

# Simultaneous Clustering and Estimation of Multiple Graphical Models

Yufeng Liu

University of North Carolina at Chapel Hill

<http://www.unc.edu/~yfliu>

Joint work with

Wonyul Lee (FDA); Yuying Xie (Michigan State)

Will Sun (U. Miami), Botao Hao and Guang Cheng (Purdue)

# Introduction: Graphical modeling framework

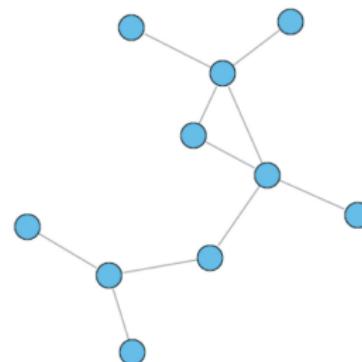
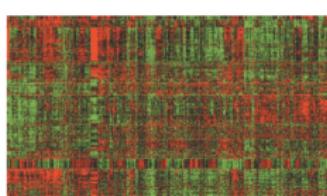
## Undirected graphical models

- ▶ Summarizes conditional independence among variables
- ▶ Each variable represented by a node
- ▶ No edge between two nodes if variables are independent given all others
- ▶ Undirected edges since this relationship is symmetric

# Introduction: Graphical modeling framework

## Gene expression network model

- ▶ Each gene represented by node in graph
- ▶ Genes share an edge when they are conditionally dependent



# Gaussian Graphical Models

- ▶  $\mathbf{X} = (X_1, \dots, X_p)' \sim \mathbf{N}(\mu, \Sigma)$ .
- ▶  $\Sigma^{-1} = \Omega = (\omega_{i,j})_{p \times p}$ .
- ▶  $\omega_{i,j} = 0$  iff  $X_i$  and  $X_j$  are conditionally independent.

$$\text{Corr}(x_i, x_j | \text{other variables}) = -\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}, i \neq j$$

- ▶ Estimating  $\Omega \Leftrightarrow$  Estimating a Gaussian graphical model.

# Gaussian Graphical Models

$$\Omega = \begin{bmatrix} 3 & 0 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ 3 & 2 & 0 & 0 & 0 \\ 4 & 1 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 \end{bmatrix}$$

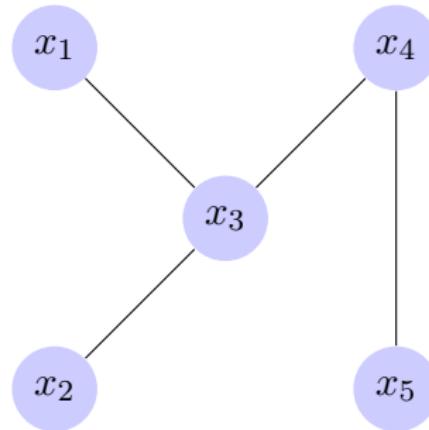


Figure: The undirected graph of  $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$ .

# Gaussian Graphical Models

- ▶ Connect with Lasso:  $x_i = \sum_{j \neq i} \beta_{ij} x_j + \epsilon_i$ ;  $\beta_{ij} = -(\omega_{ij}/\omega_{ii})$  (Meinshausen and Bühlmann, 2006; Peng et al., 2009)
- ▶  $L_1$  penalized maximum log-likelihood (Yuan and Lin, 2007, Banerjee et al., 2007, d'Aspremont et al., 2008, Friedman et al., 2008, Rothman et al. 2008, Lam and Fan 2009, etc.)

$$\hat{\Omega} = \underset{\Omega \succ 0}{\operatorname{argmin}} \operatorname{tr}(\hat{\Sigma}\Omega) - \log \det(\Omega) + \lambda \sum_{i \neq j} |\omega_{i,j}|$$

where  $\hat{\Sigma}$  is sample covariance matrix.

- ▶ Sparse estimate of  $\Omega$ .

## Direct estimation of $\Omega$

- ▶  $L_1$  minimization for inverse matrix estimation (CLIME) (Cai et al., 2011)
- ▶ Optimization using linear programming without specifying the likelihood.
- ▶ The CLIME estimator of  $\Omega$  solves

$$\min \|\Omega\|_1 \text{ subject to: } |\hat{\Sigma}\Omega - I|_\infty \leq \lambda. \quad (1)$$

# Glioblastoma multiforme (GBM) Cancer Data

- ▶ Four subtypes: Classical, Mesenchymal, Neural, and Proneural
- ▶ **Homogeneity:** All about Glioblastoma cancer.
- ▶ **Heterogeneity:** Four different subtypes.
- ▶ **Labels:** Given by separate clustering

# Joint Estimation of Multiple Graphical Models

- ▶ Multiple independent data sets share some common structure.
- ▶ Gene networks describing different sources.

$$\operatorname{argmin}_{\boldsymbol{\Omega}^{(k)}} \sum_{k=1}^K [\text{tr}(\hat{\boldsymbol{\Sigma}}^{(k)} \boldsymbol{\Omega}^{(k)}) - \log \det(\boldsymbol{\Omega}^{(k)})] + P(\boldsymbol{\Omega}^{(1)}, \dots, \boldsymbol{\Omega}^{(K)})$$

- ▶ Encourage common structure through joint regularization  
(Varoquaux et al, 2007; Guo et al., 2011; Chiquet et al., 2011;  
Danaher et al., 2012; Saegusa and Shojaie (2015), etc )

▶ Guo et al.: Encourage both group sparsity and within group  
sparsity  $P(\boldsymbol{\Omega}^{(1)}, \dots, \boldsymbol{\Omega}^{(K)}) = \lambda \sum_{i \neq j} \sum_{\ell=1}^k \sqrt{|\omega_{i,j}^{(\ell)}|}$ .

