = \{(\#\mathcal{C}_l)^{-1} + (\#\mathcal{C}_m)^{-1}\}^{1/2} \qquad \forall 1 \le l < m \le K_n.$$

▷ Eigenvectors $\boldsymbol{Q}$ contain info. about the community membership matrix $\boldsymbol{Z}$

▷ SC with $k$-means:

$$(\hat{\boldsymbol{Z}}, \hat{\boldsymbol{W}}) = \arg \min_{\boldsymbol{Z} \in \mathbb{M}_{p_n \times K_n}, \boldsymbol{W} \in \mathbb{R}_{K_n \times K_n}} ||\boldsymbol{ZW} - \hat{\boldsymbol{Q}}||_F^2$$

where $\hat{\boldsymbol{Q}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{Q}}^{\top}$ is the $K_n$-dimensional eigen decomp. of $\boldsymbol{X}$.

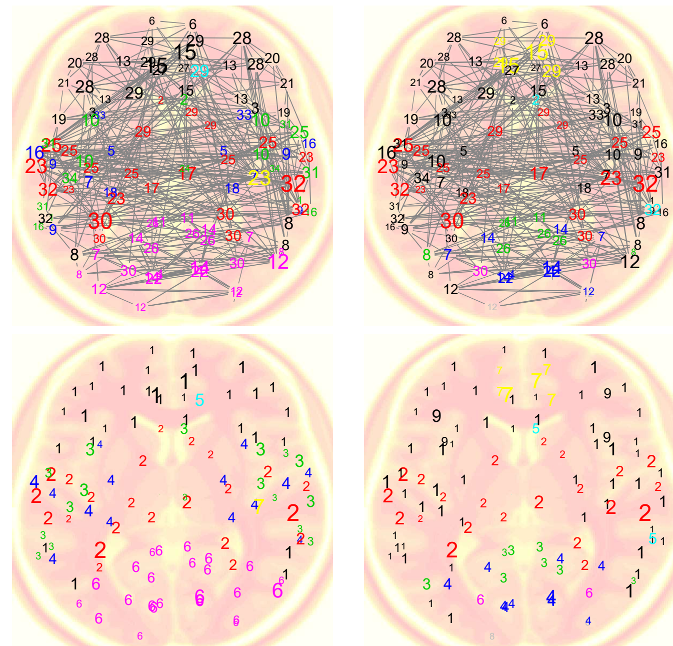
Labeling consistency:

(4) Let $S_k$ denote the sets of misclassified nodes from the $k$th community and $\hat{\boldsymbol{Z}}$ be the result of the spectral clustering. There exists a constant $c > 0$ s.t.

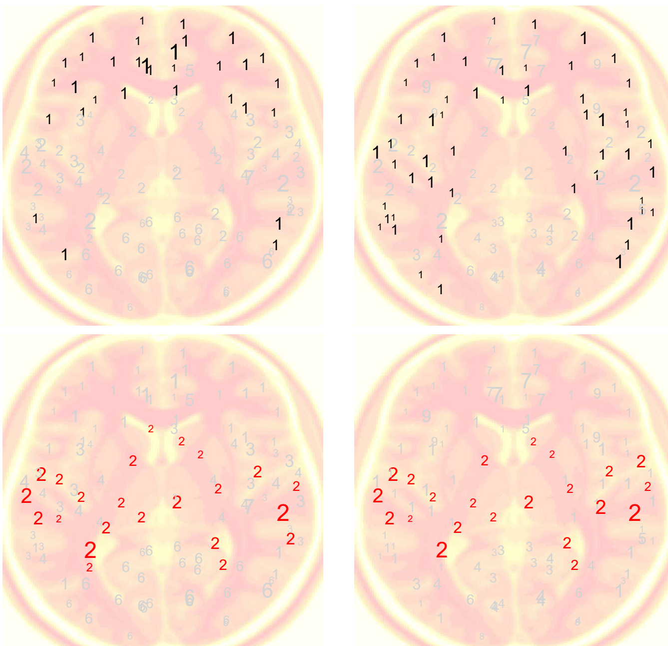
$$\sum_{k=1}^{K_n} \#S_k / \#\mathcal{C}_k \le c^{-1}(2+\xi) n^{-1/2} (p_n^2 - K_n p_n)^{1/2}.$$



