Supplementary document for the article:
A two-stage sequential conditional selection approach to sparse high-dimensional multivariate regression models

Zehua Chen[1] AND Yiwei Jiang[2]

In this document, we provide some supplementary materials to the article "A two-stage sequential conditional selection approach to sparse high-dimensional multivariate regression models". The first part of the document contains the technical details of the proofs for the asymptotic properties of the TASCS approach. The second part contains some miscellaneous materials.

1 Technical proofs

In order to make the document self-contained for the convenience of the reader, the conditions and the results in Section 3 of the article are re-stated in this section.

A1 $\ln p = O(n^{\kappa})$, where $0 < \kappa < 1/3$; $\max_j p_{0j} = O(n^c)$, for some $0 < c < 1/6$;

A2 For any $s_j \subset s_{0j}$ but $s_j \neq s_{0j}$, $\max_{k \in s_{0j} \setminus s_j} |\gamma_j(k, s_j)| > \max_{k \notin s_{0j}} |\gamma_j(k, s_j)|$;

A3 $\min_{1 \leq j \leq q} \left\{\lambda_{min}[\frac{1}{n} X^{\top}(s_{0j}) X(s_{0j})] \min_{k \in s_{0j}} |\beta_{jk}|\right\} \geq C n^{-1/6+\delta}$, where δ is an arbitrary small positive number.

A4 $\lim_{n\to\infty} \min_{1\leq j\leq q} \min_{s_j : s_{0j} \not\subset s_j, |s_j| \leq k p_{0j}} \frac{\Delta(s_j, \boldsymbol{\mu}_j)}{p_{0j} \ln p} \to \infty$, where $\Delta(s_j, \boldsymbol{\mu}_j) = \boldsymbol{\mu}_j^{\top}[I - H^{\mathrm{X}}(s_j)]\boldsymbol{\mu}_j$, $\boldsymbol{\mu}_j = X\boldsymbol{\beta}_j$, and $k > 1$ is a fixed constant.

Theorem 1. We have

(i) Assume conditions $A1 - A4$,

$$P(\hat{s}_{0j}^{\mathrm{M}} = s_{0j}, j = 1, \ldots, q) \to 1,$$

as $n \to \infty$.

(ii) In addition, suppose $P(\hat{\mathcal{T}}_0^{\mathrm{M}} = \mathcal{T}_0) \to 1$. Then

$$P(\hat{s}_{0j}^{\mathrm{C}} = s_{0j}, j = 1, \ldots, q) \to 1,$$

as $n \to \infty$.

The theorem is proved by establishing the Lemmas s1 — s4 given below.

Lemma s1. Let $\hat{s}_{j1}^{\mathrm{M}} \subset \cdots \subset \hat{s}_{jk}^{\mathrm{M}} \subset \cdots$ be the sequence of sets of selected features for the jth marginal model of the sparse high-dimensional multivariate regression model obtained in the first iteration of TASCS. Under conditions $A1 - A3$,

$$P(\hat{s}_{j1}^{\mathrm{M}} \subset \cdots \subset \hat{s}_{jk}^{\mathrm{M}} \subset \cdots \subset \hat{s}_{jp_{0j}-1}^{\mathrm{M}} \subset \hat{s}_{jp_{0j}}^{\mathrm{M}} = s_{0j}; \quad j = 1, \ldots, q) > 1 - r_{1n}, \text{ as } n \to \infty,$$

where $r_{1n} = \frac{2 \max_{1 \le j \le q} \sigma_j}{C_n^{1/2} \ln p} \exp\left\{ -\frac{(\ln p)^2}{2} + \ln p + \ln q + \ln(\max p_{0j}) \right\}$, and

$$C_n = \min_{1 \le j \le q} \left\{ \frac{\sqrt{n}}{\ln p} \lambda_{min}[\frac{1}{n} X^\top(s_{0j}) X(s_{0j})] \min_{k \in s_{0j}} |\beta_{jk}| \right\}.$$

Lemma s2. Under conditions $A1 - A3$,

$$P\left(\mathrm{EBIC}(\hat{s}_{jk}^{\mathrm{M}}) > \mathrm{EBIC}(\hat{s}_{jk+1}^{\mathrm{M}}), 1 \le k < p_{0j}; \quad j = 1, \ldots, q\right) > 1 - r_{2n},$$

where $r_{2n} = O(n^{-1+\omega})$ for some $\omega < 1$.

Lemma s3. Let s_j be any set of features for the jth marginal model satisfying $|s_j| \le k p_{0j}$ for a given $k > 1$. Under conditions $A1 - A4$,

$$P\left(\min_{s_j} \mathrm{EBIC}(s_j) > \mathrm{EBIC}(s_{0j}), \quad j = 1, \ldots, q \right) > 1 - r_{3n},$$

where $r_{3n} = \frac{Cq}{R_n \ln p}$, R_n diverging to ∞ and C being a generic constant.

