

Statistical Inference with Multi-layered Graphical Models

Subhabrata Majumdar

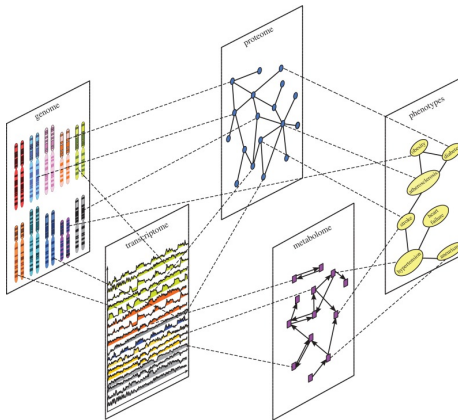
University of Florida Informatics Institute

AT&T Labs - Research, Bedminster, NJ

April 1, 2018

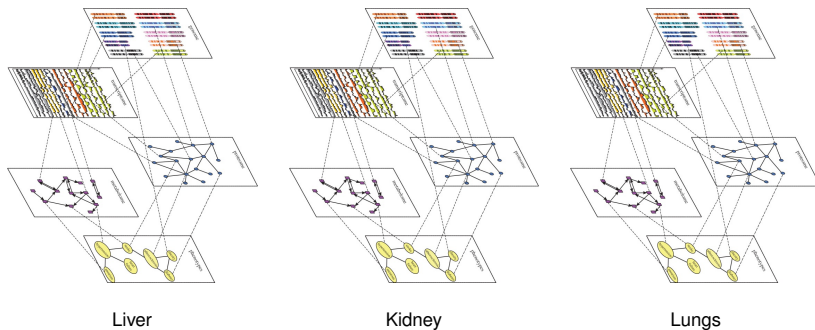
Summary

- Biological processes in the body have a natural hierarchical structure, e.g. **Gene > Protein > Metabolite**;
- There are within layer and between-layer connections in this structure.



Source: **Gligorijević and Pržulj (2015)**

These connections can be different inside different organs, experimental conditions, or for different subtypes of the same disease;



A general model for data integration.

In this work we propose a general statistical framework based on graphical models for *horizontal* (i.e. across conditions or subtypes) *and vertical* (i.e. across different layers containing data on molecular compartments) *integration of information* in data from such complex biological structures.

Specifically, we perform *joint estimation and hypothesis testing* for all the connections in these structures.

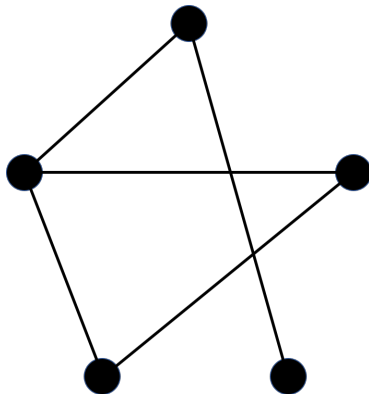
1 **Multiple multi-level graphical models**

2 Model formulation

3 Hypothesis testing

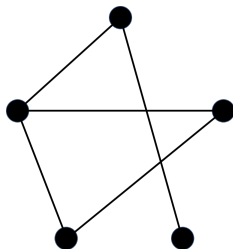
4 Simulation studies

$$\mathbb{X} = (X_1, \dots, X_p)^T \sim \mathcal{N}_p(0, \Sigma_x); \quad \Omega_x = \Sigma_x^{-1}$$

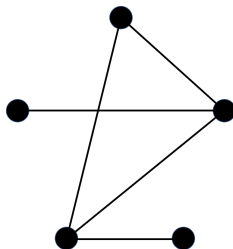


Sparse estimation of Ω_x : [Meinshausen and Bühlmann \(2006\)](#)
Multiple testing and error control: [Drton and Perlman \(2007\)](#).