## Graphs and Communities: ComGGL vs. CORD

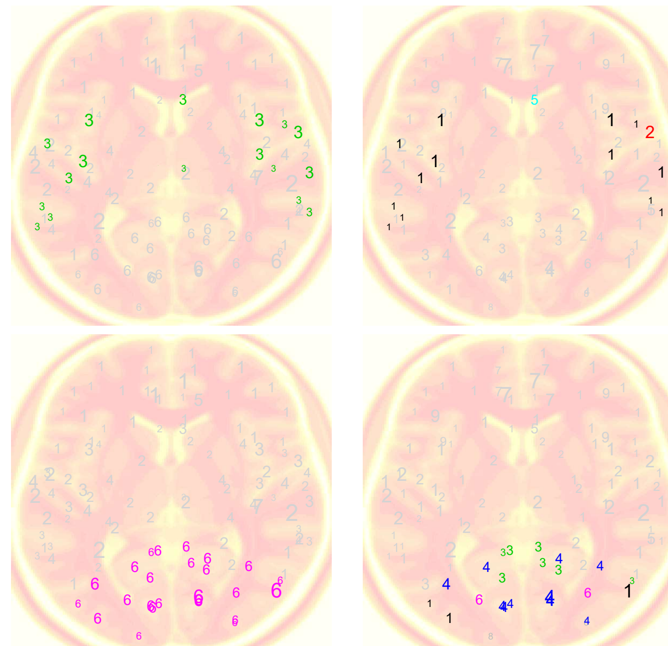


# Graphs and Communities: ComGGL vs. CORD



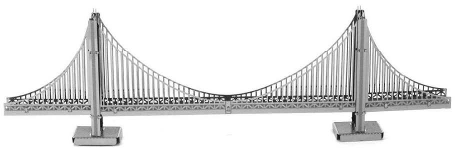


# Graphs and Communities: ComGGL vs. CORD



12 / 13

## Take-home message

- Estimation of brain pathways with penalized undirected graphs
- Joint estimation of the graph and underlying communities
- Account for communities when estimating the graph
- Account for graph when estimating the communities
- PGM  Network Analysis

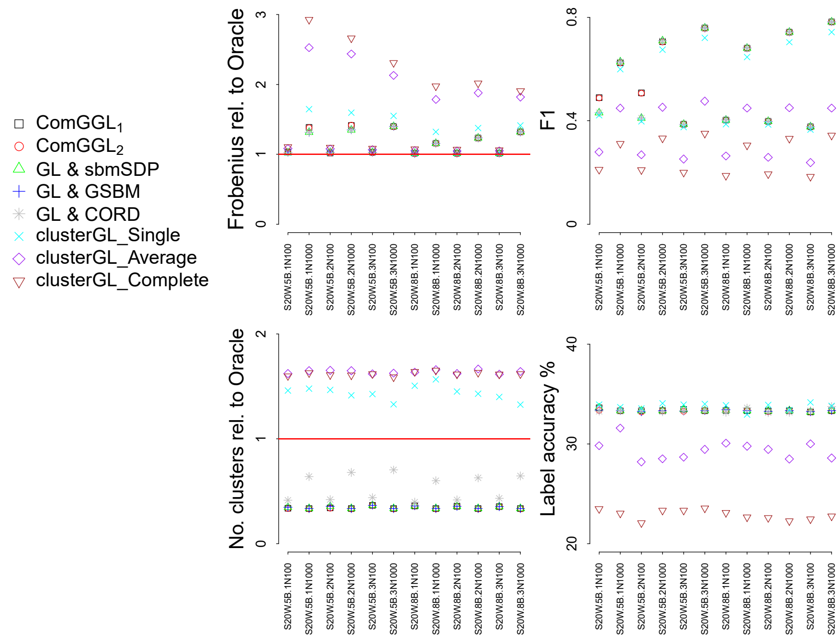
## Simulation settings

Sample size $n = 100$ or $1000$
Number of communities $K = 3$
Number of nodes $(p^1, p^2, p^3) = (20, 20, 20)$
Prob. edges within community $\pi_w = .5$ or $.8$
Prob. edges between communities $\pi_b = .1$, $.2$ or $.3$

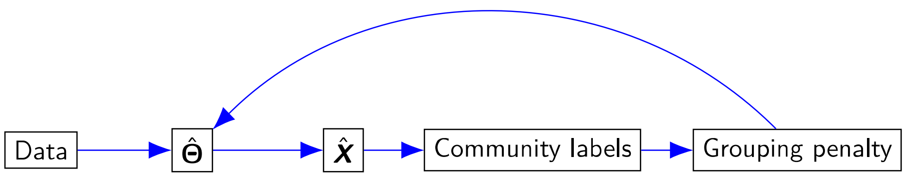
Performance measures (closer to 1 is better):