Lemma s4. Assume conditions $A1 - A3$ and, in addition, suppose that $P(\hat{\mathcal{T}}_0^{\mathrm{M}} = \mathcal{T}_0) \to 1$ where $\hat{\mathcal{T}}_0^{\mathrm{M}}$ is the estimate of $\mathcal{T}_0$ from the first stage. Let $\hat{s}_{j1}^{\mathrm{C}} \subset \cdots \subset \hat{s}_{jk}^{\mathrm{C}} \subset \cdots$ be the sequence of sets of selected features for the jth marginal model of the sparse high-dimensional multivariate regression model obtained in the second stage of TASCS. Then

$$P(\hat{s}_{j1}^{\mathrm{C}} \subset \cdots \subset \hat{s}_{jk}^{\mathrm{C}} \subset \cdots \subset \hat{s}_{jp_{0j}-1}^{\mathrm{C}} \subset \hat{s}_{jp_{0j}}^{\mathrm{C}} = s_{0j}; \quad j = 1, \ldots, q) \to 1, \text{ as } n \to \infty.$$

Lemma s1 states that the true set s_{0j} is among the increasing sequence of the selected sets of features. Lemma s2 states that the EBIC values of the sequence of the selected sets are decreasing until the sequence reaches the true set. Lemma s3 implies that the EBIC values of the selected sets after the true set are larger than that of the true one. By the nature of the sequential procedure, the procedure will stop at exactly the step when the true set is selected. The probability of these events happen simultaneously converges to 1. Thus Thorem 1 (i) follows. Lemma s2 can be established for the sequence $\hat{s}_{jk}^{\mathrm{C}}$ in exactly the same argument as that for the sequence $\hat{s}_{jk}^{\mathrm{M}}$ with a convergence rate at least not slower. Then, Lemma s2, Lemma s3 together with Lemma s4 establish that $P(\hat{s}_{0j}^{\mathrm{C}} = s_{0j}, j = 1, \ldots, q) \to 1$. This proves Theorem 1 (ii). By the way, Proposition 1 is obvious from Lemma s1.

In what follows, we provide the proofs for Lemmas s1 — s4.

Proof of Lemma s1

Proof. We establish that

$$P(s_{j1}^{*\mathrm{M}} \subset \cdots \subset s_{jk}^{*\mathrm{M}} \subset \cdots \subset s_{0j} \subset \cdots; \quad j = 1, \ldots, q) > 1 - r_{1n}, \text{ as } n \to \infty. \tag{1}$$

In what follows, we denote $s_{jk}^{*\mathrm{M}}$ by s_{jk}^{*} for convenience. Let

$$\hat{\gamma}_j(l, s_{jk}^*) = \frac{1}{n}\boldsymbol{x}_l^\top [I_n - H^{\mathrm{X}}(s_{jk}^*)]\boldsymbol{y}_j = \gamma_j(l, s_{jk}^*) + \frac{1}{n}\boldsymbol{x}_l^\top [I_n - H^{\mathrm{X}}(s_{jk}^*)]\boldsymbol{e}_j.$$

A slight modification of the proof of Theorem 3.1 in Luo and Chen (2014) yields

$$P\left(\max_{l \in s_{jk}^{*c}} \frac{1}{n}\boldsymbol{x}_l^\top [I - H^{\mathrm{X}}(s_{jk}^*)]\boldsymbol{e}_j \geq C_{jn}^{1/2} n^{-1/2} \ln p\right) \leq \frac{2\sigma_j}{C_{jn}^{1/2} \ln p} \exp\left\{-\frac{(\ln p)^2}{2} + \ln p\right\},$$

where $C_{jn} = \frac{\sqrt{n}}{\ln p}\lambda_{\min}(\frac{1}{n}X^\tau(s_{0j})X(s_{0j})) \min_{l \in s_{0j}} |\beta_{jl}|$. By Bonferroni inequality, we have

$$\begin{aligned} & P\left(\max_{0<k<p_{0j}} \max_{l \in s_{jk}^{*c}} \frac{1}{n}\boldsymbol{x}_l^\top [I - H^{\mathrm{X}}(s_{jk}^*)]\boldsymbol{e}_j \geq C_{jn}^{1/2} n^{-1/2} \ln p, \ j = 1, \ldots, q\right) \\ \leq \quad & \frac{2\max_j \sigma_j}{\min C_{jn}^{1/2} \ln p} \exp\left\{-\frac{(\ln p)^2}{2} + \ln p + \ln q + \ln(\max p_{0j})\right\} \\ = \quad & r_{1n}. \end{aligned} \tag{2}$$

It was shown in Luo and Chen (2014) that

$$\max_{l \in s_{jk}^{*-}} |\gamma_j(l, s_{jk}^*)| \geq C_{jn} n^{-1/2} \ln p, \tag{3}$$

where $s_{jk}^{*-} = s_{jk}^{*c} \cap s_{0j}$. By conditions A1 and A3, $\min_j C_{jn} \to \infty$. Thus, (2), (3) and condition A2 imply that

$$P\left(\max_{l \in s_{jk}^{*-}} |\hat{\gamma}_j(l, s_{jk}^*)| > \max_{l \in s_{0j}^c} |\hat{\gamma}_j(l, s_{jk}^*)|, 0 \leq k \leq p_{0j}, j = 1, \dots, q \right) > 1 - r_{1n}. \tag{4}$$

Given $s_{jk}^* \subset s_{0j}$, $\max_{l \in s_{jk}^{*-}} |\hat{\gamma}_j(l, s_{jk}^*)| > \max_{l \in s_{0j}^c} |\hat{\gamma}_j(l, s_{jk}^*)|$ implies that $s_{jk+1}^* \subset s_{0j}$. Since the sequential procedure starts with $s_{j0}^* = \emptyset \subset s_{0j}$, the lemma is implied by (4). □

