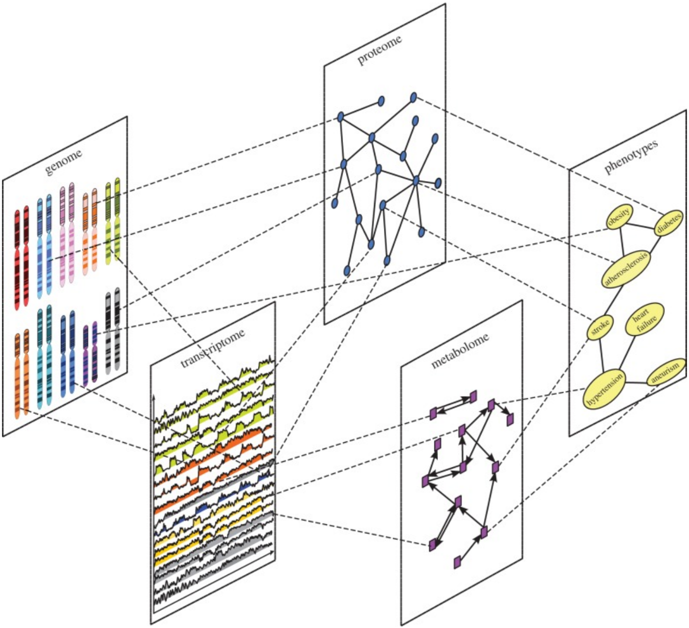
Statistical Inference with Multi-layered Graphical Models

Subhabrata Majumdar
Joint work with George Michailidis
University of Florida Informatics Institute

AT&T Labs - Research, Bedminster, NJ
April 5, 2018

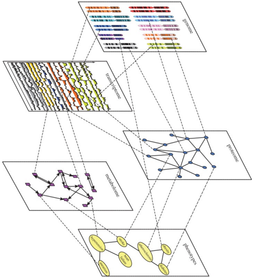
Summary

- Estimation of graphs from high-dimensional data is of importance for biological processes, financial systems or social interactions;
- Nodes in such data can have a natural hierarchical structure, e.g. Genes affecting proteins affecting metabolites, or macroeconomic indicators like interest rates or price indices affecting stock prices;
- There are within layer and between-layer connections in such structures.

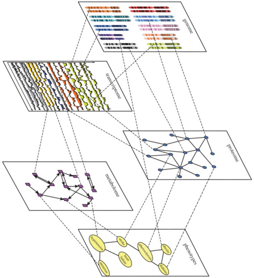


Source: Gligorijević and Pržulj (2015)

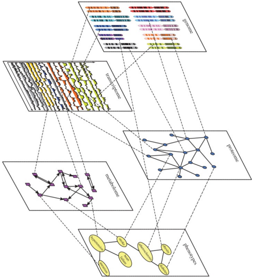
These connections can be different inside different organs, experimental conditions, or for different subtypes of the same disease;



Liver



Kidney



Lungs

Statistical inference for hierarchical graphical models.

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.

Table of contents

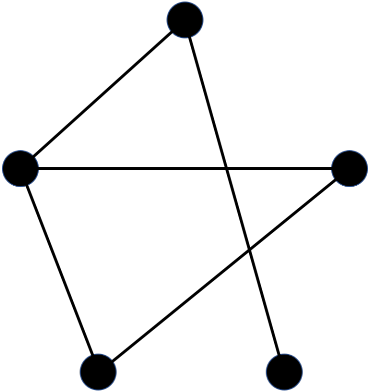
1. Multiple multi-level graphical models
2. Model formulation
3. Joint Multiple Multi-Level Estimation
4. Hypothesis testing in multi-layer models
5. Numerical experiments

Outline

1. Multiple multi-level graphical models
2. Model formulation
3. Joint Multiple Multi-Level Estimation
4. Hypothesis testing in multi-layer models
5. Numerical experiments

Gaussian Graphical models

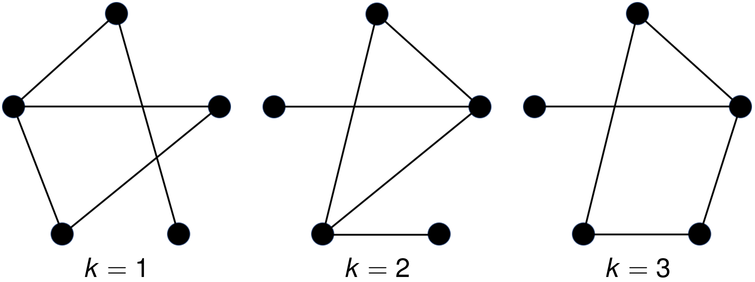
$$\mathbb{X} = (X_1, \ldots, X_p)^T \sim \mathcal{N}_p(0, \Sigma_x); \quad \Omega_x = \Sigma_x^{-1}$$