▶ Danaher et al.: Graphical group lasso penalty

$$\lambda_1 \sum_{\ell=1}^k \sum_{i \neq j} |\omega_{i,j}^{(\ell)}| + \lambda_2 \sum_{\ell=1}^k \sqrt{\sum_{i \neq j} \omega_{i,j}^{(\ell)2}}$$

Fused graphical lasso penalty

$$\lambda_1 \sum_{\ell=1}^k \sum_{i \neq j} |\omega_{i,j}^{(\ell)}| + \lambda_2 \sum_{\ell < \ell'} \sum_{i,j} |\omega_{i,j}^{(\ell)} - \omega_{i,j}^{(\ell')}|$$

## Direct estimation of $\Omega^{(1)}, \dots, \Omega^{(K)}$

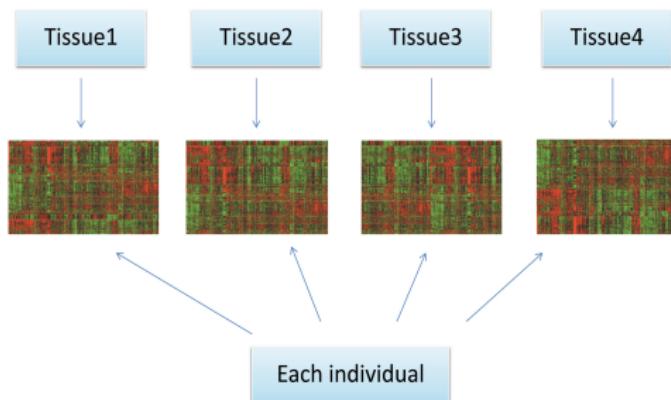
- ▶ Extend non-likelihood method CLIME for multiple graph estimation
- ▶ Take the multi-task learning perspective to better identify the common structure (Lee and Liu, JMLR 2015)
- ▶ Decompose parameters into common structure  $M$  and individual structures  $R^{(k)}$

$$M := \frac{1}{K} \sum_{k=1}^K \Omega^{(k)}, R^{(k)} := \Omega^{(k)} - M; k = 1, \dots, K.$$

- ▶ Encourage sparsity among  $\{R^{(1)}, \dots, R^{(K)}\}$  and for  $M$ .

# Joint Estimation of Dependent Graphical Models

- ▶ Multiple datasets are dependent.
- ▶ Time varying graphs (Zhou et al.; Kolar et al., 2010)
- ▶ Datasets are naturally dependent: e.g. from multiple tissues for a group of mice.
- ▶ Estimating two layers of graphs  
(Xie, Liu, Valdar; Biometrika, 2016)
  1. Category specific layer: tissue-specific graphs
  2. Systemic layer: whole-body systemic graph



# Dependent Graphical Models: Problem Formulation

Xie, Liu, Valdar (2014) considered two layers of graphs:

1. Category specific layer: tissue-specific graphs
  2. Systemic layer: whole-body systemic graph
- ▶ Let  $\mathbf{y}_{k,i} = (y_{k,i1}, \dots, y_{k,ip})^T$  be the  $i$ -th observed data vector for the  $k$ -th category.
  - ▶ For given  $i$ -th mouse,  $\mathbf{y}_{k,i}$  are not independent among different tissues  $k$ .

# Dependent Graphical Models: Problem Formulation

- ▶ We model

$$\mathbf{y}_{k,i} = \mathbf{x}_{k,i} + \mathbf{z}_i, \quad i = 1, \dots, n \quad k = 1, \dots, K$$

where  $\mathbf{z}_i$  is the shared **systemic** random effect, and  $\mathbf{x}_{k,i}$  is the random effect for  $k$ -th category.

- ▶  $\mathbf{x}_{k,i} \stackrel{\text{i.i.d.}}{\sim} \mathbf{N}(0, \Sigma_k)$ ,  $\mathbf{z}_i \stackrel{\text{i.i.d.}}{\sim} \mathbf{N}(0, \Sigma_0)$  and  $\mathbf{x}_{k,i} \perp\!\!\!\perp \mathbf{z}_i$ .
- ▶ Only  $\mathbf{y}_{k,i}$ 's are available.
- ▶ Although  $\mathbf{x}_k$  and  $\mathbf{z}$  are latent variables,

$\Omega_k = \Sigma_k^{-1}; k = 1, \dots, K$  are identifiable with  $K \geq 2$ .

