Joint Estimation and Inference for Multiple Multi-layered Gaussian Graphical Models

Subhabrata Majumdar

Abstract: The rapid development of high-throughput technologies has enabled generation of data from biological processes that span multiple layers, like genomic, proteomic or metabolomic data; and pertain to multiple sources, like disease subtypes or experimental conditions. 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 such datasets. We start with decomposing the multi-layer problem into a series of two-layer problems. For each two-layer problem, we model the outcomes at a node in the lower layer as dependent on those of other nodes in that layer, as well as all nodes in the upper layer. Following the biconvexity of our objective function, this estimation problem decomposes into two parts, where we use neighborhood selection and subsequent refitting of the precision matrix to quantify the dependency of two nodes in a single layer, and use group-penalized least square estimation to quantify the directional dependency of two nodes in different layers. Finally, to test for differences in these directional dependencies across multiple sources, we devise a hypothesis testing procedure that utilizes already computed neighborhood selection coefficients for nodes in the upper layer. We establish theoretical results for the validity of this testing procedure and the consistency of our estimates, and also evaluate their performance through simulations and a real data application.

Keywords: Data integration; Gaussian Graphical Models; Neighborhood selection; Group lasso

1 Notations

We shall denote scalars by small letters, vectors by bold small letters and matrices by bold capital letters. For any matrix $\mathbf{A}$, $(\mathbf{A})_{ij}$ denote its element in the $(i,j)^{\text{th}}$ position. For $a, b \in \mathbb{N}$, we denote the set of all $a \times b$ real matrices by $\mathbb{M}(a,b)$. For any positive integer c, define $\mathcal{I}_c = \{1, \ldots, c\}$.

2 Model

Consider the two -layered setup:

$$\mathbb{X}^k = (X_1^k, \ldots, X_p^k)^T \sim \mathcal{N}(0, \Sigma_x^k) \tag{2.1}$$

$$\mathbb{Y}^k = \mathbb{X}^k \mathbf{B}^k + \mathbb{E}^k; \quad \mathbb{E}^k = (E_1^k, \ldots, E_p^k)^T \sim \mathcal{N}(0, \Sigma_y^k) \tag{2.2}$$

$$\mathbf{B}^k \in \mathbb{M}(p,q); \quad \Omega_x^k = (\Sigma_x^k)^{-1}; \quad \Omega_y^k = (\Sigma_y^k)^{-1} \tag{2.3}$$

Want to estimate $\{(\Omega_x^k, \Omega_y^k, \mathbf{B}^k); k \in \mathcal{I}_K$ from data $\mathcal{Z}^k = \{(\mathbf{Y}^k, \mathbf{X}^k); \mathbf{Y}^k \in \mathbb{M}(n,q), \mathbf{X}^k \in \mathbb{M}(n,p), k \in \mathcal{I}_K\}$. in presence of known grouping structures $\mathcal{G}_x, \mathcal{G}_y, \mathcal{H}$ respectively.

Estimation of $\{\Omega_x^k\}$ done using JSEM. For the other part, we use the following two-step procedure:

1. Run neighborhood selection on y-network incorporating effects of x-data and an additional block-wise group penalty:

$$\min_{\mathcal{B},\Theta} \left\{ \sum_{j=1}^q \frac{1}{n_k} \left[\sum_{k=1}^K \|\mathbf{Y}_j^k - (\mathbf{Y}_{-j}^k - \mathbf{X}^k \mathbf{B}_{-j}^k)\boldsymbol{\theta}_j^k - \mathbf{X}^k \mathbf{B}_j^k\|^2 + \sum_{j \neq i} \sum_{g \in \mathcal{G}_y^{ij}} \lambda_{ij}^g \|\boldsymbol{\theta}_{ij}^{[g]}\| \right] + \sum_{b \in \mathcal{G}_x \times \mathcal{G}_y \times \mathcal{H}} \eta^b \|\mathbf{B}^{[b]}\| \right\} \tag{2.4}$$

$$= \min \left\{ f(\mathcal{Y}, \mathcal{X}, \mathcal{B}, \Theta) + P(\Theta) + Q(\mathcal{B}) \right\} \tag{2.5}$$

   where $\Theta = \{\Theta_i\}, \mathcal{B} = \{\mathbf{B}^k\}, \mathcal{Y} = \{\mathbf{Y}^k\}, \mathcal{X} = \{\mathbf{X}^k\}, \mathcal{E} = \{\mathbf{E}^k\}$.

   This estimates $\mathcal{B}$ **(possibly refit and/or within-group threshold)** .

2. Step I part 2 and step II of JSEM (see 15-656 pg 6) follows to estimate $\{\Omega_y^k\}$.

The objective function is bi-convex, so we are going to do the following in step 1-

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

- Iterate:

$$\Theta^{(t+1)} = \arg\min \left\{ f(\mathcal{Y}, \mathcal{X}, \mathcal{B}^{(t)}, \Theta^{(t)}) + P(\Theta^{(t)}) \right\} \tag{2.6}$$

$$\mathcal{B}^{(t+1)} = \arg\min \left\{ f(\mathcal{Y}, \mathcal{X}, \mathcal{B}^{(t)}, \Theta^{(t+1)}) + Q(\mathcal{B}^{(t)}) \right\} \tag{2.7}$$

- Continue till convergence.