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

Multiple Gaussian Graphical models

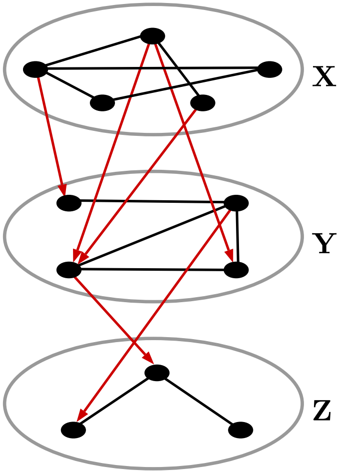
$$\mathbb{X}^k = (X_1^k, \ldots, X_p^k)^T \sim \mathcal{N}_p(0, \Sigma_x^k); \quad \Omega_x^k = (\Sigma_x^k)^{-1}$$
$$k = 1, 2, \ldots, K$$



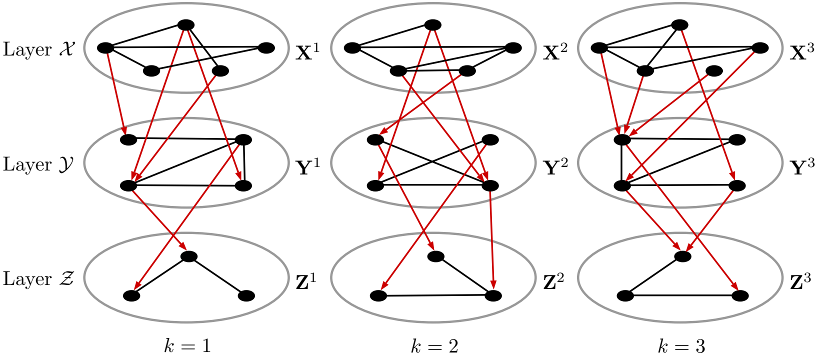
- Joint estimation of $\{\Omega_x^k\}$: Guo et al. (2011); Ma and Michailidis (2016)
- Difference and similarity testing with FDR control: Liu (2017)

$$\mathbb{E} = (E_1, \ldots, E_q)^T \sim \mathcal{N}_q(0, \Sigma_y);$$
$$\mathbb{F} = (F_1, \ldots, F_r)^T \sim \mathcal{N}_r(0, \Sigma_z);$$
$$\Omega_y = (\Sigma_y)^{-1}, \Omega_z = (\Sigma_z)^{-1}$$
$$\mathbb{Y} = \mathbb{X}^T \mathbf{B} + \mathbb{E},$$
$$\mathbb{Z} = \mathbb{Y}^T \mathbf{C} + \mathbb{F}.$$

- $\Omega_x, \Omega_y, \Omega_z$ 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



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

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

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

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

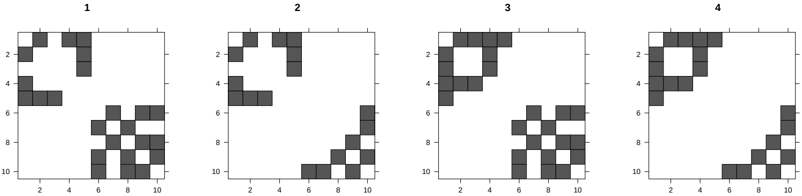
What we do

- We decompose the multi-layer problem into a series of two layer problems.
- We estimate $\{\Omega_x^k, \Omega_y^k, \mathbf{B}^k\}$ jointly for all k from a single model;
  - Incorporate structural informartion 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 $\mathbf{b}_i^k, i \in \{1, 2, \ldots, p\}$ 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 $b_{ij}^1 = b_{ij}^2, j = 1, 2, \ldots, q$ within a row.
- Use simulations for performance evaluation.

1. Multiple multi-level graphical models
2. Model formulation
3. Joint Multiple Multi-Level Estimation
4. Hypothesis testing in multi-layer models
5. Numerical experiments

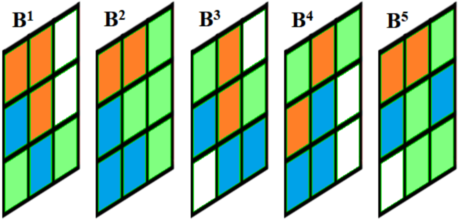
- $\mathcal{Y} = \{\mathbf{Y}^1, \ldots, \mathbf{Y}^K\}, \mathcal{X} = \{\mathbf{X}^1, \ldots, \mathbf{X}^K\}$;
- $\Omega_x = \{\Omega_x^1, \ldots, \Omega_x^K\}, \Omega_y = \{\Omega_y^1, \ldots, \Omega_y^K\}, \mathcal{B} = \{\mathbf{B}^1, \ldots, \mathbf{B}^K\}$;

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 $\{1, \ldots, K\}$ 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 $1 \leq h_i \leq p, 1 \leq h_j \leq q, 1 \leq h_k \leq K$.