- Frobenius norm relative to the oracle (knows $G(E, V)$, $K$ and $\boldsymbol{Z}$):
  $\text{Fr} = || \hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^0 ||_F = \sqrt{\sum_{i=1}^{p} \sum_{j=1}^{p} |\hat{\theta}_{ij} - \theta_{ij}|^2}$
- $F_1 = \frac{2PR}{P+R}$ where $\begin{cases} P = \frac{\#\text{estimated edges that are true edges}}{\#\text{estimated edges}} \\ R = \frac{\#\text{estimated edges that are true edges}}{\#\text{true edges}} = TPR \end{cases}$
- $K$ accuracy relative to oracle
- Labeling accuracy.

## Simulation results

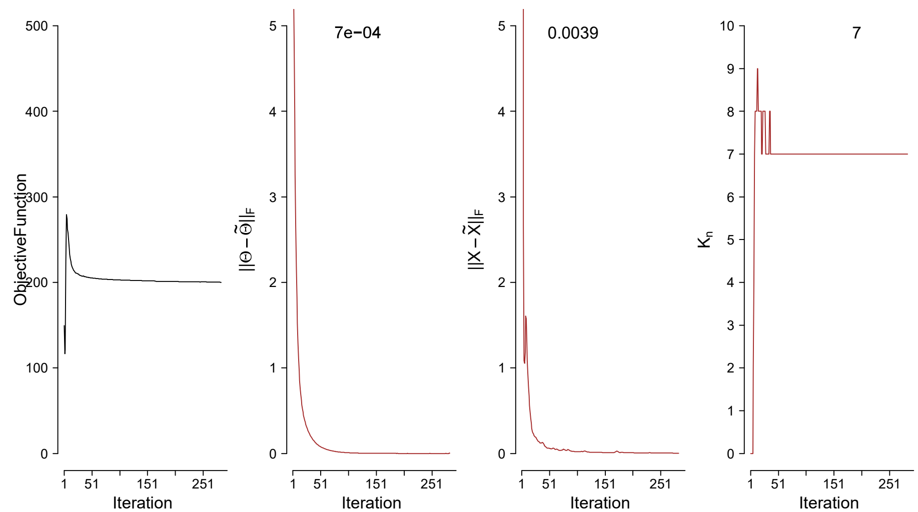


## Computational aspects for ComGGL




- $\boldsymbol{\Theta}$ and the community structure depend on one another
- structure of the communities is informative for the estimation of $\boldsymbol{\Theta}$
- to estim. communities we need $\boldsymbol{\Theta}$ & to estim. $\boldsymbol{\Theta}$ we need the communities
- ADMM algorithm where we update $\boldsymbol{\Theta}|\boldsymbol{X}$ and $\boldsymbol{X}|\boldsymbol{\Theta}$ until convergence
- Convergence follows due to biconvexity
- Complexity $O(p^3)$ (due to eigen decomposition)

## Convergence fMRI example



## ROI names

| ROI | Name | ROI | Name | ROI | Name |
|---|---|---|---|---|---|
| 1 | Bankssts | 12 | Lateraloccipital | 23 | Postcentral |
| 2 | Caudalanteriorcingulate | 13 | Lateralorbitofrontal | 24 | Posteriorcingulate |
| 3 | Caudalmiddlefrontal | 14 | Lingual | 25 | Precentral |
| 4 | Cuneus | 15 | Medialorbitofrontal | 26 | Precuneus |
| 5 | Entorhinal | 16 | Middletemporal | 27 | Rostralanteriorcingulate |
| 6 | Frontalpole | 17 | Paracentral | 28 | Rostralmiddlefrontal |
| 7 | Fusiform | 18 | Parahippocampal | 29 | Superiorfrontal |
| 8 | Inferiorparietal | 19 | Parsopercularis | 30 | Superiorparietal |
| 9 | Inferiortemporal | 20 | Parsorbitalis | 31 | Superiortemporal |
| 10 | Insula | 21 | Parstriangularis | 32 | Supramarginal |
| 11 | Isthmuscingulate | 22 | Pericalcarine | 33 | Temporalpole |
| | | | | 34 | Transversetemporal |

Table: fMRI data. Correspondence between numbers and names of the regions of interest.