- ▶ Terminologies

systemic network:  $\Omega_0 = \Sigma_0^{-1}$

category-specific network:  $\Omega_k = \Sigma_k^{-1}$

aggregate network:  $\Omega_{Y_k} = (\Omega_k^{-1} + \Omega_0^{-1})^{-1}$

# Sparsity Notion for Dependent Graphical Models

- ▶ Sparse systemic network  $\Omega_0$ : whole-body systemic graph characterizes the body wide dependence structure among genes.
- ▶ Sparse category-specific network  $\Omega_k$ : tissue-specific graphs characterize the dependence structure among genes within each tissue, after removing the common systemic variation.
- ▶ Aggregate network  $\Omega_{Y_k} = (\Omega_k^{-1} + \Omega_0^{-1})^{-1}$  may not be sparse, given sparse  $\Omega_0, \Omega_k; k = 1, \dots, K$ .
- ▶ Goal: Estimate sparse  $\Omega_0, \Omega_k; k = 1, \dots, K$ .

## Application to Mouse Gene Expression Data

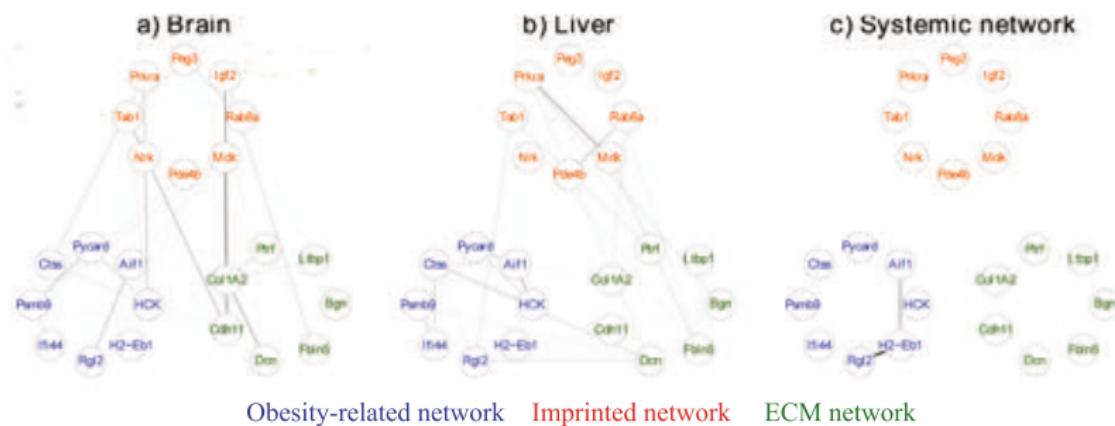
- ▶ 301 mice from  $F_2$  cross, varying genes for fat composition (Dobrin et al. 2009).
- ▶ Acts like randomized allocation of fat-inducing treatment.
- ▶ Gene expression for fat, brain, liver, and muscle.
- ▶ For each tissue, over 20,000 gene expression values.
- ▶ Three groups of gene networks: the obesity-related network, the imprinting related-network, and the extracellular matrix (ECM)-related network.

# Application to Mouse Gene Expression Data

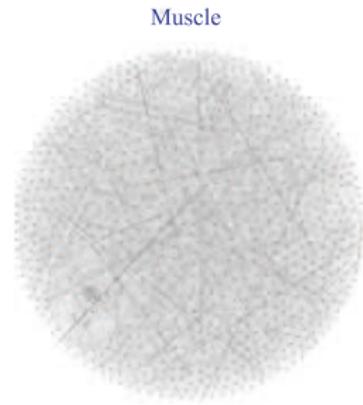
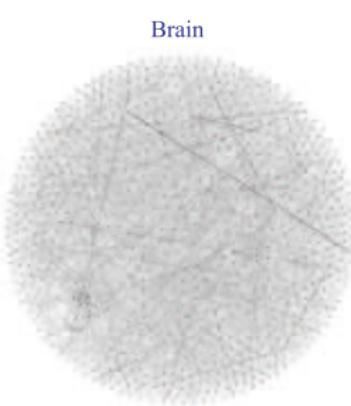
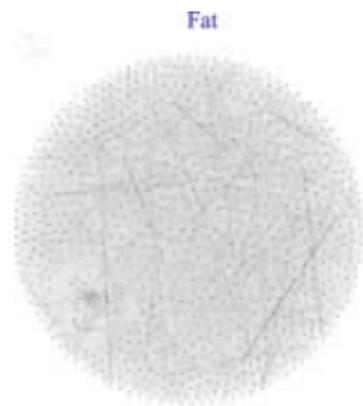
Systemic network should has links that:

1. affect the whole individual
2. reflect the treatment allocation.

Systemic network only has edges for obesity-related network.