Estimation of Ω_x

1. Estimate neighborhood coefficients of each X-node, say $\zeta_i = (\zeta_i^1, \ldots, \zeta_i^K)$ using the group sparsity information $\mathcal{G}_x$:

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

2. Recover precision matrices as MLE over the restricted supports:

$$\widehat{E}_x^k = \{(i,i') : 1 \leq i < i' \leq p, \widehat{\zeta}_{ii'}^k \neq 0 \text{ OR } \widehat{\zeta}_{i'i}^k \neq 0\},$$
$$\widehat{\Omega}_x^k = \underset{\Omega_x^k \in \mathbb{S}_+(\widehat{E}_x^k)}{\text{argmin}} \left\{ \text{Tr}(\widehat{\mathbf{S}}_x^k \Omega_x^k) - \log\det(\Omega_x^k) \right\}.$$

Joint Structural Estimation Method (JSEM)
Ma and Michailidis (2016)

Estimation strategy of $\mathcal{B}, \Omega_y$

1. Estimate regression coefficient matrices using group information $\mathcal{H}$ and neighborhood coefficients of each Y-node, say $\Theta_j = (\boldsymbol{\theta}_j^1, \ldots, \boldsymbol{\theta}_j^K)$,
2. Estimate neighborhood coefficients of each Y-node, i.e. $\widehat{\Theta}_j$ using the group sparsity information $\mathcal{G}_y$ and regression matrices $\mathbf{B}^k$,
3. Estimate Y-layer precision matrices as MLE over the restricted supports.

1. Estimate regression coefficient matrices using group information $\mathcal{H}$ and neighborhood coefficients of each Y-node, say $\Theta_j = (\boldsymbol{\theta}_j^1, \ldots, \boldsymbol{\theta}_j^K)$,
2. Estimate neighborhood coefficients of each Y-node, i.e. $\widehat{\Theta}_j$ using the group sparsity information $\mathcal{G}_y$ and regression coefficient matrices $\mathbf{B}^k$,
3. Estimate Y-layer precision matrices as MLE over the restricted supports.

Solution: Iterate!

A note on testing

Our testing methodology only requires *generic* estimates that satisfy the following convergence rates:

$$\|\widehat{\zeta}^k - \zeta_0^k\|_1 \leq O\left(\sqrt{\frac{\log p}{n}}\right)$$

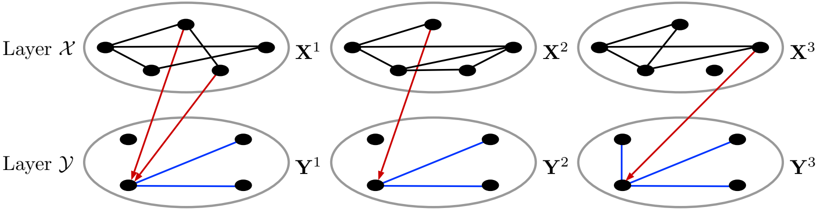
$$\|\widehat{\Omega}_y^k - \Omega_{y0}^k\|_\infty \leq O\left(\sqrt{\frac{\log(pq)}{n}}\right)$$

$$\|\widehat{\mathbf{B}}^k - \mathbf{B}_0^k\|_1 \leq O\left(\sqrt{\frac{\log(pq)}{n}}\right)$$

The estimation procedure gives one suitable set of estimates.

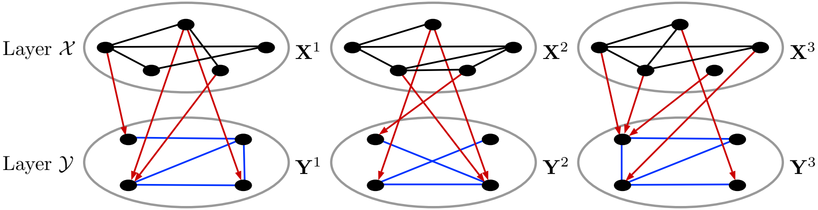
Outline

1. Multiple multi-level graphical models
2. Model formulation
3. **Joint Multiple Multi-Level Estimation**
4. Hypothesis testing in multi-layer models
5. Numerical experiments



$$\mathbf{Y}_j^k \quad - \quad \mathbf{E}_{-j}^k \boldsymbol{\theta}_j^k \quad - \quad \mathbf{X}^k \mathbf{B}_j^k$$

The objective function



$$\sum_{k=1}^{K} \frac{1}{n_k} \sum_{j=1}^{q} \left\| \mathbf{Y}_j^k - (\mathbf{Y}_{-j}^k - \mathbf{X}^k \mathbf{B}_{-j}^k) \boldsymbol{\theta}_j^k - \mathbf{X}^k \mathbf{B}_j^k \right\|^2$$

$$+ \lambda_n \sum_{h \in \mathcal{H}} \|\mathbf{B}^{[h]}\| + \gamma_n \sum_{j' \neq j, g \in \mathcal{G}_{jj'}} \|\boldsymbol{\theta}_{jj'}^{[g]}\|$$

Joint Multiple Multi-Level Estimation (JMMLE)