3 Two-sample testing

Suppose there are two disease subtypes: $k = 1, 2$, and we are interested in testing whether the downstream effect of a predictor is X-data is same across both subtypes, i.e. if $\mathbf{b}_i^1 = \mathbf{b}_i^2$ for some $i \in \{1, \ldots, p\}$. For this we consider the modified optimization problem:

$$\min_{\mathcal{B},\Theta} \frac{1}{n} \left\{ \sum_{j=1}^{q} \sum_{k=1}^{2} \|\mathbf{Y}_j^k - \mathbf{Y}_{-j}^k \boldsymbol{\theta}_j^k - \mathbf{X}^k \mathbf{B}_j^k\|^2 + \sum_{j \neq j'} \lambda_{jj'} \|\boldsymbol{\theta}_{jj'}^*\| + \sum_{i=1}^{p} \eta_i \|\mathbf{B}_{i*}^*\| \right\} \tag{3.1}$$

$$= \min \left\{ f(\mathcal{Y}, \mathcal{X}, \mathcal{B}, \Theta) + P(\Theta) + Q(\mathcal{B}) \right\} \tag{3.2}$$

with $n_1 = n_2 = n$ for simplicity; and $\mathbf{B}^k = (\mathbf{b}_1^k, \ldots, \mathbf{b}_q^k), (\mathbf{B}_{i*}^*) \in \mathbb{R}^{q \times K}$

4 Conditions

Conditions A1 from JSEM paper holds for $\mathcal{X}$ and $\mathcal{E}$. Also A2, A3 from JSEM paper.

5 Results

To prove the results in this section, we shall use a reparametrization of the neighborhood coefficients at the lower level. Specifically, notice that for $j \in \mathcal{I}_q, k \in \mathcal{I}_K$, the corresponding summand in $f(\mathcal{Y}, \mathcal{X}, \mathcal{B}, \Theta)$ can be rearranged as

$$\begin{aligned} \|\mathbf{Y}_j^k - \mathbf{X}^k \mathbf{B}_j^k - (\mathbf{Y}_{-j}^k - \mathbf{X}^k \mathbf{B}_{-j}^k) \boldsymbol{\theta}_j^k\|^2 &= \|\mathbf{Y}_j^k - \mathbf{Y}_{-j}^k \boldsymbol{\theta}_j^k - (\mathbf{X}^k \mathbf{B}_j^k - \mathbf{X}^k \mathbf{B}_{-j}^k \boldsymbol{\theta}_j^k)\|^2 \\ &= \|(\mathbf{Y} - \mathbf{X}\mathbf{B}) \mathbf{T}_j^k\|^2 \end{aligned}$$

where

$$T_{jj'}^k = \begin{cases} 1 \text{ if } j = j' \\ -\theta_{jj'}^k \text{ if } j \neq j' \end{cases}$$

Thus, with $\mathbf{T}^k := (\mathbf{T}_j^k)_{j \in \mathcal{I}_q}$, we have

$$f(\mathcal{Y}, \mathcal{X}, \mathcal{B}, \Theta) = \frac{1}{n} \sum_{j=1}^p \sum_{k=1}^K \|(\mathbf{Y}^k - \mathbf{X}^k \mathbf{B}^k)\mathbf{T}_j^k\|^2 = \frac{1}{n} \sum_{k=1}^K \|\mathbf{Y}^k - \mathbf{X}^k \mathbf{B}^k)\mathbf{T}^k\|_F^2 = \sum_{k=1}^K \operatorname{Tr}(\mathbf{S}^k (\mathbf{T}^k)^2)$$

where $\mathbf{S}^k = (1/n)(\mathbf{Y}^k - \mathbf{X}^k \mathbf{B}^k)(\mathbf{Y}^k - \mathbf{X}^k \mathbf{B}^k)^T$ is the sample covariance matrix.

Now suppose $\boldsymbol{\beta} = \operatorname{vec}(\mathbf{B})$, and any subscript or superscript on $\mathbf{B}$ will be passed on to $\boldsymbol{\beta}$. Denote by $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Theta}$ the generic estimators given by

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{pq}}{\arg\min} \left\{ -2\boldsymbol{\beta}^T \widehat{\boldsymbol{\gamma}} + \boldsymbol{\beta}^T \widehat{\boldsymbol{\Gamma}} \boldsymbol{\beta} + \lambda_n \sum_{g \in \mathcal{G}} \|\boldsymbol{\beta}^{[g]}\| \right\} \quad (5.1)$$

$$\widehat{\Theta}_j = \underset{\Theta_j \in \mathbb{M}(q-1,K)}{\arg\min} \left\{ \frac{1}{n} \sum_{k=1}^K \|\mathbf{Y}_j^k - \mathbf{X}^k \widehat{\mathbf{B}}_j^k - (\mathbf{Y}_{-j}^k - \mathbf{X}^k \widehat{\mathbf{B}}_{-j}^k)\boldsymbol{\theta}_j^k\|^2 + \gamma_n \sum_{j \neq j'} \sum_{g \in \mathcal{G}_y^{jj'}} \|\boldsymbol{\theta}_{jj'}^{[g]}\| \right\} \quad (5.2)$$

where

$$\widehat{\boldsymbol{\Gamma}} = \begin{bmatrix} (\widehat{\mathbf{T}}^1)^2 \otimes \frac{(\mathbf{X}^1)^T \mathbf{X}^1}{n} & & \\ & \ddots & \\ & & (\widehat{\mathbf{T}}^K)^2 \otimes \frac{(\mathbf{X}^K)^T \mathbf{X}^K}{n} \end{bmatrix}; \quad \widehat{\boldsymbol{\gamma}} = \begin{bmatrix} (\widehat{\mathbf{T}}^1)^2 \otimes \frac{(\mathbf{X}^1)^T}{n} \\ \vdots \\ (\widehat{\mathbf{T}}^K)^2 \otimes \frac{(\mathbf{X}^K)^T}{n} \end{bmatrix} \begin{bmatrix} \operatorname{vec}(\mathbf{Y}^1) \\ \vdots \\ \operatorname{vec}(\mathbf{Y}^K) \end{bmatrix}$$

with $\widehat{\mathbf{T}}^k$ defined the same way using $\widehat{\boldsymbol{\theta}}_j^k$ as we defined $\mathbf{T}^k$ using $\boldsymbol{\theta}_j^k$.