Proof of Lemma s2

Proof. Let $D_{jk} = \text{EBIC}(s_{jk}^*) - \text{EBIC}(s_{jk+1}^*)$. We have

$$\begin{aligned} D_{jk} &= n \ln \left(\frac{\|(I - H^{\text{X}}(s_{jk}^*))\boldsymbol{y}_j\|_2^2}{\|(I - H^{\text{X}}(s_{jk+1}^*))\boldsymbol{y}_j\|_2^2} \right) - (\ln n + 2\gamma \ln p) \\ &= n \ln \left(1 + \frac{\|(I - H^{\text{X}}(s_{jk}^*))\boldsymbol{y}_j\|_2^2 - \|(I - H^{\text{X}}(s_{jk+1}^*))\boldsymbol{y}_j\|_2^2}{\|(I - H^{\text{X}}(s_{*k+1}))\boldsymbol{y}_j\|_2^2} \right) - (\ln n + 2\gamma \ln p) \\ &= T_{jk} - (\ln n + 2\gamma \ln p), \text{ say.} \end{aligned}$$

The lemma holds if $\min_{1 \leq k \leq p_{0j}, j=1,\dots,q} T_{jk} > \ln n + 2\gamma \ln p$ with probability bigger than $1 - r_{2n}$, which is implied by

$$\begin{aligned} &\min_{jk} \left\{ \|(I - H^{\text{X}}(s_{jk}^*))\boldsymbol{y}_j\|_2^2 - \|(I - H^{\text{X}}(s_{jk+1}^*))\boldsymbol{y}_j\|_2^2 \right\} \\ \geq\ & Cn \min_j (\lambda_{\min}[\frac{1}{n} X(s_{0j})^\top X(s_{0j})] \min_{l \in s_{0j}} |\beta_{jl}|)^2, \end{aligned} \tag{5}$$

and

$$\max_{jk} \|(I - H^{\text{X}}(s_{jk+1}^*))\boldsymbol{y}_j\|_2^2 \leq Cn \max\{p_{0j}^2\}, \tag{6}$$

with probability bigger than $1 - r_{2n}$, where C is a generic constant. It is because that, when (5) and (6) hold, we have

$$\begin{aligned}
\min_{1\le k<p_{0j},j=1,\dots q} T_{jk} &\ge n\ln\left(1+C\frac{\min_j[\lambda_{\min}(\frac{1}{n}X(s_{0j})^\top X(s_{0j}))\min_{l\in s_{0j}}|\beta_{jl}|]^2}{\max_j p_{0j}^2}\right)\\
&\ge \frac{nC}{2\max p_{0j}^2}[\min_j \lambda_{\min}(\frac{1}{n}X(s_{0j})^\top X(s_{0j}))\min_{l\in s_{0j}}|\beta_{jl}|]^2\\
&> Cn^{1/3+2\delta} > \ln n + 2\gamma\ln p,
\end{aligned}$$

where the last two inequalities hold by condition A1 and A3.

In the following, we verify (5) and (6). For any given set s, vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, let $\Delta(s,\boldsymbol{u}) = \boldsymbol{u}^\top[I-H^{\mathrm{X}}(s)]\boldsymbol{u}$ and $\Delta(s,\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{u}^\top[I-H^{\mathrm{X}}(s)]\boldsymbol{v}$. To verify (5), express

$$\begin{aligned}
&\|(I-H^{\mathrm{X}}(s_{jk}^*))\boldsymbol{y}_j\|_2^2 - \|(I-H^{\mathrm{X}}(s_{jk+1}^*))\boldsymbol{y}_j\|_2^2\\
=\ &[\Delta(s_{jk}^*,\boldsymbol{\mu}_j)-\Delta(s_{jk+1}^*,\boldsymbol{\mu}_j)] + 2[\Delta(s_{jk}^*,\boldsymbol{\mu}_j,\boldsymbol{e}_j)-\Delta(s_{jk+1}^*,\boldsymbol{\mu}_j,\boldsymbol{e}_j)] + [\Delta(s_{jk}^*,\boldsymbol{e}_j)-\Delta(s_{jk+1}^*,\boldsymbol{e}_j)].
\end{aligned}$$

It has been shown in Luo and Chen (2014) that

$$\min_k[\Delta(s_{jk}^*,\boldsymbol{\mu}_j)-\Delta(s_{jk+1}^*,\boldsymbol{\mu}_j)] \ge n(\lambda_{\min}[\frac{1}{n}X(s_{0j})^\top X(s_{0j})]\min_{l\in s_{0j}}|\beta_{jl}|)^2.$$

Therefore

$$\min_{jk}[\Delta(s_{jk}^*,\boldsymbol{\mu}_j)-\Delta(s_{jk+1}^*,\boldsymbol{\mu}_j)] \ge n\min_j(\lambda_{\min}[\frac{1}{n}X(s_{0j})^\top X(s_{0j})]\min_{l\in s_{0j}}|\beta_{jl}|)^2 \equiv d_n. \tag{7}$$

Since $\Delta(s_{*k},\boldsymbol{e}_j)-\Delta(s_{*k+1},\boldsymbol{e}_j)$ follows a χ^2 distribution with degree of freedom 1, by Bonferroni inequality, we have,

$$\begin{aligned}
&P(\max_{1\le k<p_{0j};j=1,\dots,q}[\Delta(s_{jk}^*,\boldsymbol{e}_j)-\Delta(s_{jk+1}^*,\boldsymbol{e}_j)] \ge d_n^{1/2})\\
\le\ & q\max_j p_{0j}P(\chi_1^2\ge d_n^{1/2}) = 2q\max_j p_{0j}[1-\Phi(d_n^{1/4})]\\
\le\ & \frac{C}{d_n^{1/4}}\exp\{-\frac{d_n^{1/2}}{2}+\ln q+\ln\max_j p_{0j}\}.
\end{aligned} \tag{8}$$