1. Solve for $\{\mathcal{B}, \Theta\}$:

$$\{\widehat{\mathcal{B}}, \widehat{\Theta}\} = \underset{\mathcal{B},\Theta}{\text{argmin}} \left\{ \sum_{k=1}^{K} \frac{1}{n_k} \sum_{j=1}^{q} \left\| \mathbf{Y}_j^k - (\mathbf{Y}_{-j}^k - \mathbf{X}^k \mathbf{B}_{-j}^k) \boldsymbol{\theta}_j^k - \mathbf{X}^k \mathbf{B}_j^k \right\|^2 \right.$$
$$\left. + \lambda_n \sum_{h \in \mathcal{H}} \| \mathbf{B}^{[h]} \| + \gamma_n \sum_{j' \neq j, g \in \mathcal{G}_{jj'}} \| \boldsymbol{\theta}_{jj'}^{[g]} \| \right\}$$

2. Recover Y-precision matrices as MLE over the restricted supports:

$$\widehat{E}_y^k = \{(j, j') : 1 \leq j < j' \leq q, \widehat{\theta}_{jj'}^k \neq 0 \text{ OR } \widehat{\theta}_{j'j}^k \neq 0\},$$
$$\widehat{\Omega}_y^k = \underset{\Omega_y^k \in \mathbb{S}_+(\widehat{E}_y^k)}{\text{argmin}} \left\{ \text{Tr}(\widehat{\mathbf{S}}_y^k \Omega_y^k) - \log\det(\Omega_y^k) \right\}.$$

Computational algorithm

$$\{\widehat{\mathcal{B}}, \widehat{\Theta}\} = \underset{\mathcal{B},\Theta}{\text{argmin}} \{f(\mathcal{Y}, \mathcal{X}, \mathcal{B}, \Theta) + P(\mathcal{B}) + Q(\Theta)\}$$

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:

$$\mathcal{B}^{(t+1)} = \underset{\mathcal{B}}{\text{argmin}} \left\{ f(\mathcal{Y}, \mathcal{X}, \mathcal{B}, \Theta^{(t)}) + Q(\mathcal{B}) \right\}$$
$$\Theta^{(t+1)} = \underset{\Theta}{\text{argmin}} \left\{ f(\mathcal{Y}, \mathcal{X}, \mathcal{B}^{(t+1)}, \Theta) + P(\Theta) \right\}$$

3. Continue till convergence.

The two subproblems

$$\widehat{\mathcal{B}} = \underset{\mathcal{B}}{\text{argmin}} \left\{ \sum_{k=K}^{q} \frac{1}{n_k} \sum_{j=1}^{q} \left\| \mathbf{Y}_j^k - (\mathbf{Y}_{-j}^k - \mathbf{X}^k \mathbf{B}_{-j}^k) \widehat{\boldsymbol{\theta}}_j^k - \mathbf{X}^k \mathbf{B}_j^k \right\|^2 + \lambda_n \sum_{h \in \mathcal{H}} \|\mathbf{B}^{[h]}\| \right\}$$

$$= \underset{\mathcal{B}}{\text{argmin}} \left\{ \sum_{k=1}^{K} \frac{1}{n_k} \|(\mathbf{Y}^k - \mathbf{X}^k \mathbf{B}^k) \widehat{\mathbf{T}}^k\|_F^2 + + \lambda_n \sum_{h \in \mathcal{H}} \|\mathbf{B}^{[h]}\| \right\}$$

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

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

$$= \underset{\Theta}{\text{argmin}} \left\{ \sum_{k=K}^{q} \frac{1}{n_k} \sum_{j=1}^{q} \|\widehat{\mathbf{E}}_j^k - \widehat{\mathbf{E}}_{-j}^k \boldsymbol{\theta}_j^k\|^2 + \gamma_n \sum_{j' \neq j, g \in \mathcal{G}_{jj'}} \|\boldsymbol{\theta}_{jj'}^{[g]}\| \right\}$$

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

Non-asymptotic error bounds for $\widehat{\mathcal{B}}$

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

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_1 \leq \frac{48\sqrt{|h_{\max}|}s_\beta \lambda_n}{\psi^*}$$

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \leq \frac{12\sqrt{s_\beta}\lambda_n}{\psi^*}$$

$$\sum_{h \in \mathcal{H}} \|\boldsymbol{\beta}^{[h]} - \boldsymbol{\beta}_0^{[h]}\| \leq \frac{48 s_\beta \lambda_n}{\psi^*}$$

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

Non-asymptotic error bounds for $\widehat{\mathcal{B}}$

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

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_1 \leq \frac{48\sqrt{|h_{\max}|}s_\beta \lambda_n}{\psi^*}$$

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \leq \frac{12\sqrt{s_\beta}\lambda_n}{\psi^*}$$

$$\sum_{h \in \mathcal{H}} \|\boldsymbol{\beta}^{[h]} - \boldsymbol{\beta}_0^{[h]}\| \leq \frac{48 s_\beta \lambda_n}{\psi^*}$$