Theorem 5.1. *Assume fixed $\mathcal{X}, \mathcal{E}$ and deterministic $\widehat{\mathcal{B}} = \{\widehat{\mathbf{B}}^k\}$. Also for $k = 1, \ldots, K$,*

(T1) $\|\widehat{\mathbf{B}}^k - \mathbf{B}_0^k\|_1 \leq v_\beta$, *where $v_\beta = \eta_\beta \sqrt{\frac{\log(\mathbf{pq})}{\mathbf{n}}}$ with η_β being a quantity depending on $\mathcal{B}$ only;*

(T2) *Denote $\widehat{\mathbf{E}}^k = \mathbf{Y}^k - \mathbf{X}^k \widehat{\mathbf{B}}^k, k \in \mathcal{I}_K$. Then for all $j \in \mathcal{I}_q$,*

$$\frac{1}{n} \left\| (\widehat{\mathbf{E}}_{-j}^k)^T \widehat{\mathbf{E}}^k \mathbf{T}_{0,j}^k \right\|_\infty \leq \mathbb{Q}\left(v_\beta, \Sigma_x^k, \Sigma_y^k\right)$$

where $\mathbb{Q}\left(v_\beta, \Sigma_x^k, \Sigma_y^k\right)$ is a $O(1/\sqrt{n})$ deterministic function depending on the population parameters $\mathcal{B}, \Sigma_x^k$ and Σ_y^k.

(T3) *Denote $\widehat{\mathbf{S}}^k = (\widehat{\mathbf{E}}^k)^T \widehat{\mathbf{E}}^k / n$. Then $\widehat{\mathbf{S}}^k \sim RE(\psi^k, \phi^k)$ with $Kq\phi \leq \psi/2$ where $\psi = \min_k \psi^k, \phi =$*

$\max_k \phi^k$;

(T4) *Assumption (A2) holds for* Σ_y^k.

Then, given the choice of tuning parameter

$$\gamma_n \geq 4\sqrt{q}\mathbb{Q}_{\max}; \quad \mathbb{Q}_{\max} := \max_{k \in \mathcal{I}_K} \mathbb{Q}\left(v_\beta, \Sigma_x^k, \Sigma_y^k\right)$$

the following holds

$$\frac{1}{K}\sum_{k=1}^{K} \|\widehat{\Omega}_y^k - \Omega_y^k\|_F^2 \leq \mathbf{O}\left(\frac{\mathbf{48c_0\sqrt{q}|g_{\max}|S}\mathbb{Q}_{\max}}{\psi}\right) \tag{5.3}$$

where $|g_{\max}|$ *is the maximum group size.*

Proof of Theorem 5.1. The proof has three parts, where we prove the consistency of the neighborhood regression coefficients, selection of edge sets, and finally the refitting step, respectively. This is the same structure as the proof of Theorem 1 in **?**, where they prove consistency of the JSEM estimates. The derivation of the first part is different from that in the JSEM proof, which we shall show in detail. The second and third parts follow similar lines, incorporating the updated quantities from part 1. For these we provide a rough sketch and leave the details to the reader.

Step 1: consistency of neighborhood regression. The following proposition establishes error bounds for the estimated y-neighborhood coefficients.

Proposition 5.2. *Consider the estimation problem in (5.2) and choose* $\gamma_n \geq 4\sqrt{q}\mathbb{Q}_{\max}$. *Given the conditions (T2) and (T3) hold, for any solution of (5.2) we shall have*

$$\|\widehat{\Theta}_j - \Theta_{0,j}\|_F \leq 12\sqrt{|g_{\max}|s_j}\gamma_n/\psi \tag{5.4}$$

$$\sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \|\hat{\boldsymbol{\theta}}_{jj'}^{[g]} - \boldsymbol{\theta}_{0,jj'}^{[g]}\| \leq 48|g_{\max}|s_j\gamma_n/\psi \tag{5.5}$$

Also denote the non-zero support of $\widehat{\Theta}_j$ *by* $\widehat{\mathcal{S}}_j$, *i.e.* $\widehat{\mathcal{S}}_j = \{(j', g) : \hat{\boldsymbol{\theta}}_{jj'}^{[g]} \neq \mathbf{0}\}$. *Then*

$$|\widehat{\mathcal{S}}_j| \leq 128|g_{\max}|s_j/\psi \tag{5.6}$$

Proof of Proposition 5.2. In its reparametrized version, (5.2) becomes

$$\widehat{\mathbf{T}}_j = \arg\min_{\mathbf{T}_j} \left\{ \frac{1}{n} \sum_{k=1}^{K} \|(\mathbf{Y}^k - \mathbf{X}^k \widehat{\mathbf{B}}^k) \mathbf{T}_j^k\|^2 + \gamma_n \sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \|\mathbf{T}_{jj'}^{[g]}\| \right\} \tag{5.7}$$

with $\mathbf{T}_{jj'}^{[g]} := (T_{jj'}^k)_{k \in g}$. Now for any $\mathbf{T}_j \in \mathbb{M}(q, K)$ we have

$$\frac{1}{n} \sum_{k=1}^{K} \|(\mathbf{Y}^k - \mathbf{X}^k \widehat{\mathbf{B}}^k) \widehat{\mathbf{T}}_j^k\|^2 + \gamma_n \sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \|\widehat{\mathbf{T}}_{jj'}^{[g]}\| \leq \frac{1}{n} \sum_{k=1}^{K} \|(\mathbf{Y}^k - \mathbf{X}^k \widehat{\mathbf{B}}^k) \mathbf{T}_j^k\|^2 + \gamma_n \sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \|\mathbf{T}_{jj'}^{[g]}\|$$