Note that $\frac{\Delta(s_{jk}^*,\boldsymbol{\mu}_j,\boldsymbol{e}_j)-\Delta(s_{jk+1}^*,\boldsymbol{\mu}_j,\boldsymbol{e}_j)}{\sqrt{\Delta(s_{jk}^*,\boldsymbol{\mu}_j)-\Delta(s_{jk+1}^*,\boldsymbol{\mu}_j)}} = Z_{jk}$ follows a standard normal distribution. The same argument as above yields

$$P(\max_{1\le k<p_{0j};j=1,\dots,q}|Z_{jk}|\ge d_n^{1/4}) \le \frac{C}{d_n^{1/4}}\exp\{-\frac{d_n^{1/2}}{2}+\ln q+\ln\max_j p_{0j}\}. \tag{9}$$

Inequality (5) follows from (7), (8) and (9).

To verify (6), note that

$$\begin{aligned}\|(I-H^{\mathrm{X}}(s_{*k+1}))\boldsymbol{y}_j\|_2^2 &= \Delta(s_{*k+1},\boldsymbol{\mu}_j)+\Delta(s_{*k+1},\boldsymbol{e}_j)+2\Delta(s_{*k+1},\boldsymbol{\mu}_j,\boldsymbol{e}_j)\\ &\le \Delta(s_{*k+1},\boldsymbol{\mu}_j)+\Delta(s_{*k+1},\boldsymbol{e}_j)+2\sqrt{\Delta(s_{*k+1},\boldsymbol{\mu}_j)\Delta(s_{*k+1},\boldsymbol{e}_j)}. \quad (10)\end{aligned}$$

It was shown in Luo and Chen (2014) that, for any j and $k<p_{0j}$, $\Delta(s^*_{jk+1},\mu)\le Cnp_{0j}^2$. Therefore

$$\max_{jk}\Delta(s^*_{jk+1},\mu)\le Cn\max\{p_{0j}^2\}. \quad (11)$$

Since $\Delta(s_{*k+1},\boldsymbol{\epsilon})\sim\chi^2_{n-k-1}$, we have, for any $\delta>0$,

$$\begin{aligned}&P\left(\max_{jk}\left|\frac{\Delta(s_{jk+1}*,\boldsymbol{e}_j)}{n-k-1}-1\right|\ge n^{-\delta/2}\right)\\ \le\ &\sum_{jk}P\left(\left|\frac{\Delta(s_{jk+1}*,\boldsymbol{e}_j)}{n-k-1}-1\right|\ge n^{-\delta/2}\right)\\ \le\ &\sum_{jk}\frac{n^\delta\,\mathrm{Var}(Z^2)}{n-k-1}\le q\max_j p_{0j}\frac{n^\delta\,\mathrm{Var}(Z^2)}{n-\max_j p_{0j}}<Cn^{-(1-\delta)}q\max_j p_{0j},\end{aligned}$$

where Z is a standard normal variable and C is a generic constant. Therefore,

$$\begin{aligned}&P\left(\max_{jk}\Delta(s_{jk+1}*,\boldsymbol{e}_j)<n(1+n^{-\delta/2})\right)\\ \ge\ &P\left(\max_{jk}\Delta(s_{jk+1}*,\boldsymbol{e}_j)<(n-k-1)(1+n^{-\delta/2})\right)\\ \ge\ &P\left(\cap_{jk}\left\{\left|\frac{\Delta(s_{jk+1}*,\boldsymbol{e}_j)}{n-k-1}-1\right|<n^{-\delta/2}\right\}\right)\\ =\ &P\left(\max_{jk}\left|\frac{\Delta(s_{jk+1}*,\boldsymbol{e}_j)}{n-k-1}-1\right|<n^{-\delta/2}\right)\\ \ge\ &1-Cn^{-(1-\delta)}q\max_j p_{0j}. \quad (12)\end{aligned}$$

Combining (9) and (12), we have that (6) holds with probability greater than $1-Cn^{-(1-\delta)}q\max_j p_{0j}$.

Let

$$r_{2n}=\frac{2C}{d_n^{1/4}}\exp\{-\frac{d_n^{1/2}}{2}+\ln q+\ln\max_j p_{0j}\}+Cn^{-(1-\delta)}q\max_j p_{0j}.$$

It follows from A1 and A3 that $r_{2n}=O(n^{-1+\omega})$ for some $\omega<1$. Lemma s2 is established. □

Proof of Lemma s3

Proof. It was shown in Luo and Chen (2013) that

$$P(\min_{s_j, |s_j| \le k_{nj}} \text{EBIC}(s_j) > \text{EBIC}(s_{0j})) \to 1,$$

for fixed j, where $k_{nj} = kp_{0j}$. However, Lemma 3 requires the uniform convergence of the above probability for all j. In order to establish the uniform convergence, we derive a convergence rate for the above probability in the following. Express

$$\begin{aligned} \text{EBIC}_\gamma(s_j) - \text{EBIC}_\gamma(s_{0j}) &= n \ln \frac{\Delta(s_j, \boldsymbol{y}_j)}{\Delta(s_{0j}, \boldsymbol{y}_j)} + (|s_j| - p_{0j})(\ln n + 2\gamma \ln p) \\ &\equiv T_1 + T_2. \end{aligned}$$