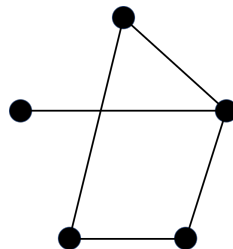
$$\mathbb{X}^k = (X_1^k, \dots, X_p^k)^T \sim \mathcal{N}_p(0, \Sigma_x^k); \quad \Omega_x^k = (\Sigma_x^k)^{-1}$$
$$k = 1, 2, \dots, K$$



$k = 1$



$k = 2$



$k = 3$

- Joint estimation of $\{\Omega_x^k\}$: Guo et al. (2011); Ma and Michailidis (2016)
- Difference and similarity testing with FDR control: Liu (2017+)

Multi-Layered Gaussian Graphical models

$$\mathbb{E} = (E_1, \dots, E_q)^T \sim \mathcal{N}_p(0, \Sigma_y);$$

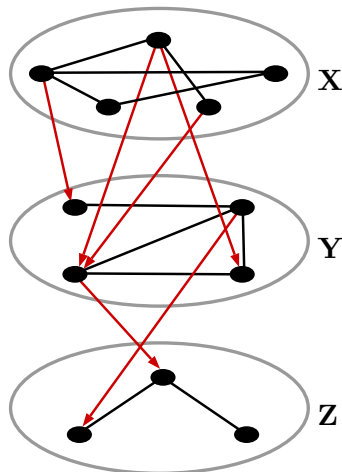
$$\mathbb{F} = (F_1, \dots, F_r)^T \sim \mathcal{N}_r(0, \Sigma_z);$$

$$\Omega_y = (\Sigma_y)^{-1}, \Omega_z = (\Sigma_z)^{-1}$$

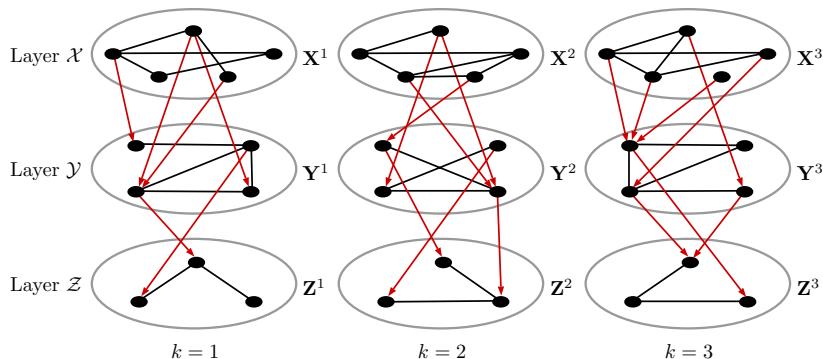
$$\mathbb{Y} = \mathbb{X}\mathbf{B} + \mathbb{E},$$

$$\mathbb{Z} = \mathbb{Y}\mathbf{C} + \mathbb{F}.$$

- $\Omega_x, \Omega_y, \Omega_z$ give undirected within-layer edges, while \mathbf{B}, \mathbf{C} gives directed between-layer edges.
- Sparse estimation of the components: [Lin et al. \(2016\)](#).
- Testing: ??



Multiple Multi-layered Gaussian Graphical models



$$\mathbb{E}^k = (E_1^k, \dots, E_q^k)^T \sim \mathcal{N}_q(0, \Sigma_y^k); \quad \Omega_y^k = (\Sigma_y^k)^{-1}$$

$$\mathbb{F}^k = (F_1^k, \dots, F_r^k)^T \sim \mathcal{N}_r(0, \Sigma_z^k); \quad \Omega_z^k = (\Sigma_z^k)^{-1}$$

$$\mathbb{Y}^k = \mathbb{X}^k \mathbf{B}^k + \mathbb{E}^k,$$

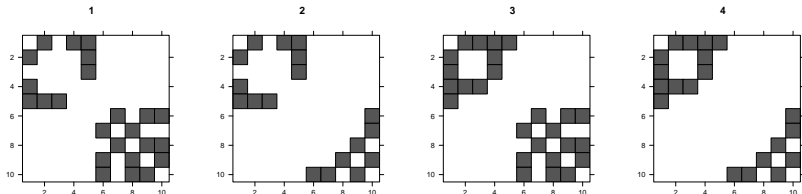
$$\mathbb{Z}^k = \mathbb{Y}^k \mathbf{C}^k + \mathbb{F}^k; \quad k = 1, 2, \dots, K$$

- We estimate $\{\Omega_x^k, \Omega_y^k, \mathbf{B}^k\}$ jointly for all k from a single model;
 - Incorporate structural information using group sparsity,
 - Propose alternating block algorithm to compute solutions,
 - Derive convergence properties of solutions.
- Devise a full pairwise testing procedure for rows of \mathbf{B}^k ;
 - Propose a debiasing method for estimates of individual rows of \mathbf{B}^k , say $\mathbf{b}_i^k, i \in \{1, 2, \dots, p\}$ and derive its asymptotic distribution;
 - For $K = 2$, propose a test for row-wise differences $\mathbf{b}_i^1 - \mathbf{b}_i^2$;
 - Perform multiple testing for elementwise differences $b_{ij}^1 = b_{ij}^2, j = 1, 2, \dots, q$ within a row.
- Use simulations for performance evaluation.