For $\mathbf{T}_j = \mathbf{T}_{0,j}$ this reduces to

$$\sum_{k=1}^{K} (\mathbf{d}_j^k)^T \widehat{\mathbf{S}}^k \mathbf{d}_j^k \leq -2 \sum_{k=1}^{K} (\mathbf{d}_j^k)^T \widehat{\mathbf{S}}^k \mathbf{T}_{0,j}^k + \gamma_n \sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \left(\|\mathbf{T}_{jj'}^{[g]}\| - \|\mathbf{T}_{jj'}^{[g]} + \mathbf{d}_{jj'}^{[g]}\| \right) \tag{5.8}$$

with $\widehat{\mathbf{T}}_j^k := \mathbf{T}_{0,j}^k + \mathbf{d}_j^k$ etc. For the k^{th} summand in the first term on the right hand side, since $d_{jj}^k = 0$, $\widehat{\mathbf{E}}^k \mathbf{d}_j^k = \widehat{\mathbf{E}}_{-j}^k \mathbf{d}_{-j}^k$. Thus by cauchy-schwarz inequality

$$\left| (\mathbf{d}_j^k)^T \widehat{\mathbf{S}}^k \mathbf{T}_{0,j}^k \right| \leq \|(\mathbf{d}_j^k\| \left\| \frac{(\widehat{\mathbf{E}}^k)^T \widehat{\mathbf{E}}^k}{n} \mathbf{T}_{0,j}^k \right\| \leq \sqrt{q} \|\mathbf{d}_j^k\| \left\| \frac{1}{n} (\widehat{\mathbf{E}}_{-j}^k)^T \widehat{\mathbf{E}}^k \mathbf{T}_{0,j}^k \right\|_\infty$$

which is $\leq \|\mathbf{d}_j^k\| \sqrt{q} \mathbb{Q}(v_\beta, \Sigma_x^k, \Sigma_y^k) \leq \|\mathbf{d}_j^k\| \gamma_n / 4$ by assumption (T2) and choice of γ_n. For the second term, suppose $\mathcal{S}_{0,j}$ is the support of $\Theta_{0,j}$, i.e. $\mathcal{S}_{0,j} = \{(j', g) : \boldsymbol{\theta}_{jj'}^{[g]} \neq 0\}$. Then

$$\begin{aligned}
\sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \left(\|\mathbf{T}_{jj'}^{[g]}\| - \|\mathbf{T}_{jj'}^{[g]} + \mathbf{d}_{jj'}^{[g]}\| \right) &\leq \sum_{(j', g) \in \mathcal{S}_{0,j}} \left(\|\mathbf{T}_{jj'}^{[g]}\| - \|\mathbf{T}_{jj'}^{[g]} + \mathbf{d}_{jj'}^{[g]}\| \right) - \sum_{(j', g) \notin \mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| \\
&\leq \sum_{(j', g) \in \mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| - \sum_{(j', g) \notin \mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\|
\end{aligned}$$

so that (5.8) reduces to

$$
\begin{aligned}
\sum_{k=1}^{K} (\mathbf{d}_j^k)^T \widehat{\mathbf{S}}^k \mathbf{d}_j^k &\leq \frac{\gamma_n}{2} \left[\sum_{(j',g)\in\mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| + \sum_{(j',g)\notin\mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| \right] + \gamma_n \left[\sum_{(j',g)\in\mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| - \sum_{(j',g)\notin\mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| \right] \\
&= \frac{3\gamma_n}{2} \sum_{(j',g)\in\mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| - \frac{\gamma_n}{2} \sum_{(j',g)\notin\mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| \\
&\leq \frac{3\gamma_n}{2} \sum_{j\neq j', g\in\mathcal{G}_y^{jj'}} \|\mathbf{d}_{jj'}^{[g]}\| \qquad (5.9)
\end{aligned}
$$

Since the left hand side is ≥ 0, this also implies

$$
\sum_{(j',g)\notin\mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| \leq 3 \sum_{(j',g)\in\mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| \quad \Rightarrow \sum_{j\neq j', g\in\mathcal{G}_y^{jj'}} \|\mathbf{d}_{jj'}^{[g]}\| \leq 4 \sum_{(j',g)\in\mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| \leq 4\sqrt{|g_{\max}| s_j} \|\mathbf{D}_j\|_F
$$

with $\mathbf{D}_j = (\mathbf{d}_j^k)_{k\in\mathcal{I}_K}$. Now the RE condition on $\widehat{\mathbf{S}}^k$ means that

$$
\sum_{k=1}^{K} (\mathbf{d}_j^k)^T \widehat{\mathbf{S}}^k \mathbf{d}_j^k \geq \sum_{k=1}^{K} \left(\psi_k \|\mathbf{d}_j^k\|^2 - \phi_k \|\mathbf{d}_j^k\|_1^2 \right) \geq \psi \|\mathbf{D}_j\|_F^2 - \phi \|\mathbf{D}_j\|_1^2 \geq (\psi - Kq\phi) \|\mathbf{D}_j\|_F^2 \geq \frac{\psi}{2} \|\mathbf{D}_j\|_F^2
$$

by assumption (T3). Thus we finally have

$$
\frac{\psi}{3} \|\mathbf{D}_j\|_F^2 \leq \gamma_n \sum_{j\neq j', g\in\mathcal{G}_y^{jj'}} \|\mathbf{d}_{jj'}^{[g]}\| \leq 4\gamma_n \sqrt{|g_{\max}| s_j} \|\mathbf{D}_j\|_F \qquad (5.10)
$$

Since

$$
(\mathbf{D}_j)_{j',k} = \widehat{T}_{jj'}^k - T_{0,jj'}^k = \begin{cases} 0 \text{ if } j = j' \\ -(\widehat{\theta}_{jj'}^k - \theta_{0,jj'}^k) \text{ if } j \neq j' \end{cases}
$$

The bounds in (5.4) and (5.5) are obtained by replacing the corresponding elements in (5.10).