Note that $\Delta(s_{0j}, \boldsymbol{y}_j) = \Delta(s_{0j}, \boldsymbol{e}_j)$ which follows a χ^2-distribution with degrees of freedom $n - p_{0j}$ and can be written as $\sum_{i=1}^{n-p_{0n}} Z_i^2$ where Z_i's are i.i.d. standard normal variables. By Chebyshev inequality, for any small constant δ,

$$P\left(\left| \frac{1}{n - p_{0j}} \sum_{i=1}^{n-p_{0n}} Z_i^2 - 1 \right| > n^{-\delta/2} \right) \le \frac{2n^\delta}{n - p_{0j}}.$$

Therefore

$$P(\Delta(s_{0j}, \boldsymbol{e}_j) = (n - p_{0j})(1 + O(n^{-\delta/2}))) \ge 1 - \frac{2n^\delta}{n - p_{0j}} \equiv 1 - \omega_n^{[1]}. \tag{13}$$

In what follows, we consider separately two cases: $s_j \not\subset s_{0j}$ and $s_j \subset s_{0j}$.

Case I: $s_j \not\subset s_{0j}$. In this case, express

$$T_1 = n \ln \left(1 + \frac{\Delta(s_j, \boldsymbol{y}_j) - \Delta(s_{0j}, \boldsymbol{e}_j)}{\Delta(s_{0j}, \boldsymbol{e}_j)} \right).$$

Note that

$$\Delta(s_j, \boldsymbol{y}_j) - \Delta(s_{0j}, \boldsymbol{e}_j) = \Delta(s_j, \boldsymbol{\mu}_j) + 2\Delta(s_j, \boldsymbol{\mu}_j, \boldsymbol{e}_j) + \boldsymbol{e}_j^\top H^{\mathrm{X}}(s_{0j})\boldsymbol{e}_j - \boldsymbol{e}_j^\top H^{\mathrm{X}}(s_j)\boldsymbol{e}_j.$$

First, by Chebyshev inequality we have

$$P(\boldsymbol{e}_j^\top H^{\mathrm{X}}(s_{0j})\boldsymbol{e}_j \le \Delta(s_j, \boldsymbol{\mu}_j)) \ge 1 - \frac{1}{(\ln p)\ D_{nj}}, \quad \text{(I)}$$

where $D_{nj} = \Delta(s_j, \boldsymbol{\mu}_j)/(p_{0j} \ln p) \to \infty$ by condition A4.

Next, we are going to show, for some constant C,

$$P(\max_{s_j, |s_j| \le k_{nj}} \boldsymbol{e}_j^\top H^{\mathrm{X}}(s_j)\boldsymbol{e}_j \le C k_{nj} \ln p) > 1 - \omega_{1n}, \quad \text{(II)}$$

and

$$P\left(|\max_{s_j, |s_j| \le k_{nj}} \Delta(s_j, \boldsymbol{\mu}_j, \boldsymbol{e}_j)| \le \sqrt{\Delta(s_j, \boldsymbol{\mu}_j) k_{nj} \ln p)} \right) > 1 - \omega_{2n}, \quad \text{(III)}$$

where ω_{1n} and ω_{2n} will be given later. Then, condition A4, (I), (II) and (III) imply that

$$P\left(\Delta(s_j, \boldsymbol{y}_j) - \Delta(s_{0j}, \boldsymbol{e}_j) = C\Delta(s_j, \boldsymbol{\mu}_j), \text{for all } s_j : |s_j| \le k_{nj} \right) > 1 - \omega_n^{[2]}, \tag{14}$$

where $\omega_n^{[2]} = \omega_{1n} + \omega_{2n} + \frac{1}{(\ln p) \ D_n}$, $D_n = \min_j D_{nj}$. It then follows from (13) and (14) that

$$T_1 = n \ln \left(1 + \frac{C\Delta(s_j, \boldsymbol{\mu}_j)}{n} \right), \tag{15}$$

uniformly for all s_j such that $|s_j| \le k_{nj}$ with probability greater than $1 - \omega_n$ where $\omega_n = \omega_n^{[1]} + \omega_n^{[2]}$.

Inequalities (II) and (III) are verified in the following. Let $d > 1$ be a constant and $m_j = 2dk_{nj}[\ln p + \ln(dk_{nj} \ln p)]$. Since $\boldsymbol{e}_j^\top H(s_j)\boldsymbol{e}_j$ follows a χ^2-distribution with degrees of freedom $l = |s_j|$. Let $\mathcal{S}_l$ be the class of models consisting of exactly l covariates. Denote by $\tau(\mathcal{S}_l)$ the size of $\mathcal{S}_l$, i.e., $\tau(\mathcal{S}_l) = \binom{p}{l}$. By the Bonferroni inequality, we have

$$\begin{aligned} &P(\max_{s_j, |s_j| \le k_{nj}} \boldsymbol{e}_j^\top H^{\mathrm{X}}(s_j)\boldsymbol{e}_j \ge m_j) \\ =&P(\max\{\chi_l^2(s) : s \in \mathcal{S}_l, l \le k_{nj}\} \ge m_j) \le \sum_{l=1}^{k_{nj}} \tau(\mathcal{S}_l) P(\chi_l^2 \ge m_j). \end{aligned}$$