- 1 Multiple multi-level graphical models
- 2 Model formulation**
- 3 Hypothesis testing
- 4 Simulation studies

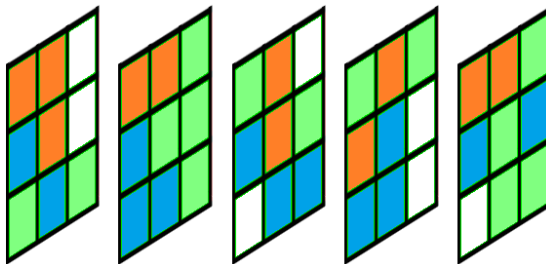
- $\mathcal{Y} = \{\mathbf{Y}^1, \dots, \mathbf{Y}^K\}, \mathcal{X} = \{\mathbf{X}^1, \dots, \mathbf{X}^K\};$
- $\Omega_x = \{\Omega_x^1, \dots, \Omega_x^K\}, \Omega_y = \{\Omega_y^1, \dots, \Omega_y^K\}, \mathcal{B} = \{\mathbf{B}^1, \dots, \mathbf{B}^K\};$

Group structures in X-network is denoted by $\mathcal{G}_x = \{\mathcal{G}_{x,ii'} : i \neq i', 1 \leq i, i' \leq p\}$. Each $\mathcal{G}_{x,ii'}$ is a partition of $\{1, \dots, K\}$ denoting grouping over k for the $(i, i')^{\text{th}}$ elements of the X-precision matrices.



Define $\mathcal{G}_y = \{\mathcal{G}_{y,jj'} : j \neq j', 1 \leq j, j' \leq q\}$ similarly.

Group structures in \mathcal{B} is denoted by \mathcal{H} , with each $h \in \mathcal{H}$ being a collection of 3-tuples (h_i, h_j, h_k) so that $1 \leq h_i \leq p, 1 \leq h_j \leq q, 1 \leq h_k \leq K$.



- **For single GGM:** Estimate neighboring edges for each node, then refit.

$$\hat{\zeta}_i = \underset{\zeta_i}{\operatorname{argmin}} \left\{ \frac{1}{n} \|\mathbf{X}_i - \mathbf{X}_{-i} \zeta_i\|^2 + \nu_n \sum_{i' \neq i} |\zeta_{ii'}| \right\};$$

$$\hat{\Omega}_x = \underset{\Omega_x \in \cup_i \operatorname{supp}(\zeta_i)}{\operatorname{argmin}} \{ \operatorname{Tr}(\mathbf{S}_x \Omega_x) + \log \det(\Omega_x) \}$$

Can do this because $\zeta_{ii'} = -\omega_{ii'}/\omega_{ii}$, so zeros of the precision matrix and neighborhood matrix are same.

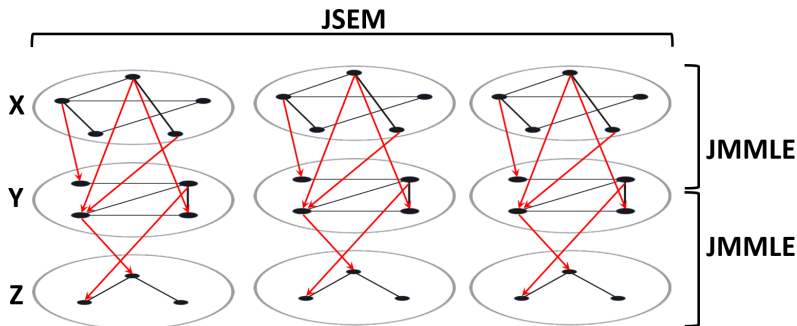
- **For multiple GGM:** Incorporate penalty across different k (JSEM: **Ma and Michailidis (2016)**).

$$\hat{\zeta}_i = \underset{\zeta_i}{\operatorname{argmin}} \left\{ \sum_{k=1}^K \frac{1}{n_k} \|\mathbf{X}_i^k - \mathbf{X}_{-i}^k \zeta_i^k\|^2 + \nu_n \sum_{i' \neq i, g \in \mathcal{G}_{x, ii'}} \|\zeta_{ii'}^{[g]}\| \right\};$$

$$\hat{\Omega}_x^k = \underset{\Omega_x^k \in \cup_i \operatorname{supp}(\zeta_i^k)}{\operatorname{argmin}} \{ \operatorname{Tr}(\mathbf{S}_x^k \Omega_x^k) + \log \det(\Omega_x^k) \};$$

Joint Multiple Multi-Level Estimation (JMMLE)

We decompose the multi-layer problem into a series of two layer problems. For within-layer connections in the topmost layer, we use JSEM. For other connections we use JMMLE.