For the bound on $|\widehat{\mathcal{S}}_j|$, notice that if $\hat{\boldsymbol{\theta}}_{jj'}^{[g]} \neq 0$ for some (j', g),

$$
\begin{aligned}
\frac{1}{n} \sum_{k\in g} \left| ((\widehat{\mathbf{E}}_{-j}^k)^T \widehat{\mathbf{E}}^k (\widehat{\mathbf{T}}_j^k - \mathbf{T}_{0,j}^k))^{j'} \right| &\geq \frac{1}{n} \sum_{k\in g} \left| ((\widehat{\mathbf{E}}_{-j}^k)^T \widehat{\mathbf{E}}^k \widehat{\mathbf{T}}_j^k)^{j'} \right| - \frac{1}{n} \sum_{k\in g} \left| ((\widehat{\mathbf{E}}_{-j}^k)^T \widehat{\mathbf{E}}^k \mathbf{T}_{0,j}^k)^{j'} \right| \\
&\geq |g|\gamma_n - \sum_{k\in g} \mathbb{Q}(v_\beta, \Sigma_x^k, \Sigma_y^k)
\end{aligned}
$$

using the KKT condition for (5.2) and assumption (T2). The choice of γ_n now ensures that the right hand side is $\geq 3|g|\gamma_n/4$. Hence

$$
\begin{aligned}
|\hat{\mathcal{S}}_j| &\leq \sum_{(j',g)\in\widehat{\mathcal{S}}_j} \frac{16}{9n^2|g|^2\gamma_n^2} \sum_{k\in g} \left| ((\widehat{\mathbf{E}}_{-j}^k)^T \widehat{\mathbf{E}}^k (\widehat{\mathbf{T}}_j^k - \mathbf{T}_{0,j}^k))^{j'} \right|^2 \\
&\leq \frac{16}{9\gamma_n^2} \sum_{k=1}^K \frac{1}{n} \left\| (\widehat{\mathbf{E}}_{-j}^k)^T \widehat{\mathbf{E}}^k (\widehat{\mathbf{T}}_j^k - \mathbf{T}_{0,j}^k) \right\|^2 \\
&= \frac{16}{9\gamma_n^2} \sum_{k=1}^K (\mathbf{d}_j^k)^T \widehat{\mathbf{S}}^k \mathbf{d}_j^k \\
&\leq \frac{8}{3\gamma_n} \sum_{j\neq j', g\in\mathcal{G}_y^{jj'}} \|\mathbf{d}_{jj'}^{[g]}\| \leq \frac{1}{\psi} 128 |g_{\max}| s_j
\end{aligned}
$$

using (5.9) and (5.10).

Step 2: Edge set selection. We denote the selected edge set for the k^{th} Y-network by $\hat{E}^k$. Denote its population version by E_0^k. Further, let

$$
\tilde{\Omega}_y^k = \text{diag}(\Omega_y^k) + \Omega_{y,E_0^k\cap\hat{E}^k}^k
$$

With similar derivations to the proof of Corollary A.1 in **?**, The following two upper bounds can be established:

$$
|\hat{E}^k| \leq \frac{128|g_{\max}|S}{\psi} \tag{5.11}
$$

$$
\frac{1}{K}\sum_{k=1}^K \|\tilde{\Omega}_y^k - \Omega_y^k\|_F \leq \frac{12c_0\sqrt{|g_{\max}|S}\gamma_n}{\sqrt{K}\psi} \tag{5.12}
$$

following which, taking $\gamma_n = 4\sqrt{q}\mathbb{Q}_{\max}$ and

$$
\Lambda_{\min}(\tilde{\Omega}_y^k) \geq d_0 - 12\sqrt{|g_{\max}|S}\gamma_n/\psi > 0; \quad \Lambda_{\max}(\tilde{\Omega}_y^k) \leq c_0 + 12\sqrt{|g_{\max}|S}\gamma_n/\psi < \infty \tag{5.13}
$$

with

$$
t_1 = \textbf{tbd}
$$

Step 3: Refitting.

□

Part II. Proof of Thm 2 in 15-656 follows. We only need a new bound for $Var(\mathbf{Y}_i^k|\mathbf{Y}_{-i}^k, \mathbf{X}^k, \widehat{\mathbf{B}}_i^k)$.

For this we have

$$Var(\mathbf{Y}_i^k|\mathbf{Y}_{-i}^k, \mathbf{X}^k, \widehat{\mathbf{B}}_i^k) = \mathbb{E}(\widehat{\boldsymbol{\epsilon}}_i^k)^2 = \mathbb{E}(\boldsymbol{\epsilon}_i^k + \boldsymbol{\delta}_i^k)^2 \leq \left(\frac{1}{d_0} + \frac{c(v_\beta)}{n} \right)^2$$

applying cauchy-schwarz inequality followed by assumption (A2). Now Replace $1/\sqrt{nd_0}$ in choice of λ, α_n in Thm 2 statement with $1/\sqrt{n}(\sqrt{1/d_0} + \sqrt{c(v_\beta)/n})$.

$\square$

Proposition 5.3. *Consider deterministic $\widehat{\mathcal{B}}$ satisfying assumption (T1). Then for sample size $n \succsim \log(pq)$ and $k \in \mathcal{I}_K$,*

1. *$\widehat{\mathbf{S}}^k$ satisfies the RE condition: $\widehat{\mathbf{S}}^k \sim RE(\psi^k, \phi^k)$, where*

$$\psi^k = \frac{\Lambda_{\min}(\Sigma_x^k)}{2}; \quad \phi^k = \frac{\psi^k \log p}{n} + 2 v_\beta c_2 [\Lambda_{\max}(\Sigma_x^k) \Lambda_{\max}(\Sigma_y^k)]^{1/2} \sqrt{\frac{\log(pq)}{n}}$$

with probability $\geq 1 - 6c_1 \exp[-(c_2^2 - 1)\log(pq)] - 2\exp(-c_3 n), c_1, c_3 > 0, c_2 > 1$.