There is a constant C close to 1, not depending on j for $j \le k_{nj}$, such that

$$\begin{aligned} \tau(\mathcal{S}_l) P(\chi_l^2 \ge m_j) \approx& \frac{C}{2^{l/2-1}\Gamma(l/2)} \frac{\tau(\mathcal{S}_l)}{p_n^{dk_{nj}}} (dk_{nj} \ln p_n)^{-k_{nj}} m_j^{l/2-1} \\ \le& \frac{C}{m_j p^{d-1}} (k_{nj} \ln p)^{-l} m_j^{l/2} = \frac{C}{m_j p^{d-1}} \left[\sqrt{\frac{m_j}{(dk_{nj} \ln p)^2}} \right]^l = \frac{C}{m_j p^{d-1}} \rho_n^l, \text{ say,} \end{aligned}$$

where

$$\rho_n = \sqrt{\frac{m_j}{(dk_{nj} \ln p)^2}} = \sqrt{\frac{2d[k_{nj} \ln p + k_{nj} \ln(dk_{nj} \ln p)]}{(dk_{nj} \ln p)^2}} (1 + o(1)) \le \rho,$$

for some ρ between 0 and 1, when n is large enough, since $\rho_n \to 0$. Thus

$$P\left(\max_{s_j, |s_j| \le k_{nj}} \boldsymbol{e}_j^\top H^{\mathrm{X}}(s_j)\boldsymbol{e}_j \ge m_j\right) \le \frac{C}{m_j p^{d-1}} \sum_{l=1}^{k_{nj}} \rho^l \le \frac{C}{m_j p^{d-1}} \frac{\rho}{1-\rho}; \tag{16}$$

that is,

$$P\left(\max_{s_j, |s_j| \le k_n} \boldsymbol{\epsilon}_j^\top H^{\mathrm{X}}(s_j)\boldsymbol{\epsilon}_j \le C k_n \ln p\right) > 1 - \frac{C}{m_j p^{d-1}},$$

which establishes (II) with $\omega_{1n} = \frac{C}{p_{0j} p^{d-1} \ln p}$.

Furthermore, we can express

$$\Delta(s_j, \boldsymbol{\mu}_j, \boldsymbol{e}_j) = \sqrt{\Delta(s_j, \boldsymbol{\mu}_j)} Z(s_j),$$

where $Z(s_j) \sim N(0,1)$. For any s_j with $|s_j| \le k_{nj}$,

$$|\Delta(s_j, \boldsymbol{\mu}_j, \boldsymbol{e}_j)| \le \sqrt{\Delta(s_j, \boldsymbol{\mu}_j)} \max\{|Z(s)| : |s| \le k_{nj}\}.$$

Let m_j be the same as above. We have

$$\begin{aligned} P(\max\{|Z(s)| : |s| \le k_{nj}\} \ge \sqrt{m_j}) =& P(\max\{|Z(s)| : s \in \mathcal{S}_l, l \le k_{nj}\} \ge \sqrt{m_j}) \\ \le& \sum_{l=1}^{k_{nj}} \tau(\mathcal{S}_l) P(Z(s) \ge \sqrt{m_j}) = \sum_{l=1}^{k_{nj}} \tau(\mathcal{S}_l) P(\chi_1^2 \ge m_j) \\ \le& \sum_{l=1}^{k_{nj}} \tau(\mathcal{S}_l) P(\chi_l^2 \ge m_j). \end{aligned}$$

Thus, (III) follows with $\omega_{2n} = \omega_{1n}$. Eventually, we have that, with probability greater than $1 - \omega_n$, for all s_j such that $|s_j| \le k_{nj}$,

$$\begin{aligned} & \mathrm{EBIC}_\gamma(s) - \mathrm{EBIC}_\gamma(s_{0j}) \\ =& n \ln\left(1 + \frac{C\Delta(s_j, \mu_j)}{n}\right) + (|s_j| - p_{0j})(\ln n + 2\gamma \ln p) \\ \ge& n \ln[1 + C p_{0j} \ln p / n] - p_{0j}(\ln n + 2\gamma \ln p), \end{aligned}$$

for an arbitrarily large C, when n is large enough, by condition A4. Then by choosing $C > 1 + 2\gamma$, the difference goes to infinity as $n \to \infty$.

Case II: $s_{0j} \subset s_j$. In this case, $\Delta(s_j, \boldsymbol{y}_j) = \Delta(s_j, \boldsymbol{e}_j)$ and

$$\Delta(s_{0j}, \boldsymbol{e}_j) - \Delta(s_j, \boldsymbol{e}_j) = \boldsymbol{e}_j^T \{H^{\mathrm{X}}(s_j) - H^{\mathrm{X}}(s_{0j})\}\boldsymbol{e}_j \equiv \chi_l^2(s_j),$$

where $\chi^2_l(s_j)$ is a χ^2 random variable depending on s_j with degrees of freedom $l = |s_j| - p_{0j}$. We express

$$-T_1 = n\log\left(\frac{\Delta(s_{0j},\boldsymbol{e}_j)}{\Delta(s_j,\boldsymbol{e}_j)}\right) = n\log\left\{1+\frac{\chi^2_l(s)}{\Delta(s_{0j},\boldsymbol{e}_j)-\chi^2_l(s)}\right\} \le \frac{n\chi^2_j(s)}{\Delta(s_{0j},\boldsymbol{e}_j)-\chi^2_j(s)}.$$