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

Error bounds for $\widehat{\Theta}, \widehat{\Omega}$

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

$$\|\widehat{\Theta}_j - \Theta_{0,j}\|_F \leq \frac{12\sqrt{s_j}\gamma_n}{\psi}$$

$$\sum_{j\neq j', g\in\mathcal{G}_y^{jj'}} \|\widehat{\boldsymbol{\theta}}_{jj'}^{[g]} - \boldsymbol{\theta}_{0,jj'}^{[g]}\| \leq \frac{48 s_j \gamma_n}{\psi}$$

$$\frac{1}{K}\sum_{k=1}^{K} \|\widehat{\Omega}_y^k - \Omega_y^k\|_F \leq O\left(\frac{\sqrt{S}\gamma_n}{\sqrt{K}}\right)$$

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

Error bounds for $\widehat{\Theta}, \widehat{\Omega}$

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

$$\|\widehat{\Theta}_j - \Theta_{0,j}\|_F \leq \frac{12\sqrt{s_j}\gamma_n}{\psi}$$

$$\sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \|\widehat{\boldsymbol{\theta}}_{jj'}^{[g]} - \boldsymbol{\theta}_{0,jj'}^{[g]}\| \leq \frac{48 s_j \gamma_n}{\psi}$$

$$\frac{1}{K}\sum_{k=1}^{K} \|\widehat{\Omega}_y^k - \Omega_y^k\|_F \leq O\left(\frac{\sqrt{S}\gamma_n}{\sqrt{K}}\right)$$

with $\psi, \mathbb{Q}_0$ 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. Joint Multiple Multi-Level Estimation
4. **Hypothesis testing in multi-layer models**
5. Numerical experiments

High-dimensional hypothesis testing

- We propose a debiased estimator for $\mathbf{b}_i^k$ that makes use of already computed model quantities, and establish asymptotic properties of its scaled version,
- We assume $K = 2$, and propose an asymptotic test for detecting differential effects of a variable in the upper layer, i.e. testing for the null hypothesis $H0 : \mathbf{b}_{0i}^1 = \mathbf{b}_{0i}^2$,
- We also propose pairwise simultaneous tests across $j = 1, \ldots, q$ for detecting the elementwise differences $b_{0ij}^1 = b_{0ij}^2$.

Debiasing estimates from a multi-layer model

- We are interested in testing if the effect of variable i in the X-data has any downstream effect. For this we use the i^{th} rows of the estimates $\widehat{\mathbf{B}}^k$.
- We debias these row vectors using the neighborhood coefficients in the X-network computed previously using JSEM:

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

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

Result

Assume we have 'good enough' estimators:

$$\|\widehat{\zeta}^k - \zeta_0^k\|_1 = O\left(\sqrt{\frac{\log p}{n}}\right)$$

$$\|\widehat{\mathbf{B}}^k - \mathbf{B}_0^k\|_1 = O\left(\sqrt{\frac{\log(pq)}{n}}\right)$$

$$\left\|\widehat{\Omega}_y^k - \Omega_y^k\right\|_\infty = O\left(\sqrt{\frac{\log pq}{n}}\right)$$

Define $\widehat{s}_i^k = \sqrt{\|\mathbf{X}_i^k - \mathbf{X}_{-i}^k \widehat{\zeta}_i^k\|^2/n}$, and $m_i^k = \sqrt{n} t_i^k / \widehat{s}_i^k$. and:

$$\widehat{\Omega}_y = \text{diag}(\widehat{\Omega}_y^1, \ldots, \widehat{\Omega}_y^K), \quad \mathbf{M}_i = \text{diag}(m_i^1, \ldots, m_i^K)$$

$$\widehat{\mathbf{C}}_i = \text{vec}(\widehat{\mathbf{c}}_i^1, \ldots, \widehat{\mathbf{c}}_i^K)^T, \quad \mathbf{D}_i = \text{vec}(\mathbf{b}_{0i}^1, \ldots, \mathbf{b}_{0i}^K)^T$$

Then under mild conditions, for sample size satisfying $\log p = o(n^{1/2}), \log q = o(n^{1/2})$ we have $\widehat{\Omega}_y^{1/2} \mathbf{M}_i (\widehat{\mathbf{C}}_i - \mathbf{D}_i) \sim \mathcal{N}_{Kq}(\mathbf{0}, \mathbf{I}) + o_P(1)$.

Result

Assume we have 'good enough' estimators:

$$\boxed{\begin{aligned} \|\widehat{\zeta}^k - \zeta_0^k\|_1 &= O\left(\sqrt{\frac{\log p}{n}}\right) \\ \|\widehat{\mathbf{B}}^k - \mathbf{B}_0^k\|_1 &= O\left(\sqrt{\frac{\log(pq)}{n}}\right) \\ \left\|\widehat{\Omega}_y^k - \Omega_y^k\right\|_\infty &= O\left(\sqrt{\frac{\log pq}{n}}\right) \end{aligned}}$$

Define $\widehat{s}_i^k = \sqrt{\|\mathbf{X}_i^k - \mathbf{X}_{-i}^k \widehat{\zeta}_i^k\|^2 / n}$, and $m_i^k = \sqrt{n} t_i^k / \widehat{s}_i^k$. and:

$$\widehat{\Omega}_y = \text{diag}(\widehat{\Omega}_y^1, \ldots, \widehat{\Omega}_y^K), \quad \mathbf{M}_i = \text{diag}(m_i^1, \ldots, m_i^K)$$
$$\widehat{\mathbf{C}}_i = \text{vec}(\widehat{\mathbf{c}}_i^1, \ldots, \widehat{\mathbf{c}}_i^K)^T, \quad \mathbf{D}_i = \text{vec}(\mathbf{b}_{0i}^1, \ldots, \mathbf{b}_{0i}^K)^T$$