## Application to Mouse Gene Expression Data: $P = 1000$



## Application to Mouse Gene Expression Data: P = 1000

- ▶ Systemic network is sparse (249 edges among 62 genes)
- ▶ These 62 genes include some obesity-related genes
- ▶ Analysis of gene ontology (GO) enrichment on the systemic network: the network is significantly enriched for genes associated with immune and metabolic processes, consistent with recent studies linking obesity to strong negative impacts on immune response to infection
- ▶ Tissue-specific networks are much denser

## Open Problems

- ▶ Existing literature focuses on joint estimation from **labeled** data sets.
- ▶ Labels are often obtained by separate clustering algorithms.
- ▶ Joint graph estimation approach can fail when cluster labels are **incorrect**.
- ▶ Known clustering labels: can use **convex** optimization.  
Unknown clustering labels: involves **non-convex** optimization.

## Our Contribution

- ▶ Methodology: a new high dimensional EM algorithm for simultaneous clustering and multiple graphical model estimation.

For each iteration,

- ▶ E-step plays the role of clustering,
  - ▶ M-step conducts the high-dimensional joint estimation.
- 
- ▶ Theory: Analysis of the estimator arising in each iteration step, which leads to an interesting *trade-off* between **statistical error** and **optimization error**.

# Simultaneous Clustering and Estimation

- ▶ New methodology
- ▶ EM algorithm
- ▶ Theoretical results
- ▶ Numerical examples and GBM applications

# Multiple Graphical Models

- ▶ Consider  $K$  clusters  $\mathcal{A}_1, \dots, \mathcal{A}_K$  with cluster assignment matrix  $\mathbf{L} \in \mathbb{R}^{n \times K}$ . Observations  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^p$  are drawn from Gaussian mixture model, which can be specified as mixed Gaussian density by the form

$$f(\mathbf{x}) = \sum_{k=1}^K \pi_k f_k(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), k = 1, \dots, K,$$

where  $\pi_k$  is  $k$ -th mixture weight,  $\boldsymbol{\mu}_k$  is  $k$ -th cluster mean and  $\boldsymbol{\Sigma}_k$  is  $k$ -the covariance matrix.

- ▶ Let  $\boldsymbol{\Theta} := \{\boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_K\}$ , where  $\boldsymbol{\Theta}_k = (\boldsymbol{\mu}_k, \boldsymbol{\Omega}_k)$ .
- ▶ Goal: **Estimation  $K$  precision matrices**  $\boldsymbol{\Omega}_1, \dots, \boldsymbol{\Omega}_K$ .

## If the cluster assignment matrix is known...

- ▶ The log-likelihood function is referred to *complete data*:

$$\log \mathcal{L}(\boldsymbol{\Theta} | \mathbf{X}, \mathbf{L}) := \sum_{i=1}^n \sum_{k=1}^K L_{ik} [\log \pi_k + \log f_k(\mathbf{x}_i; \boldsymbol{\Theta}_k)].$$

- ▶ Joint estimation is summarized as the following optimization problem:

$$\operatorname{argmax}_{\boldsymbol{\Omega}_1, \dots, \boldsymbol{\Omega}_K \succ 0} \log \mathcal{L}(\boldsymbol{\Theta} | \mathbf{X}, \mathbf{L}) - \mathcal{P}(\boldsymbol{\Theta}).$$

- ▶  $\mathcal{P}(\boldsymbol{\Theta})$  encourages common structure across different clusters.
- ▶ If we use convex penalty, the whole problem is a **convex optimization problem**.

## If the cluster assignment matrix is unknown...

- ▶ Simultaneous Clustering And Estimation (SCAN).
- ▶ The log-likelihood function for the *observed data* can be specified by

$$\log \mathcal{L}(\boldsymbol{\Theta} | \mathbf{X}) := \sum_{i=1}^n \log \left( \sum_{k=1}^K \pi_k f_k(\mathbf{x}_i; \boldsymbol{\mu}_k, (\boldsymbol{\Omega}_k)^{-1}) \right).$$

- ▶ The **non-convex** optimization problem is formulated as

$$\max_{\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Omega}_k} \log \mathcal{L}(\boldsymbol{\Theta} | \mathbf{X}) - \mathcal{P}(\boldsymbol{\Theta}).$$

- ▶ SCAN Penalty

$$\mathcal{P}(\boldsymbol{\Theta}) = \underbrace{\lambda_1 \sum_{k=1}^K \sum_{j=1}^p |\mu_{kj}|}_{\text{feature selection}} + \underbrace{\lambda_2 \sum_{k=1}^K \sum_{i \neq j} |\omega_{kij}|}_{\text{sparse estimate}} + \underbrace{\lambda_3 \sum_{i \neq j} \left( \sum_{k=1}^K \omega_{kij}^2 \right)^{1/2}}_{\text{common characteristic}}.$$

# Expectation-Maximization Algorithm

- ▶ SCAN is based on Gaussian mixture model: a **non-convex** problem.
- ▶ E-step is a clustering step, and M-step is a joint estimation step.
- ▶ The interaction between E-step and M-step makes the cluster structure more and more refined.
- ▶ Much more general than existing methods (Balakrishnan et al., 2016 studied low dimensional EM)

# Outline of Our Algorithm

---

**Input:** Training data  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , number of clusters  $K$ , tuning parameter  $\lambda_1, \lambda_2, \lambda_3$ .