We combine sparse neighborhood selection in the Y-network with sparse estimation of $\{\mathbf{B}^k\}$.

Take $\Theta_j = (\theta_j^1, \dots, \theta_j^K)$, $\Theta = \{\Theta_j\}_{j=1}^q$.

$$\{\hat{\mathcal{B}}, \hat{\Theta}\} = \operatorname{argmin}_{\mathcal{B}, \Theta} \left\{ \sum_{k=1}^K \frac{1}{n_k} \sum_{j=1}^q \left\| \mathbf{y}_j^k - (\mathbf{y}_{-j}^k - \mathbf{x}^k \mathbf{B}_{-j}^k) \theta_j^k - \mathbf{x}^k \mathbf{B}_j^k \right\|^2 \right. \\ \left. + \lambda_n \sum_{h \in \mathcal{H}} \|\mathbf{B}^{[h]}\| + \gamma_n \sum_{j' \neq j, g \in \mathcal{G}_{jj'}} \|\theta_{jj'}^{[g]}\| \right\}$$

$$\hat{\Omega}_y^k = \operatorname{argmin}_{\Omega_y^k \in \cup_i \operatorname{supp}(\theta_i^k)} \{ \operatorname{Tr}(\mathbf{S}_y^k \Omega_y^k) + \log \det(\Omega_y^k) \} \quad k = 1, 2, \dots, K$$

$$\{\hat{\mathcal{B}}, \hat{\Theta}\} = \operatorname{argmin}_{\mathcal{B}, \Theta} \{f(\mathcal{Y}, \mathcal{X}, \mathcal{B}, \Theta) + P(\mathcal{B}) + Q(\Theta)\}$$

The objective function is biconvex, so we solve the above by the following alternating iterative algorithm:

- 1 Start with initial estimates of \mathcal{B} and Θ , say $\mathcal{B}^{(0)}, \Theta^{(0)}$.
- 2 Iterate:

$$\begin{aligned}\mathcal{B}^{(t+1)} &= \operatorname{argmin}_{\mathcal{B}} \left\{ f(\mathcal{Y}, \mathcal{X}, \mathcal{B}, \Theta^{(t)}) + Q(\mathcal{B}) \right\} \\ \Theta^{(t+1)} &= \operatorname{argmin}_{\Theta} \left\{ f(\mathcal{Y}, \mathcal{X}, \mathcal{B}^{(t+1)}, \Theta) + P(\Theta) \right\}\end{aligned}$$

- 3 Continue till convergence.

The two subproblems

$$\begin{aligned}\hat{\mathbf{B}} &= \underset{\mathbf{B}}{\operatorname{argmin}} \left\{ \sum_{k=K}^q \frac{1}{n_k} \sum_{j=1}^q \left\| \mathbf{Y}_j^k - (\mathbf{Y}_{-j}^k - \mathbf{X}^k \mathbf{B}_{-j}^k) \hat{\boldsymbol{\theta}}_j^k - \mathbf{X}^k \mathbf{B}_j^k \right\|^2 + \lambda_n \sum_{h \in \mathcal{H}} \|\mathbf{B}^{[h]}\| \right\} \\ &= \underset{\mathbf{B}}{\operatorname{argmin}} \left\{ \sum_{k=1}^K \frac{1}{n_k} \left\| (\mathbf{Y}^k - \mathbf{X}^k \mathbf{B}^k) \hat{\mathbf{T}}^k \right\|_F^2 + \lambda_n \sum_{h \in \mathcal{H}} \|\mathbf{B}^{[h]}\| \right\}\end{aligned}$$

where $\hat{t}_{jj}^k = 1, \hat{t}_{jj'}^k = -\hat{\theta}_{jj'}^k$.