Then under mild conditions, for sample size satisfying $\log p = o(n^{1/2}), \log q = o(n^{1/2})$ we have $\widehat{\Omega}_y^{1/2} \mathbf{M}_i (\widehat{\mathbf{C}}_i - \mathbf{D}_i) \sim \mathcal{N}_{Kq}(\mathbf{0}, \mathbf{I}) + o_P(1)$.

Global test for $H_0 : \mathbf{b}_{0i}^1 = \mathbf{b}_{0i}^2$ at level $\alpha, 0 < \alpha < 1$

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

$$D_i = (\widehat{\mathbf{c}}_i^1 - \widehat{\mathbf{c}}_i^2)^T \left(\frac{\widehat{\Sigma}_y^1}{(m_i^1)^2} + \frac{\widehat{\Sigma}_y^2}{(m_i^2)^2} \right)^{-1} (\widehat{\mathbf{c}}_i^1 - \widehat{\mathbf{c}}_i^2)$$

   where $\widehat{\Sigma}_y^k = (\widehat{\Omega}_y^k)^{-1}, k = 1, 2$.
3. Reject H_0^i if $D_i \geq \chi^2_{q,1-\alpha}$.

Besides controlling the type-I error at a specified level, the above testing procedure maintains rate optimal power, i.e. asymptotic power goes to 1 when $\|\mathbf{b}_{0i}^1 - \mathbf{b}_{0i}^2\| \geq O(n^{-1/2})$.

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

1. Calculate the pairwise test statistics for $j = 1, \ldots, q$:

$$d_{ij} = \frac{\widehat{c}_{ij}^1 - \widehat{c}_{ij}^2}{\sqrt{\hat{\sigma}_{jj}^1/(m_i^1)^2 + \hat{\sigma}_{jj}^2/(m_i^2)^2}}$$

2. Obtain the threshold

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

3. For $j = 1, \ldots, q$, reject H_0^{ij} if $|d_{ij}| \geq \hat{\tau}$.

Based on the FDR control procedure, we can perform *within-row thresholding* in the matrices $\widehat{\mathbf{B}}^k$ to tackle group misspecification.

$$\hat{\tau}_i^k := \inf \left\{ \tau \in \mathbb{R} : 1 - \Phi(\tau) \leq \frac{\beta}{2q} \max \left(\sum_{j \in \mathcal{I}_q} \mathbb{I}(|\sqrt{\hat{\omega}_{jj}^k} m_i^k \hat{c}_{ij}^k| \geq \tau), 1 \right) \right\}$$

$$\hat{b}_{ij}^{k,\text{thr}} = \hat{b}_{ij}^k \mathbb{I} \left(|\sqrt{\hat{\omega}_{jj}^k} m_i^k \hat{c}_{ij}^k| \geq \hat{\tau}_i^k \right)$$

Even without group misspecification, this helps identify directed edges between layers that have high nonzero values.

Outline

1. Multiple multi-level graphical models
2. Model formulation
3. Joint Multiple Multi-Level Estimation
4. Hypothesis testing in multi-layer models
5. Numerical experiments

Simulation setup

- Number of categories (K) = 5;
- Structured $\{\Omega_x\}, \{\Omega_y\}, \mathcal{B}$;
- Groups in $\mathcal{B}, \Omega_x$ are non-zero with probability $5/p$, and their elements come from Unif$[-1, -0.5] \cup [0.5, 1]$;
- Groups in Ω_y are non-zero with probability $5/q$, and their elements come from 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 $\mathbf{Y}^k = \mathbf{X}^k \mathbf{B}^k + \mathbf{E}^k$;
- 50 Replications.
- Tuning parameters:

$$\gamma_n \in \{0.3, 0.4, ..., 1\} \sqrt{\frac{\log q}{n}}, \lambda_n \in \{0.4, 0.6, ..., 1.8\} \sqrt{\frac{\log p}{n}}$$

Evaluation metrics

1. True positive Rate-

$$\text{TPR}(\widehat{\mathcal{B}}) = \frac{1}{K}\sum_{k=1}^{K} \frac{|\,\text{supp}(\widehat{\mathbf{B}}^k) \cup \text{supp}(\mathbf{B}_0^k)|}{|\,\text{supp}(\mathbf{B}_0^k)|}$$

2. True negatives-

$$\text{TNR}(\widehat{\mathcal{B}}) = \frac{1}{K}\sum_{k=1}^{K} \frac{|\,\text{supp}^c(\widehat{\mathbf{B}}^k) \cup \text{supp}^c(\mathbf{B}_0^k)|}{|\,\text{supp}^c(\mathbf{B}_0^k)|}$$