**Output:** Cluster assignment  $L_{ik}$ , cluster means  $\mu_k$  and graph  $\Omega_k$ .

---

**Step 1:** Randomly initialize cluster centers  $\mu_k^{(0)}$ , precision matrices  $\Omega_k^{(0)}$  and set  $\pi_k^{(0)} = \frac{1}{K}$ .

**Step 2:** Until the termination condition is met, for  $t = 1, 2, \dots$

(a) E-step. Find the cluster assignment  $L_{\Theta^{(t-1)}, k}(\mathbf{x}_i)$

(b) M-step. Given  $L_{\Theta^{(t-1)}, k}(\mathbf{x}_i)$ , update mixture weight  $\pi_k^{(t)}$ , cluster mean  $\mu_k^{(t)}$ , and the precision matrix  $\Omega_k^{(t)}$  accordingly.

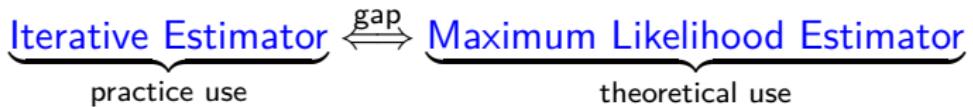
---

## Limitation of Classical Result

- ▶ Classical convergence result of EM algorithm by Wu (1983):
  - ▶ **Unimodal** → Global optimum
  - ▶ **Muti-modal** → Local optimum
- ▶ We are interested in *Maximum Likelihood Estimate* (MLE), which shows good statistical performance.

## Limitation of Classical Result

- ▶ Classical convergence result of EM algorithm by Wu (1983):
  - ▶ **Unimodal** → Global optimum
  - ▶ **Muti-modal** → Local optimum
- ▶ We are interested in *Maximum Likelihood Estimate* (MLE), which shows good statistical performance.
- ▶ There is a significant gap when we move from **practical** use of EM algorithm to its **theoretical** understanding.



## Limitation of Classical Result

- ▶ Classical convergence result of EM algorithm by Wu (1983):
  - ▶ **Unimodal** → Global optimum
  - ▶ **Muti-modal** → Local optimum
- ▶ We are interested in *Maximum Likelihood Estimate* (MLE), which shows good statistical performance.
- ▶ There is a significant gap when we move from **practical** use of EM algorithm to its **theoretical** understanding.



- ▶ Our results show that if we are given an **appropriate initialization**, the EM update will converge to a local optima or stationary point but within a **statistical precision** of a global optima.

# Population $Q$ -function & Finite-sample $Q$ -function

- ▶ Finite-sample  $Q$ -function

$$Q_n(\boldsymbol{\Theta}' | \boldsymbol{\Theta}) := \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K L_{\boldsymbol{\Theta}, k}(\mathbf{x}_i) [\log \pi_k + \log f_k(\mathbf{x}_i; \boldsymbol{\Theta}'_k)].$$

It corresponds to *statistical error*.

- ▶ Population  $Q$ -function

$$Q(\boldsymbol{\Theta}' | \boldsymbol{\Theta}) := \mathbb{E} \left[ \sum_{k=1}^K L_{\boldsymbol{\Theta}, k}(\mathbf{X}) [\log \pi'_k + \log f_k(\mathbf{X}; \boldsymbol{\Theta}'_k)] \right].$$

It corresponds to *optimization error*.

# Final Estimation Error

## Theorem

Consider the SCAN procedure with initialization  $\Theta^{(0)} \in \mathcal{B}_\alpha(\Theta^*)$  for some *constant radius*  $\alpha$ . Let  $\kappa < 1$  be a contractive parameter. If the sample size  $n$  is large enough, the iterative estimator  $\Theta^{(t)}$  satisfies

$$\left\| \Theta^{(t)} - \Theta^* \right\|_2 \lesssim \underbrace{\varepsilon(n, p, K, \Psi(\mathcal{M}))}_{\text{Statistical Error (SE)}} + \underbrace{\kappa^t \left\| \Theta^{(0)} - \Theta^* \right\|_2}_{\text{Optimization Error (OE)}}, \quad (2)$$

with high probability.  $\Psi(\mathcal{M})$  measures the sparsity of cluster means and precision matrices. Here  $\Theta^* \in \mathbb{R}^{K(p^2+p)}$  is the true parameter.

# Statistical Error

## Corollary

When the iteration step  $t$  is large enough such that

$$t \geq T = \log_{1/\kappa} \frac{\|\Theta^{(0)} - \Theta^*\|_2}{\varphi(n, p, K)},$$

the optimization error is dominated by the statistical error, namely

$$\|\Theta^{(T)} - \Theta^*\|_2 = O_P \left( \underbrace{\sqrt{\frac{K^5 d \log p}{n}}}_{\text{Cluster means error}} + \underbrace{\sqrt{\frac{K^3 (Ks + p) \log p}{n}}}_{\text{Precision matrices error}} \right),$$

with probability converging to 1. Here  $d$  and  $s$  represent the sparsity of cluster mean and precision matrix for a single cluster.