$$\begin{aligned}\hat{\boldsymbol{\Theta}} &= \underset{\boldsymbol{\Theta}}{\operatorname{argmin}} \left\{ \sum_{k=K}^q \frac{1}{n_k} \sum_{j=1}^q \left\| \mathbf{Y}_j^k - (\mathbf{Y}_{-j}^k - \mathbf{X}^k \hat{\mathbf{B}}_{-j}^k) \boldsymbol{\theta}_j^k - \mathbf{X}^k \hat{\mathbf{B}}_j^k \right\|^2 + \gamma_n \sum_{j' \neq j, g \in \mathcal{G}_{jj'}} \|\boldsymbol{\theta}_{jj'}^{[g]}\| \right\} \\ &= \underset{\boldsymbol{\Theta}}{\operatorname{argmin}} \left\{ \sum_{k=K}^q \frac{1}{n_k} \sum_{j=1}^q \left\| \hat{\mathbf{E}}_j^k - \hat{\mathbf{E}}_{-j}^k \boldsymbol{\theta}_j^k \right\|^2 + \gamma_n \sum_{j' \neq j, g \in \mathcal{G}_{jj'}} \|\boldsymbol{\theta}_{jj'}^{[g]}\| \right\}\end{aligned}$$

where $\hat{\mathbf{E}}^k = \mathbf{Y}^k - \mathbf{X}^k \hat{\mathbf{B}}^k$.

Non-asymptotic error bounds for $\hat{\beta}$

For $\lambda_n \geq 4\sqrt{|h_{\max}|}\mathbb{R}_0\sqrt{\frac{\log(pq)}{n}}$, the following hold with probability approaching 1 as $n \rightarrow \infty$,

$$\|\hat{\beta} - \beta_0\|_1 \leq \frac{48\sqrt{|h_{\max}|}s_\beta\lambda_n}{\psi^*}$$

$$\|\hat{\beta} - \beta_0\| \leq \frac{12\sqrt{s_\beta}\lambda_n}{\psi^*}$$

$$\sum_{h \in \mathcal{H}} \|\beta^{[h]} - \beta_0^{[h]}\| \leq \frac{48s_\beta\lambda_n}{\psi^*}$$

with ψ^*, \mathbb{R}_0 being constants, and $\beta = (\text{vec}(\mathbf{B}^1)^T, \dots, \text{vec}(\mathbf{B}^K)^T)^T$, $|h_{\max}|$ the maximum group size in β_0 (the true β) and s_β the sparsity of β_0 .

For $\gamma_n = 4\sqrt{|g_{\max}|}\mathbb{Q}_0\sqrt{\frac{\log(pq)}{n}}$, the following hold with probability approaching 1 as $n \rightarrow \infty$,

$$\begin{aligned}\|\hat{\Theta}_j - \Theta_{0,j}\|_F &\leq \frac{12\sqrt{s_j}\gamma_n}{\psi} \\ \sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \|\hat{\theta}_{jj'}^{[g]} - \theta_{0,jj'}^{[g]}\| &\leq \frac{48s_j\gamma_n}{\psi} \\ \left| \text{supp}(\hat{\Theta}_j) \right| &\leq \frac{128s_j}{\psi} \\ \frac{1}{K} \sum_{k=1}^K \|\hat{\Omega}_y^k - \Omega_y^k\|_F &\leq O\left(\frac{\sqrt{S}\gamma_n}{\sqrt{K}}\right)\end{aligned}$$

with ψ, \mathbb{Q}_0 being constants, $|g_{\max}|$ the maximum group size in Θ_0 , s_j the sparsity of Θ_j and $S = \sum_j s_j$.

- 1 Multiple multi-level graphical models
- 2 Model formulation
- 3 Hypothesis testing**
- 4 Simulation studies

- Consider the case $K = 2$, and suppose we are interested in testing if the effect of variable i in the X-data is different across the two populations.
- For this we use the i^{th} rows of the estimates $\hat{\mathbf{B}}^1$ and $\hat{\mathbf{B}}^2$.
- We debias these row vectors using the neighborhood coefficients in the X-network computed previously using JSEM: In this setup, define the desparsified estimate of \mathbf{b}_i^k as

$$\hat{\mathbf{c}}_i^k = \hat{\mathbf{b}}_i^k + \frac{1}{nt_i^k} \left(\mathbf{X}_i^k - \mathbf{X}_{-i}^k \hat{\zeta}_i^k \right)^T (\mathbf{Y}^k - \mathbf{X}^k \hat{\mathbf{B}}^k)$$

for $k = 1, 2$, where $t_i^k := (\mathbf{X}_i^k - \mathbf{X}_{-i}^k \hat{\zeta}_i^k)^T \mathbf{X}_{-i}^k / n$.