3. Relative error in Frobenius norm-

$$\text{RF}(\widehat{\mathcal{B}}) = \frac{1}{K}\sum_{k=1}^{K} \frac{\|\widehat{\mathbf{B}}^k - \mathbf{B}_0^k\|_F}{\|\mathbf{B}_0^k\|_F}$$

4. Matthews correlation coefficient (MCC).

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

(π_x, π_y)	(p, q, n)	Method	TPR	TNR	MCC	RF
$(5/p, 5/q)$	(60,30,100)	JMMLE	**0.97(0.02)**	0.99(0.003)	**0.96(0.014)**	0.24(0.033)
		Separate	0.96(0.018)	0.99(0.004)	0.93(0.014)	0.22(0.029)
	(30,60,100)	JMMLE	0.97(0.013)	0.99(0.002)	**0.96(0.008)**	0.27(0.024)
		Separate	0.99(0.009)	0.99(0.003)	0.93(0.017)	0.18(0.021)
	(200,200,150)	JMMLE	0.98(0.011)	**1.0(0)**	**0.99(0.005)**	**0.16(0.025)**
		Separate	0.99(0.001)	0.99 (0.001)	0.88(0.009)	0.18(0.007)
	(300,300,150)	JMMLE	**1.0(0.001)**	**1.0(0)**	**0.99(0.001)**	**0.14 (0.015)**
		Separate	1.0(0.001)	0.99(0.001)	0.84(0.01)	0.21(0.007)
$(30/p, 30/q)$	(200,200,100)	JMMLE	**0.97(0.017)**	**1.0(0)**	**0.98(0.008)**	**0.21(0.032)**
		Separate	0.32(0.01)	0.99(0.001)	0.49(0.009)	0.85(0.06)
	(200,200,200)	JMMLE	**0.99(0.006)**	**1.0(0)**	**0.99(0.007)**	**0.13(0.016)**
		Separate	0.97(0.004)	0.98(0.001)	0.93(0.002)	0.19(0.07)

Table of outputs for estimation of regression matrices, giving empirical mean and standard deviation (in brackets) of each evaluation metric over 50 replications.

(π_x, π_y)	(p, q, n)	Method	TPR	TNR	MCC	RF
$(5/p, 5/q)$	(60,30,100)	JMMLE	0.76(0.018)	0.90(0.006)	**0.61(0.024)**	**0.32(0.008)**
		Separate	0.77(0.031)	0.92(0.007)	0.56(0.03)	0.51(0.017)
		JSEM	0.24(0.013)	0.8(0.003)	0.05(0.015)	1.03(0.002)
	(30,60,100)	JMMLE	0.7(0.018)	**0.94(0.002)**	0.55(0.018)	**0.3(0.005)**
		Separate	0.76(0.041)	0.89(0.015)	0.59(0.039)	0.49(0.014)
		JSEM	0.13(0.005)	0.9(0.001)	0.03(0.007)	1.04(0.001)
	(200,200,150)	JMMLE	0.68(0.017)	**0.98(0)**	0.48(0.013)	**0.26(0.002)**
		Separate	0.78(0.019)	0.97(0.001)	0.55(0.012)	0.6(0.007)
		JSEM	0.05(0.002)	0.97(0)	0.02(0.002)	1.01(0)
	(300,300,150)	JMMLE	**0.71(0.014)**	**0.98(0)**	0.44(0.008)	**0.25(0.002)**
		Separate	0.71(0.017)	0.98(0.001)	0.51(0.011)	0.59(0.005)
		JSEM	0.04(0.002)	0.98(0)	0.02(0.002)	1.01(0)
$(30/p, 30/q)$	(200,200,100)	JMMLE	**0.77(0.016)**	**0.98(0)**	**0.46(0.013)**	**0.31(0.003)**
		Separate	0.57(0.027)	0.44(0.007)	0.04(0.008)	0.84(0.002)
		JSEM	0.05(0.002)	0.97(0)	0.01(0.002)	1.01(0)
	(200,200,200)	JMMLE	**0.76(0.018)**	**0.98(0)**	**0.55(0.015)**	**0.27(0.004)**
		Separate	0.73(0.023)	0.94(0.003)	0.39(0.017)	0.62(0.011)
		JSEM	0.05(0.002)	0.97(0)	0.03(0.003)	1.01(0)

Table of outputs for estimation of lower layer precision matrices over 50 replications.

- Set $K = 2$, then randomly assign each element of $\mathbf{B}_0^1$ as non-zero w.p. π, then draw their values from Unif$\{[-1, -0.5] \cup [0.5, 1]\}$ independently.
- Generate a matrix of differences $\mathbf{D}$, where $(\mathbf{D})_{ij}$ takes values –1, 1, 0 w.p. 0.1, 0.1 and 0.8, respectively. Finally set $\mathbf{B}_0^2 = \mathbf{B}_0^1 + \mathbf{D}$.
- Identical sparsity structures for the pairs of X- and Y-precision matrices.
- Type-I error set at 0.05, FDR controlled at 0.2.
- Empirical sizes of global tests are calculated from estimators obtained from a separate set of data generated by setting all elements of $\mathbf{D}$ to 0.