## Some Remarks

- ▶ Given an **appropriate initialization**, our EM update will converge to a local optima or stationary point within a **statistical precision** of a global optima.
- ▶ In our framework, the number of cluster  $K$  can diverge with  $n$  and  $p$ , which fits our big data set up.
- ▶ When  $K$  is fixed:
  - ▶ If the cluster label is given in advance, the rate  $O_P(\sqrt{(s + p) \log p / n})$  is the optimal rate for precision matrix estimation.
  - ▶ If the covariance matrix is an identity matrix,  $O_P(\sqrt{d \log p / n})$  is the optimal rate for cluster mean estimation.

## Tuning Parameter Selection

- ▶ Choice of three tuning parameters  $\Lambda := \{\lambda_1, \lambda_2, \lambda_3\}$  to optimize the clustering and network estimation performance.
- ▶ Use the adaptive BIC criterion defined as

$$\text{BIC}(\Lambda) = -2 \log \hat{L}(\Lambda) + \log(n) \text{df}_\Lambda(\boldsymbol{\mu}) + 2 \text{df}_\Lambda(\boldsymbol{\Omega})$$

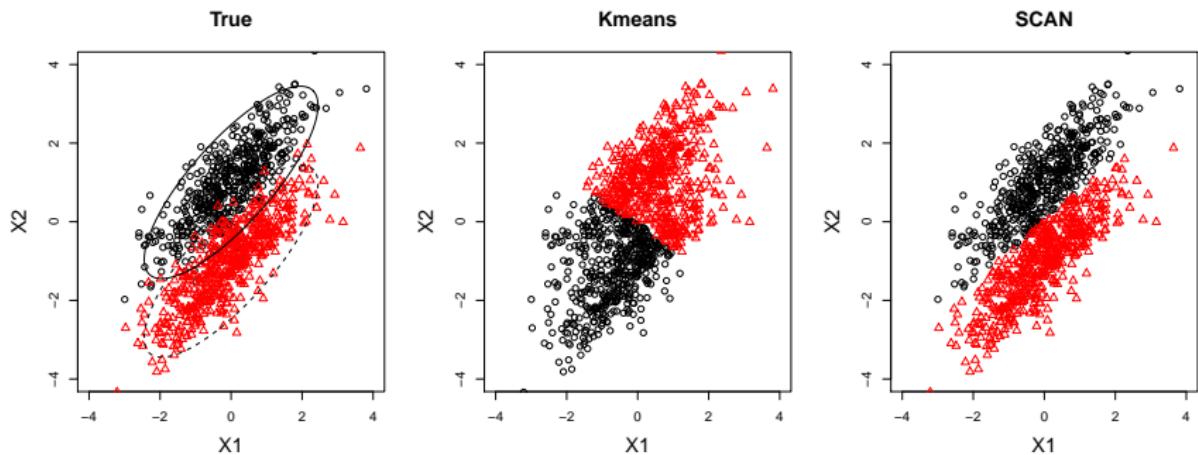
- ▶  $\{\text{df}_\Lambda(\boldsymbol{\mu}), \text{df}_\Lambda(\boldsymbol{\Omega})\}$ : the degrees of freedom of the model; can be approximated by the size of selected variables in the final estimator.

## Methods for Comparisons

- ▶ K-means (for clustering)
- ▶ K-means (for clustering) + JGL (for graphical model estimation; Danahar et al., 2014)
- ▶ Zhou et al. (2009): improved K-means by using a graphical lasso in precision matrix estimation.
- ▶ SCAN: simultaneous clustering and estimation of graphical models

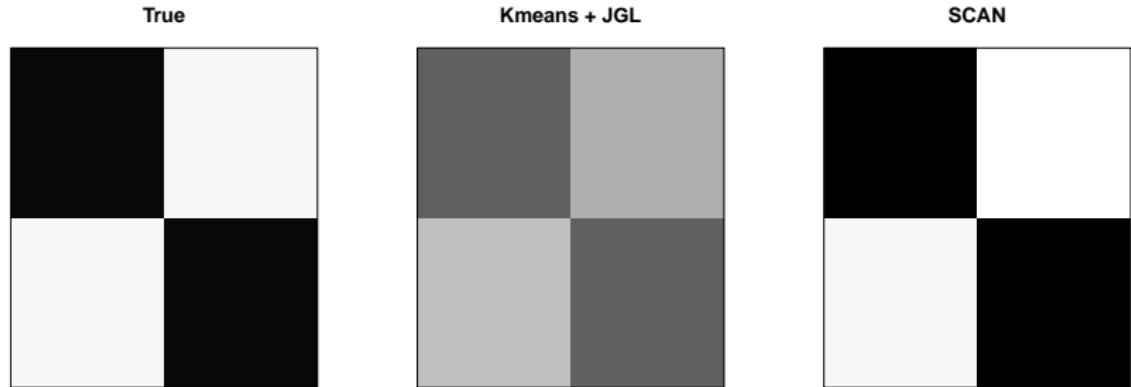
# Numerical Illustration: Clustering Results

Figure: True clusters (left); clusters obtained from standard  $K$ -means clustering (Kmeans; middle) and SCAN (right).