2. *The following deviation bound is satisfied for any $j \in \mathcal{I}_q$*

$$\left\| \frac{1}{n} (\widehat{\mathbf{E}}_{-j}^k)^T \widehat{\mathbf{E}}^k \mathbf{T}_{0,j}^k \right\|_\infty \leq \mathbb{Q}\left(v_\beta, \Sigma_x^k, \Sigma_y^k \right)$$

with probability $\geq 1 - 1/p^{\tau_1 - 2} - 6c_1 \exp[-(c_2^2 - 1)\log(pq)] - 6c_4 \exp[-(c_5^2 - 1)\log(pq)], c_4 > 0, c_5 > 1$, where

$$\begin{aligned}\mathbb{Q}\left(v_\beta, \Sigma_x^k, \Sigma_y^k \right) = 2 v_\beta^2 & \left[\sqrt{\frac{\log 4 + \tau_1 \log p}{c_x^k n}} + \max_i \sigma_{x,ii}^k \right] + \\ & 4 v_\beta c_2 [\Lambda_{\max}(\Sigma_x^k)\Lambda_{\max}(\Sigma_y^k)]^{1/2} \sqrt{\frac{\log(pq)}{n}} + \\ & c_5 \left[\Lambda_{\max}(\Sigma_{y,-j}^k)(\sigma_{y,jj}^k + (\boldsymbol{\theta}_{0,j}^k)^T \Sigma_{y,-j}^k \boldsymbol{\theta}_{0,j}^k - 2(\boldsymbol{\sigma}_{y,j,-j}^k)^T \boldsymbol{\theta}_{0,j}^k) \right]^{1/2} \sqrt{\frac{\log(q-1)}{n}}\end{aligned}$$

Proof of Proposition 5.3. We drop the superscript k since there is no scope of ambiguity. For part 1, we start with an auxiliary lemma:

Lemma 5.4. *For a sub-gaussian design matrix $\mathbf{X} \in \mathbb{M}(n, p)$ with columns having mean $\mathbf{0}_p$ and covari-*

ance matrix Σ_x, the sample covariance matrix $\widehat{\Sigma}_x = \mathbf{X}^T\mathbf{X}/n$ satisfies the RE condition

$$\widehat{\Sigma}_x \sim RE\left(\frac{\Lambda_{\min}(\Sigma_x)}{2}, \frac{\Lambda_{\min}(\Sigma_x)\log p}{2n}\right)$$

with probability $\geq 1 - 2\exp(-c_3 n)$ for some $c_3 > 0$.

This is same as Lemma 2 in Appendix B of **?** and its proof can be found there. Now denote $\widehat{\mathbf{E}} = \mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}}$. For $\mathbf{v} \in \mathbb{R}^q$, we have

$$\begin{aligned}
\mathbf{v}^T\widehat{\mathbf{S}}\mathbf{v} &= \frac{1}{n}\|\widehat{\mathbf{E}}\mathbf{v}\|^2 \\
&= \frac{1}{n}\|(\mathbf{E} + \mathbf{X}(\mathbf{B}_0 - \widehat{\mathbf{B}}))\mathbf{v}\|^2 \\
&= \mathbf{v}^T\mathbf{S}\mathbf{v} + \frac{1}{n}\|\mathbf{X}(\mathbf{B}_0 - \widehat{\mathbf{B}})\mathbf{v}\|^2 + 2\mathbf{v}^T(\mathbf{B}_0 - \widehat{\mathbf{B}})^T\left(\frac{(\mathbf{X})^T\mathbf{E}}{n}\right)\mathbf{v} \qquad (5.14)
\end{aligned}$$

For the first summand, $\mathbf{v}^T\mathbf{S}^k\mathbf{v} \geq \psi_y\|\mathbf{v}\|^2 - \phi_y\|\mathbf{v}\|_1^2$ with $\psi_y = \Lambda_{\min}(\Sigma_y)/2, \phi_y = \psi_y \log p/n$ by applying Lemma 5.4 on $\mathbf{S}$. The second summand is greater than or equal to 0. For the third summand,

$$2\mathbf{v}^T(\mathbf{B}_0 - \widehat{\mathbf{B}})^T\left(\frac{(\mathbf{X})^T\mathbf{E}}{n}\right)\mathbf{v} \geq -2v_\beta\left\|\frac{(\mathbf{X})^T\mathbf{E}}{n}\right\|_\infty \|\mathbf{v}\|_1^2$$

by assumption (T1). Now we use another lemma:

Lemma 5.5. *For zero-mean independent sub-gaussian matrices $\mathbf{X} \in \mathbb{M}(n,p), \mathbf{E} \in \mathbb{M}(n,q)$ with parameters (Σ_x, σ_x^2) and (Σ_e, σ_e^2) respectively, given that $n \succsim \log(pq)$ the following holds with probability $\geq 1 - 6c_1\exp[-(c_2^2 - 1)\log(pq)]$ for some $c_1 > 0, c_2 > 1$:*

$$\frac{1}{n}\|\mathbf{X}^T\mathbf{E}\|_\infty \leq c_2[\Lambda_{\max}(\Sigma_x)\Lambda_{\max}(\Sigma_e)]^{1/2}\sqrt{\frac{\log(pq)}{n}}$$

This is a part of Lemma 3 of Appendix B in **?**, and is proved therein. Subsequently we collect all summands in (5.14) and get

$$\mathbf{v}^T\widehat{\mathbf{S}}\mathbf{v} \geq \psi_y\|\mathbf{v}\|^2 - \left(\phi_y + 2v_\beta c_2[\Lambda_{\max}(\Sigma_x)\Lambda_{\max}(\Sigma_y)]^{1/2}\sqrt{\frac{\log(pq)}{n}}\right)\|\mathbf{v}\|_1^2$$

with probability $\geq 1 - 2\exp(-c_3 n) - 6c_1\exp[-(c_2^2 - 1)\log(pq)]$. This concludes the proof of part 1.