Let $\tilde{\mathcal{S}}_j = \{s : s \in \mathcal{S}_{j+p_{0j}}, s_0 \subset s\}$ and $m_l = 2dl[\ln p + \ln(dl\ln p)]$. By a similar argument leading to (16), we have

$$P\left(\max_{1\le l\le k_{nj}-p_{0j}} \frac{\max\{\chi^2_l(s) : s\in\tilde{\mathcal{S}}_l\}}{m_l} \ge 1\right) \le \frac{C}{p^{d-1}\ln p}. \tag{17}$$

Combining (13) and (17), we have, with probability greater than $1 - C[n^{-(1-\delta)} + 1/(p^{d-1}\ln p)]$,

$$-T_1 \le \frac{nm_l}{n-m_l} \le m_l(1+o(1)) = 2d(|s_j| - p_{0j})\ln p(1+o(1)),$$

uniformly for all s_j such that $|s_j| \le k_{nj}$ and $s_{0j} \subset s_j$. Thus

$$T_1 \ge -2d(|s_j| - p_{0j})\ln p(1+o(1)).$$

Finally we have

$$\begin{aligned}&\mathrm{EBIC}_\gamma(s_j) - \mathrm{EBIC}_\gamma(s_{0j})\\ \ge&(|s_j| - p_{0j})[\ln n + 2\gamma\ln p] - 2d(|s_j| - p_{0j})\ln p(1+o(1)) > 0,\end{aligned}$$

uniformly for all s_j with $|s_j| \le k_n$ and $s_{0j} \subset s_j$, if n is big enough, when $\gamma > d - \frac{\ln n}{2\ln p}$.

Let $r_n = \omega_n + C[n^{-(1-\delta)} + 1/(p^{d-1}\ln p)]$. Summarizing Case I and II, we conclude that

$$P(\min_{s_j,|s_j|\le kp_{0j}} \mathrm{EBIC}(s_j) > \mathrm{EBIC}(s_{0j})) \ge 1 - r_n.$$

It is easy to see that r_n is of the order $1/(R_n\ln p)$ where $R_n \to \infty$. Then, by Bonferroni inequality, Lemma s3 holds with $r_{3n} = \frac{q}{R_n\ln p}$. □

Proof of Lemma s4

Proof. For convenience, denote $\mathcal{B}_{j^-}$ by $\mathcal{B}_j$ and $(Y_{j^-} - X\mathcal{B}_{j^-})$ by $Z(\mathcal{B}_j)$. Thus $\tilde{\boldsymbol{y}}_j = \boldsymbol{y}_j - Z(\mathcal{B}_j)\boldsymbol{\xi}_j$. Let $\hat{\boldsymbol{y}}_j = \boldsymbol{y}_j - Z(\hat{\mathcal{B}}_j)\hat{\boldsymbol{\xi}}_j$, where $\hat{\mathcal{B}}_j$ and $\hat{\boldsymbol{\xi}}_j$ are the estimates from the previous iteration. Now write

$$\hat{\boldsymbol{y}}_j = \boldsymbol{y}_j - Z(\mathcal{B}_j)\boldsymbol{\xi}_j + Z(\mathcal{B}_j)\boldsymbol{\xi}_j - Z(\hat{\mathcal{B}}_j)\hat{\boldsymbol{\xi}}_j = \tilde{\boldsymbol{y}} - Z(\hat{\mathcal{B}}_j)(\hat{\boldsymbol{\xi}}_j - \boldsymbol{\xi}_j) + X(\hat{\mathcal{B}}_j - \mathcal{B}_j)\boldsymbol{\xi}_j.$$

By an abuse of notation, denote $\frac{1}{n}\boldsymbol{x}_l^\top[I_n - H^{\mathrm{X}}(s_{jk}^*)]\hat{\boldsymbol{y}}_j$ still by $\hat{\gamma}_j(l, s_{jk}^*)$. Decompose $\hat{\gamma}_j(l, s_{jk}^*)$ as follows.

$$\begin{aligned}\hat{\gamma}_j(l, s_{jk}^*) &= \gamma_j(l, s_{jk}^*) + \frac{1}{n}\boldsymbol{x}_l^\top[I_n - H^{\mathrm{X}}(s_{jk}^*)]\boldsymbol{\epsilon}_j + \frac{1}{n}\boldsymbol{x}_l^\top[I_n - H^{\mathrm{X}}(s_{jk}^*)]Z(\hat{\mathcal{B}}_j)(\hat{\boldsymbol{\xi}}_j - \boldsymbol{\xi}_j) \\ &\quad + \frac{1}{n}\boldsymbol{x}_l^\top[I_n - H^{\mathrm{X}}(s_{jk}^*)]X(\hat{\mathcal{B}}_j - \mathcal{B}_j)\boldsymbol{\xi}_j. \end{aligned} \tag{18}$$

We only need to show that $\gamma_j(l, s_{jk}^*)$ dominates all the other components of $\hat{\gamma}_j(l, s_{jk}^*)$. Replacing $\boldsymbol{e}_j$ in the proof of Lemma 1 by $\boldsymbol{\epsilon}_j$, we have

$$P\left(\max_{l\in s_{jk}^{*c}} \frac{1}{n}\boldsymbol{x}_l^\top[I - H^{\mathrm{X}}(s_{jk}^*)]\boldsymbol{\epsilon}_j \geq C_{jn}^{1/2} n^{-1/2}\ln p\right) \leq \frac{2\tau_j}{C_{jn}^{1/2}\ln p}\exp\left\{-\frac{(\ln p)^2}{2} + \ln p\right\}. \tag{19}$$