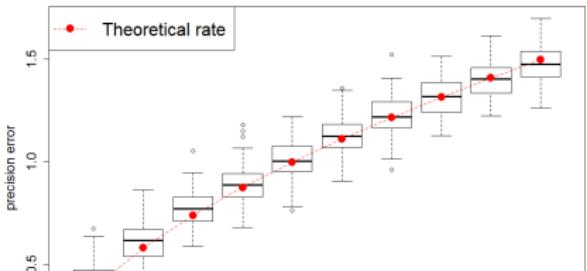
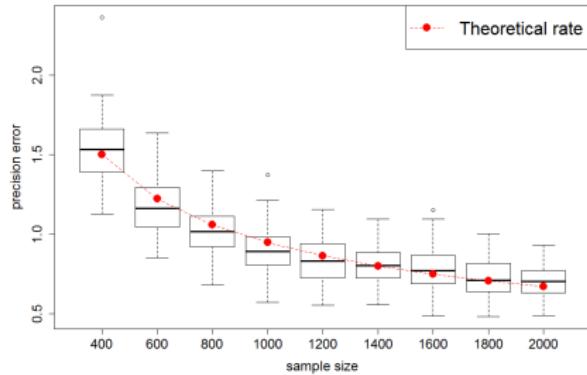
## Numerical Illustration: Precision Matrix Estimation

Figure: The true precision matrix and the estimated precision matrices from the two stage method (Kmeans + JGL) and our SCAN method



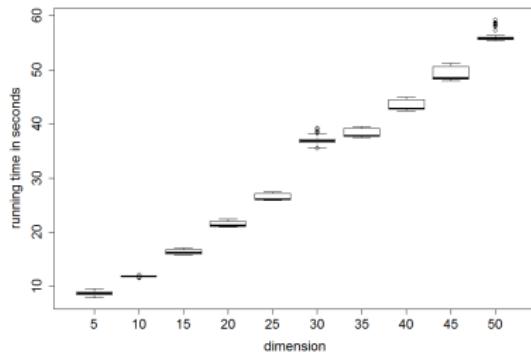
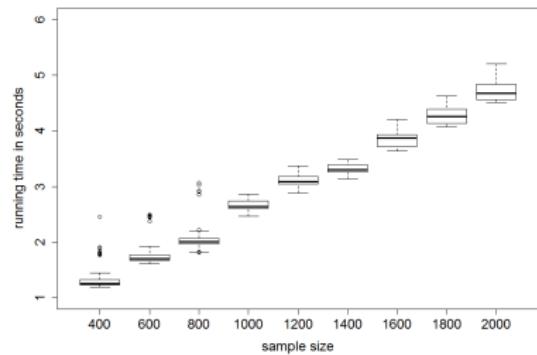
# Numerical Illustration: Errors

Figure: Comparison of the numerical error and the theoretical error rates of our SCAN method.



# Numerical Illustration: Computational Time

Figure: Running time of our algorithm (Left: varying sample sizes and fixed  $p = 10$ ; Right: varying dimensions and fixed  $n = 5000$ )



## Simulation Set Up

- ▶  $K = 3, n = 300, p = 100$ . Label  $Y_i$  is uniformly generated from  $\{1, 2, 3\}$ .
- ▶  $\mathbf{X}_i \sim \mathcal{N}(\boldsymbol{\mu}(Y_i), \Omega^{-1}(Y_i))$ .
- ▶ The first 10 variables of  $\boldsymbol{\mu}(Y_i)$  are

$$\begin{cases} (\mu \mathbf{1}_5^T, -\mu \mathbf{1}_5^T)^T & \text{if } Y_i = 1 \\ \mu \mathbf{1}_{10} & \text{if } Y_i = 2, \\ (-\mu \mathbf{1}_5^T, -\mu \mathbf{1}_5^T)^T & \text{if } Y_i = 3 \end{cases}$$

and the last 90 variables of  $\boldsymbol{\mu}(Y_i)$  are zeros.

- ▶  $\Omega_i$  are tridiagonal matrix with diagonal entries 1. Off-diagonal terms of  $\Omega_1, \Omega_2, \Omega_3$  is  $\eta, 0.9 * \eta, 1.1 * \eta$ , respectively.
- ▶ **Model 1:**  $\mu = 1, \eta = 0.4$ ; **Model 2:**  $\mu = 0.8, \eta = 0.3$ ; **Model 3:**  $\mu = 0.8, \eta = 0.4$ .