To prove part 2, we decompose the quantity in question:

$$
\begin{aligned}
\left\| \frac{1}{n} \widehat{\mathbf{E}}_{-j}^T \widehat{\mathbf{E}} \mathbf{T}_{0,j} \right\|_\infty &= \left\| \frac{1}{n} \left[\mathbf{E}_{-j} + \mathbf{X}(\mathbf{B}_{0,j} - \widehat{\mathbf{B}}_j) \right]^T \left[\mathbf{E} + \mathbf{X}(\mathbf{B}_0 - \widehat{\mathbf{B}}) \right] \mathbf{T}_{0,j} \right\|_\infty \\
&\leq \left\| \frac{1}{n} \mathbf{E}_{-j}^T \mathbf{E} \mathbf{T}_{0,j} \right\|_\infty + \left\| \frac{1}{n} \mathbf{E}_{-j}^T \mathbf{X}(\mathbf{B}_0 - \widehat{\mathbf{B}}) \mathbf{T}_{0,j} \right\|_\infty \\
&+ \left\| \frac{1}{n} (\mathbf{B}_{0,j} - \widehat{\mathbf{B}}_j)^T \mathbf{X}^T \mathbf{X} (\mathbf{B}_0 - \widehat{\mathbf{B}}) \mathbf{T}_{0,j} \right\|_\infty + \left\| \frac{1}{n} (\mathbf{B}_{0,j} - \widehat{\mathbf{B}}_j)^T \mathbf{X}^T \mathbf{E} \mathbf{T}_{0,j} \right\|_\infty \\
&= \|\mathbf{W}_1\|_\infty + \|\mathbf{W}_2\|_\infty + \|\mathbf{W}_3\|_\infty + \|\mathbf{W}_4\|_\infty \qquad (5.15)
\end{aligned}
$$

Now

$$
\mathbf{W}_1 = \frac{1}{n} \mathbf{E}_{-j}^T (\mathbf{E}_j - \mathbf{E}_{-j} \boldsymbol{\theta}_{0,j})
$$

For node j in the y-network, $\mathbb{E}_{-j}$ and $E_j - \mathbb{E}_{-j}\boldsymbol{\theta}_{0,j}$ are the neighborhood regression coefficients and residuals, respectively. Thus they are orthogonal, so we can apply Lemma 5.5 on $\mathbf{E}_{-j}$ and $\mathbf{E}_j - \mathbf{E}_{-j}\boldsymbol{\theta}_{0,j}$ to obtain that for $n \succsim \log(q-1)$,

$$
\|\mathbf{W}_1\|_\infty \leq c_5 \left[\Lambda_{\max}(\Sigma_{y,-j})(\sigma_{y,jj} + \boldsymbol{\theta}_{0,j}^T \Sigma_{y,-j} \boldsymbol{\theta}_{0,j} - 2\boldsymbol{\sigma}_{y,j,-j}^T \boldsymbol{\theta}_{0,j}) \right]^{1/2} \sqrt{\frac{\log(q-1)}{n}} \qquad (5.16)
$$

holds with probability $\geq 1 - 6c_4 \exp[-(c_5^2 - 1)\log(pq)]$ for some $c_4 > 0, c_5 > 1$.

The same bounds hold for $\mathbf{W}_2$ and $\mathbf{W}_4$:

$$
\begin{aligned}
\|\mathbf{W}_2\|_\infty &\leq \left\| \frac{1}{n} \mathbf{E}_{-j}^T \mathbf{X}(\mathbf{B}_0 - \widehat{\mathbf{B}}) \right\|_\infty \|\mathbf{T}_{0,j}\|_1 \leq \left\| \frac{1}{n} \mathbf{E}^T \mathbf{X} \right\|_\infty \|\mathbf{B}_0 - \widehat{\mathbf{B}}\|_1 \|\mathbf{T}_{0,j}\|_1 \\
\|\mathbf{W}_4\|_\infty &\leq \left\| \frac{1}{n} (\mathbf{B}_{0,j} - \widehat{\mathbf{B}}_j)^T \mathbf{X}^T \mathbf{E} \right\|_\infty \|\mathbf{T}_{0,j}\|_1 \leq \left\| \frac{1}{n} \mathbf{E}^T \mathbf{X} \right\|_\infty \|\mathbf{B}_0 - \widehat{\mathbf{B}}\|_1 \|\mathbf{T}_{0,j}\|_1
\end{aligned}
$$

Now since Ω_y is diagonally dominant, $|\omega_{y,jj}| \geq \sum_{j \neq j'} |\omega_{y,jj'}|$ for any $j \in \mathcal{I}_q$. Hence

$$
\|\mathbf{T}_{0,j}\|_1 = \sum_{j'=1}^{q} |T_{jj'}| = 1 + \sum_{j \neq j'} |\theta_{jj'}| = 1 + \frac{1}{\omega_{y,jj}} \sum_{j \neq j'} |\omega_{y,jj'}| \leq 2
$$

so that for $n \succsim \log(pq)$,

$$
\|\mathbf{W}_2\|_\infty + \|\mathbf{W}_4\|_\infty \leq 4 v_\beta c_2 [\Lambda_{\max}(\Sigma_x)\Lambda_{\max}(\Sigma_y)]^{1/2} \sqrt{\frac{\log(pq)}{n}} \qquad (5.17)
$$

with probability $\geq 1 - 6c_1 \exp[-(c_2^2 - 1)\log(pq)]$ by applying Lemma 5.5 and assumption (T1).