Assume we have 'good enough' estimators:

$$\|\hat{\zeta}^k - \zeta_0^k\|_1 = O\left(\sqrt{\frac{\log p}{n}}\right); \quad \|\hat{\mathbf{B}}^k - \mathbf{B}_0^k\|_1 = O\left(\sqrt{\frac{\log(pq)}{n}}\right)$$

$$\left\|(\hat{\Omega}_y^k)^{1/2} - (\Omega_y^k)^{1/2}\right\|_\infty = O\left(\sqrt{\frac{\log q}{n}}\right)$$

Also define

$$\hat{s}_i^k := \sqrt{\|\mathbf{X}_i^k - \mathbf{X}_{-i}^k \hat{\zeta}_i^k\|^2 / n}; \quad m_i^k := \sqrt{nt_i^k} / \hat{s}_i^k$$

Then for sample size satisfying $n \gtrsim \log(pq)$, $\log p = o(n^{1/2})$, $\log q = o(n^{1/2})$ we have

$$\begin{bmatrix} \hat{\Omega}_y^1 & \\ & \hat{\Omega}_y^2 \end{bmatrix}^{1/2} \begin{bmatrix} m_i^1(\hat{\mathbf{c}}_i^1 - \mathbf{b}_i^1) & \\ & m_i^2(\hat{\mathbf{c}}_i^2 - \mathbf{b}_i^2) \end{bmatrix} \sim \mathcal{N}_{2q}(\mathbf{0}, \mathbf{I}) + o_P(1)$$

- 1 Obtain the debiased estimators $\hat{\mathbf{c}}_i^1, \hat{\mathbf{c}}_i^2$;
- 2 Calculate the test statistic

$$D_i = \left\| m_i^1 (\hat{\Omega}_y^1)^{1/2} \hat{\mathbf{c}}_i^1 - m_i^2 (\hat{\Omega}_y^2)^{1/2} \hat{\mathbf{c}}_i^2 \right\|^2$$

- 3 Reject H_0 if $D_i \geq \chi_{2q, 1-\alpha}^2$.

Simultaneous tests for $H_0^j : b_{0ij}^1 = b_{0ij}^2$ at level $\alpha, 0 < \alpha < 1$

- 1 Calculate the pairwise test statistics d_{ij} for $j = 1, \dots, q$:

$$d_{ij} = \frac{\tau_{ij}^1 \hat{\mathbf{c}}_{ij}^1 - \tau_{ij}^2 \hat{\mathbf{c}}_{ij}^2}{1/m_i^1 + 1/m_i^2}$$

where τ_{ij}^k is the $(i, j)^{\text{th}}$ element of $(\hat{\Omega}_y^k)^{1/2}, k = 1, 2$.

- 2 Obtain the threshold

$$\hat{\tau} = \inf \left\{ \tau \in \mathbb{R} : 1 - \Phi(\tau) \leq \frac{\alpha}{2q} \max \left(\sum_{j \in \mathcal{I}_q} \mathbb{I}(|d_{ij}| \geq \tau), 1 \right) \right\}$$

- 3 For $j \in \mathcal{I}_q$, reject H_0^j if $|d_{ij}| \geq \hat{\tau}$.

- 1 Multiple multi-level graphical models
- 2 Model formulation
- 3 Hypothesis testing
- 4 **Simulation studies**

- Number of categories (K) = 5;
- Structured $\{\Omega_x\}, \{\Omega_y\}, \mathcal{B}$;
- Groups in \mathcal{B}, Ω_x are non-zero with probability $5/p$, and their elements come from $\text{Unif}(-1, -0.5) \cup (0.5, 1)$;
- Groups in Ω_y are non-zero with probability $5/q$, and their elements come from $\text{Unif}(-1, -0.5) \cup (0.5, 1)$;
- We generate size- n i.i.d. samples \mathbf{X}^k from $\mathcal{N}_p(0, \Sigma_x^k)$, and \mathbf{E}^k from $\mathcal{N}_p(0, \Sigma_y^k)$, then obtain $\mathbf{Y}^k = \mathbf{X}^k \mathbf{B}^k + \mathbf{E}^k$;
- 100 Replications.