Results

(π_x, π_y)	(p, q)	n	Global test		Simultaneous tests	
			Power	Size	Power	FDR
$(5/p, 5/q)$	(60,30)	100	0.977 (0.018)	0.058 (0.035)	0.937 (0.021)	0.237 (0.028)
		200	0.987 (0.016)	0.046 (0.032)	0.968 (0.013)	0.218 (0.032)
	(30,60)	100	0.985 (0.018)	0.097 (0.069)	0.925 (0.022)	0.24 (0.034)
		200	0.990 (0.02)	0.119 (0.059)	0.958 (0.024)	0.245 (0.041)
	(200,200)	150	0.987 (0.005)	0.004 (0.004)	0.841 (0.13)	0.213 (0.007)
	(300,300)	150	0.988 (0.002)	0.002 (0.003)	0.546 (0.035)	0.347 (0.017)
		300	0.998 (0.003)	0.000 (0.001)	0.989 (0.003)	0.117 (0.006)
$(30/p, 30/q)$	(200,200)	100	0.994 (0.005)	0.262 (0.06)	0.479 (0.01)	0.557 (0.006)
		200	0.998 (0.004)	0.020 (0.01)	0.962 (0.003)	0.266 (0.007)
		300	0.999 (0.002)	0.011 (0.008)	0.990 (0.004)	0.185 (0.009)

Table of outputs for hypothesis testing.

Future work

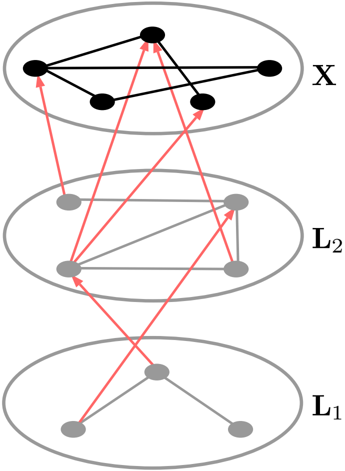
- Beyond pairwise testing: global and simultaneous tests for $K > 2$;
- Multi-level estimation and testing for model assumptions other than structured sparsity;
- Non-gaussian data;
- **Graphical models with non-linear interactions**.

Graphical models with non-linear interactions

- Take the multi-layer structure as a generative model.
- Only the top layer is observed, other layers are composed of latent data.

$$
\begin{aligned}
\mathbb{L}_1 &= (L_{11}, \ldots, L_{1r})^T \sim \mathcal{N}_r(0, \Sigma_1);\\
\mathbb{L}_2 &= \phi(\mathbb{L}_1^T \mathbf{B}) + \mathbb{E},\\
\mathbb{X} &= \phi(\mathbb{L}_2^T \mathbf{C}) + \mathbb{F},\\
\mathbb{E} &= (E_1, \ldots, E_q)^T \sim \mathcal{N}_q(0, \Sigma_2);\\
\mathbb{F} &= (F_1, \ldots, F_p)^T \sim \mathcal{N}_p(0, \Sigma_x).
\end{aligned}
$$

where ϕ is a known activation function.

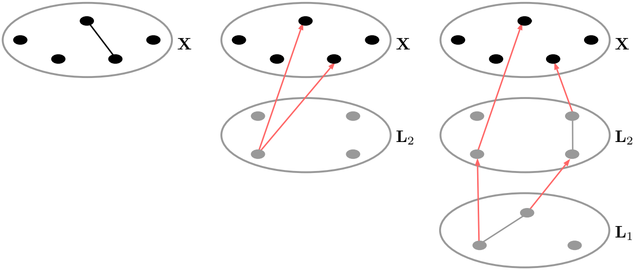


References in literature

- Non-linear generalization of a factor model.
- A general version: $\mathbb{L}_2 = f_1(\mathbb{L}_1) + \mathbb{E}$ etc. for unknown function f_1, has been proposed as Deep Latent Gaussian Model (Rezende et al., 2014).
- The choice $\phi(\mathbb{L}^T\mathbf{B}) \equiv \phi(\mathbb{L})^T\mathbf{B}$ corresponds to Non-linear Gaussian belief networks (Frey and Hinton, 1999).

Our plan

Incorporate sparse estimation of the model parameters to model non-linear interactions.



- Monte-Carlo Sequential EM, backpropagation
- Theoretical properties of estimates

References

M. Drton and M. D. Perlman. Multiple Testing and Error Control in Gaussian Graphical Model Selection. *Statist. Sci.*, 22(3):430–449, 2007.

B. J. Frey and G. E. Hinton. Variational learning in nonlinear Gaussian belief networks. *Neural Comput.*, 11(1): 193–213, 1999.

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.*, 45(6):2680–2707, 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 Â¨ lasso. *Ann. Statist.*, 34(3):1436–1462, 2006.

D. J. Rezende, S. Mohamed, and D. Wierstra. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. In *ICML Proceedings*, volume 32, pages 1278–1286, 2014.

THANK YOU!

Questions?