Finally for $\mathbf{W}_3$, we apply Lemma 8 of **?** on the (sub-gaussian) design matrix $\mathbf{X}$ to obtain that for sample size

$$n \geq 512(1 + 4\Lambda_{\max}(\Sigma_x^k))^4 \max_j (\sigma_{x,jj}^k)^4 \log(4p^{\tau_1})$$

we get that with probability $\geq 1 - 1/p^{\tau_1 - 2}, \tau_1 > 2$,

$$\left\| \frac{\mathbf{X}^T\mathbf{X}}{n} \right\|_\infty \leq \sqrt{\frac{\log 4 + \tau_1 \log p}{c_x n}} + \max_i \sigma_{x,ii}; \quad c_x = \left[128(1 + 4\Lambda_{\max}(\Sigma_x))^2 \max_i (\sigma_{x,ii})^2 \right]^{-1}$$

Thus with the same probability,

$$\|\mathbf{W}_4\|_\infty \leq \left\| \frac{\mathbf{X}^T\mathbf{X}}{n} \right\|_\infty \|\widehat{\mathbf{B}} - \mathbf{B}_0\|_1^2 \|\mathbf{T}_{0,j}\|_1 \leq 2v_\beta^2 \left[\sqrt{\frac{\log 4 + \tau_1 \log p}{c_x n}} + \max_i \sigma_{x,ii} \right] \tag{5.18}$$

We now bound the right hand side of (5.15) using (5.16), (5.17) and (5.18) to complete the proof, with the leading term of the sample size requirement being $n \gtrsim \log(pq)$.

$\square$

Now concentrate on the k-population estimation problem. We want to obtain

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^{pq}} \left\{ -2\boldsymbol{\beta}^T \widehat{\boldsymbol{\gamma}} + \boldsymbol{\beta}^T \widehat{\boldsymbol{\Gamma}} \boldsymbol{\beta} + \lambda_n \sum_{g \in \mathcal{G}} \|\boldsymbol{\beta}^{[g]}\| \right\}$$

Theorem 5.6. *Assume fixed $\mathcal{X}, \mathcal{E}$, and deterministic $\widehat{\Theta} = \{\widehat{\Theta}_j\}$. Also for $j \in \mathcal{I}_q$,*

(B1) $\|\widehat{\Theta}_j - \Theta_{0,j}\|_F \leq v_\Theta \sqrt{\frac{\log q}{n}}$ *for some v_Θ dependent on Θ.*

(B2) *The deviation bound holds:*

$$\|\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\Gamma}} \boldsymbol{\beta}_0\|_\infty \leq \mathbb{R}(v_\Theta, \Sigma_y)$$

tbd

(B3) $\widehat{\boldsymbol{\Gamma}} \sim RE(\psi^*, \phi^*)$ *with $Kpq\phi^* \leq \psi^*/2$.*

Then, given the choice of tuning parameter

$$\lambda \geq 4\sqrt{pq}\mathbb{R}(v_\Theta, \Sigma_y)$$

the following holds

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \leq 12\sqrt{s_\beta}\lambda_n/\psi^* \tag{5.19}$$

$$\sum_{g\in\mathcal{G}} \|\boldsymbol{\beta}^{[g]} - \boldsymbol{\beta}_0^{[g]}\| \leq 48 s_\beta \lambda_n/\psi^* \tag{5.20}$$

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T\widehat{\boldsymbol{\Gamma}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \leq 72 s_\beta \lambda_n^2/\psi^* \tag{5.21}$$

Also denote the non-zero support of $\widehat{\boldsymbol{\beta}}$ *by* $\widehat{\mathcal{S}}_\beta$*, i.e.* $\widehat{\mathcal{S}}_\beta = \{g : \hat{\boldsymbol{\beta}}^{[g]} \neq \mathbf{0}\}$*. Then*

$$|\widehat{\mathcal{S}}_\beta| \leq 128 s_\beta/\psi^* \tag{5.22}$$

Proof of Theorem 5.6. The proof follows that of Theorem 5.1, with a different group norm structure. We shall only point our the differences.

Putting $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ in (5.1) we get

$$-2\widehat{\boldsymbol{\beta}}^T\widehat{\boldsymbol{\gamma}} + \boldsymbol{\beta}^T\widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\beta}} + \lambda_n \sum_{g\in\mathcal{G}} \|\widehat{\boldsymbol{\beta}}^{[g]}\| \leq -2\boldsymbol{\beta}_0^T\widehat{\boldsymbol{\gamma}} + \boldsymbol{\beta}_0^T\widehat{\boldsymbol{\Gamma}}\boldsymbol{\beta}_0 + \lambda_n \sum_{g\in\mathcal{G}} \|\boldsymbol{\beta}_0^{[g]}\|$$

Denote $\mathbf{b} = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$. Then we have

$$\mathbf{b}^T\widehat{\boldsymbol{\Gamma}}\mathbf{b} \leq 2\mathbf{b}^T(\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\Gamma}}\boldsymbol{\beta}_0) + \lambda_n \sum_{g\in\mathcal{G}} (\|\boldsymbol{\beta}_0^{[g]}\| - \|\boldsymbol{\beta}_0^{[g]} + \mathbf{b}^{[g]}\|)$$

Proceeding similarly as the proof of Theorem 5.1, with a different deviation bound and choice of λ_n based on that, we get expressions equivalent to (5.9) and (5.10) respectively:

$$\mathbf{b}^T\widehat{\boldsymbol{\Gamma}}\mathbf{b} \leq \frac{3}{2}\sum_{g\in\mathcal{G}} \|\mathbf{b}^{[g]}\| \tag{5.23}$$

$$\frac{\psi^*}{3}\|\mathbf{b}\|^2 \leq \lambda_n \sum_{g\in\mathcal{G}} \|\mathbf{b}^{[g]}\| \leq 4\lambda_n\sqrt{s_\beta}\|\mathbf{b}\| \tag{5.24}$$

The bounds in (5.19), (5.20), (5.21) and (5.22) follow.

□

Proposition 5.7. *Consider deterministic* $\widehat{\Theta}$ *satisfying assumption (B1). Then for sample size* $n \gtrsim \log(pq)$*,*

1. $\widehat{\mathbf{\Gamma}}$ *satisfies the RE condition:* $\widehat{\mathbf{\Gamma}} \sim RE(\psi^*, \phi^*)$, *where*

$$\textbf{tbd}$$

with probability **tbd**.

2. *The following deviation bound is satisfied:*

$$\|\widehat{\boldsymbol{\gamma}} - \widehat{\mathbf{\Gamma}}\boldsymbol{\beta}_0\|_\infty \leq \textbf{tbd}$$

with probability **tbd**.