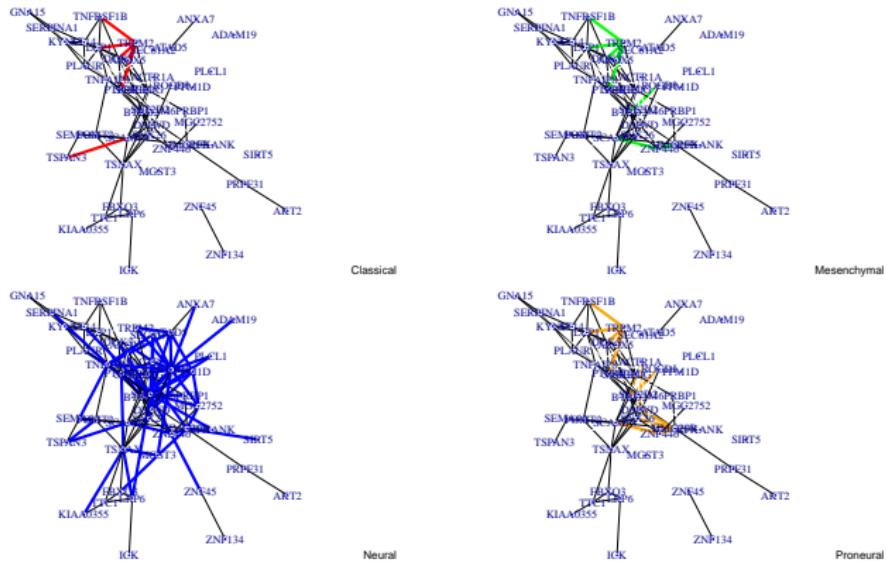
## Comparison Criteria

- ▶ The clustering errors (CE)
- ▶ Cluster mean error (CME)  $\frac{1}{K} \sum_{k=1}^K \left\| \hat{\boldsymbol{\mu}}^{(k)} - \boldsymbol{\mu}^{(k)} \right\|_2$
- ▶ Precision mean errors (PME)  $\frac{1}{K} \sum_{k=1}^K \left\| \hat{\boldsymbol{\Omega}}^{(k)} - \boldsymbol{\Omega}^{(k)} \right\|_F$
- ▶ True positive rate (TPR1) and false positive rate (FPR1) of mean estimation
- ▶ True positive rate (TPR2) and false positive rate (FPR2) of precision matrix estimation

# Simulation Results

Models	Methods	CE	CME	PME	TPR1/FPR1	TPR2 /FPR2
Model 1 $\mu = 1$ $\eta = 0.4$	K-means	0.219	3.65	NA	1/1	NA / NA
	Zhou et al.	0.134	2.33	11.62	0.99 /0.24	0.996 /0.154
	K-means + JGL	0.226	3.68	10.53	1 /1	0.991 /0.04
	SCAN	<b>0.070</b>	<b>1.97</b>	<b>8.57</b>	0.97 /0.23	0.998 /0.04
Model 2 $\mu = 0.8$ $\eta = 0.3$	K-means	0.186	2.58	NA	1 /1	NA / NA
	Zhou et al.	0.104	1.64	10.93	0.97 /0.16	0.96 /0.1
	K-means + JGL	0.112	1.79	8.05	1 /1	0.99 /0.007
	SCAN	<b>0.039</b>	<b>1.30</b>	<b>7.52</b>	1 /0.15	1 /0.006
Model 3 $\mu = 0.8$ $\eta = 0.4$	K-means	0.015	1.30	NA	1 /1	NA / NA
	Zhou et al.	0.024	1.38	10.42	0.99 /0	0.97 /0.099
	K-means + JGL	0.015	1.30	7.55	1 /1	0.999 /0.006
	SCAN	<b>0.007</b>	<b>1.29</b>	<b>7.50</b>	1 /0	0.999 /0.006

# GBM Applications: Classical, Mesenchymal, Neural and Proneural



## Summary and Extensions

- ▶ Simultaneous clustering and graphical model estimation.
- ▶ Estimation of high dimensional Gaussian mixture distributions.
- ▶ Efficient EM algorithm with theoretical guarantees.
- ▶ Numerical examples demonstrate the superior performance.
- ▶ Future investigation on the tuning parameter selection and theoretical properties.
- ▶ Extensions for other types of data.

## References

- ▶ Hao, B., Sun, W. W., Liu, Y. and Cheng, G. (2018) Simultaneous Clustering and Estimation of Heterogeneous Graphical Models. JMLR.
- ▶ Xie, Y., Liu, Y., and Valdar, W. (2016). Joint Estimation of Multiple Dependent Gaussian Graphical Models with Applications to Mouse Genetics. *Biometrika*, 103, 3, 493-511.
- ▶ Lee, W. and Liu, Y. (2015). Estimation of Multiple Graphical Models with Common Structures. *Journal of Machine Learning Research*, 16, 1035-1062.
- ▶ Balakrishnan, S., Wainwright, M. J., and Yu, B. (2016). Statistical guarantees for the em algorithm: From population to sample-based analysis. AoS.
- ▶ Danaher, P., Wang, P., and Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *JRSS-B*.
- ▶ Guo, J., Levina, E., Michailidis, G., and Zhu, J. (2011). Joint estimation of multiple graphical models. *Biometrika*.
- ▶ Zhou, H., Pan, W., and Shen, X. (2009). Penalized model-based clustering with unconstrained covariance matrices. *Electron. J. Statist.*