- 1 True positives-

$$\text{TP}(\hat{\mathcal{B}}) = \frac{\sum_k |\text{supp}(\hat{\mathbf{B}}^k) \cup \text{supp}(\mathbf{B}_0^k)|}{\sum_k |\text{supp}(\mathbf{B}_0^k)|}$$

- 2 True negatives-

$$\text{TN}(\hat{\mathcal{B}}) = \frac{\sum_k |\text{supp}^c(\hat{\mathbf{B}}^k) \cup \text{supp}^c(\mathbf{B}_0^k)|}{\sum_k |\text{supp}^c(\mathbf{B}_0^k)|}$$

- 3 Relative error in Frobenius norm-

$$\text{rel.Frob}(\hat{\mathcal{B}}) = \sum_{k=1}^K \frac{\|\hat{\mathbf{B}}^k - \mathbf{B}_0^k\|_F}{\|\mathbf{B}_0^k\|_F}$$

Same metrics are used for $\hat{\Theta}$.

Setting 1: $n = 100, p = 60, q = 30, K = 5$

Method	$TP(\hat{B})$	$TN(\hat{B})$	$rel.Frob(\hat{B})$	$TP(\hat{\Theta})$	$TN(\hat{\Theta})$	$rel.Frob(\hat{\Theta})$
Joint (JMMLE)	0.999 (2e-3)	0.99 (0.01)	0.19 (0.02)	0.66 (0.06)	0.95 (0.01)	0.33 (0.02)
Separate	0.95 (0.02)	0.99 (2e-3)	0.27 (0.03)	0.89 (0.02)	0.63 (0.01)	0.77 (0.04)

Setting 2: $n = 100, p = 30, q = 60, K = 5$

Method	$TP(\hat{B})$	$TN(\hat{B})$	$rel.Frob(\hat{B})$	$TP(\hat{\Theta})$	$TN(\hat{\Theta})$	$rel.Frob(\hat{\Theta})$
Joint (JMMLE)	0.996 (4e-3)	0.99 (6e-3)	0.21 (0.01)	0.58 (0.04)	0.98 (3e-3)	0.32 (8e-3)
Separate	0.66 (0.04)	0.994 (1e-3)	0.59 (0.03)	0.62 (0.03)	0.81 (7e-3)	0.43 (0.01)

Setting 3: $n = 150, p = 200, q = 200, K = 5$

Method	$TP(\hat{B})$	$TN(\hat{B})$	$rel.Frob(\hat{B})$	$TP(\hat{\Theta})$	$TN(\hat{\Theta})$	$rel.Frob(\hat{\Theta})$
Joint (JMMLE)	1.00 (0)	1.00 (0)	0.12 (5e-3)	0.39 (0.04)	0.996 (2e-3)	0.30 (7e-3)
Separate	0.42 (0.03)	0.99 (1e-3)	0.48 (0.02)	0.41 (0.02)	0.73 (0.01)	0.44 (0.02)

- Larger simulation: effect of signal strength and sparsity;
- Real data application, application to personalized medicine;
- Mediation analysis.

- M. Drton and M. D. Perlman. Multiple Testing and Error Control in Gaussian Graphical Model Selection. *Statist. Sci.*, 22(3):430–449, 2007.
- V. Gligorijević and N. Pržulj. Methods for biological data integration: perspectives and challenges. *J. R. Soc. Interface*, 12(112):20150571, 2015.
- J. Guo, E. Levina, G. Michailidis, and J. Zhu. Joint estimation of multiple graphical models. *Biometrika*, 98(1): 1–15, 2011.
- J. Lin, S. Basu, M. Banerjee, and G. Michailidis. Penalized Maximum Likelihood Estimation of Multi-layered Gaussian Graphical Models. *J. Mach. Learn. Res.*, 17:5097–5147, 2016.
- W. Liu. Structural similarity and difference testing on multiple sparse Gaussian graphical models. *Ann. Statist.*, To appear, 2017+.
- J. Ma and G. Michailidis. Joint Structural Estimation of Multiple Graphical Models. *J. Mach. Learn. Res.*, 17: 5777–5824, 2016.
- N. Meinshausen and P. Bühlmann. High dimensional graphs and variable selection with the ℓ_1 lasso. *Ann. Statist.*, 34(3):1436–1462, 2006.

THANK YOU!

Questions?