We still have

$$\max_{l\in s_{jk}^{*-}} |\gamma_j(l, s_{jk}^*)| \geq C_{jn} n^{-1/2}\ln p,$$

where $C_{jn} \to \infty$. It suffices to verify that the third and fourth component in (18) are less than $Cn^{-1/2}\ln p$ for some constant C. Let the the third and fourth component be denoted respectively by U_{1n} and U_{2n}. Let $H^{\mathrm{Z}}(\hat{\mathcal{B}}_j)$ and $H(X)$ be the projection matrices of $Z(\hat{\mathcal{B}}_j)$ and X respectively.

Let $v_1^2 = \boldsymbol{x}_l^\top[I - H^{\mathrm{X}}(s_{jk}^*)]H^{\mathrm{Z}}(\hat{\mathcal{B}}_j)[I - H^{\mathrm{X}}(s_{jk}^*)]\boldsymbol{x}_l$ and $Z_n = \frac{1}{v_1}\boldsymbol{x}_l^\top[I_n - H^{\mathrm{X}}(s_{jk}^*)]Z(\hat{\mathcal{B}}_j)(\hat{\boldsymbol{\xi}}_j - \boldsymbol{\xi}_j)$. We can express $U_{1n} = \frac{v_1}{n}Z_n$. By Theorem 4 of Luo and Chen (2014) and Sluskey's theorem, Z_n has an asymptotic normal distribution with mean zero and variance τ_j^2, and $Z_n/\ln p$ converges to zero in probability. Note that $v_1^2 \leq \boldsymbol{x}_l^\top\boldsymbol{x}_l = n$. Thus, there is constant C such that

$$|U_{1n}| = \frac{v_1}{n}\ln p(|Z_n|/\ln p) \leq C\ln p/\sqrt{n}. \tag{20}$$

Let $v_2^2 = \boldsymbol{x}_l^\top[I - H^{\mathrm{X}}(s_{jk}^*)]H(X)[I - H^{\mathrm{X}}(s_{jk}^*)]\boldsymbol{x}_l$ and $Z_{ni} = \frac{1}{v_2}\boldsymbol{x}_l^\top[I - H^{\mathrm{X}}(s_{jk}^*)]X(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)$. Similarly, Z_{ni} has an asymptotic normal distribution with mean zero and variance σ_i^2 and $v_2^2 \leq n$. We then have

$$|U_{2n}| = \frac{v_2}{n}\left|\sum_{1\leq i\leq q, i\neq j}\xi_{jl}Z_{ni}\right| \leq \frac{\ln p}{q\sqrt{n}}\sum_{1\leq i\leq q, i\neq j}|\xi_{ji}|\left|\frac{q}{\ln p}Z_{ni}\right| \leq C\ln p/\sqrt{n}, \tag{21}$$

for some constant C. Note that $q/\ln p \to 0$. Combining (19) – (21), we have that the probability in Lemma s4 converges to 1. It should be noted that (20) and (21) hold on a set with probability converging to 1 since $\hat{\mathcal{B}}_j$ and $\hat{\boldsymbol{\xi}}_j$ are estimates from stage 1. This affects the convergence rate of the probability in Lemma s4 but not the convergence. □

Theorem 2. Assume conditions $B1 - B4$. Let $\hat{\mathcal{T}}_0$ be the index set of the identified nonzero entries of Ω in a Ω-step. Then

$$P(\hat{\mathcal{T}}_0 = \mathcal{T}_0) \to 1, \quad \text{as } n \to \infty.$$

To show Theorem 2, we need Lemma 1 which is re-stated below.

Lemma 1. Assume that the correlations $\sigma_{jk}/(\sigma_j \sigma_k)$ are bounded by a constant less than 1, and the variances σ_j^2 are bounded. Then

$$P(\max_{1 \le j,k \le q} |\frac{1}{n}\boldsymbol{z}_j^\top \boldsymbol{z}_k - \sigma_{jk}| > n^{-1/3} c_0) \to 0,$$

where c_0 is a fixed constant.

Once Lemma 1 is established, by following the same arguments in Jiang and Chen (2016), conditions $B2 - B4$ can be transfered into empirical versions in terms of Z similar to $A2 - A4$. The remainder of the proof will be exactly the same as that in Jiang and Chen (2016), and hence will be omitted. Therefore, we only give the proof of lemma 1 in the following.

Proof of Lemma 1

Proof. Recall that $\boldsymbol{z}_j = [I - H^X(\hat{s}_{0j})]\boldsymbol{y}_j$. By theorem 1, $P(\hat{s}_{0j} = s_{0j}, j = 1, \ldots, q) \to 1$. Therefore, in the proof, we can replace $\hat{s}_{0j}$ with s_{0j}. Thus, we have

$$\boldsymbol{z}_j = [I - H^X(s_{0j})]\boldsymbol{y}_j = [I - H^X(s_{0j})][X(s_{0j})\boldsymbol{\beta}_j(s_{0j}) + \boldsymbol{\epsilon}_j] = [I - H^X(s_{0j})]\boldsymbol{\epsilon}_j,$$

and

$$\begin{aligned} \frac{1}{n}\boldsymbol{z}_j^\top \boldsymbol{z}_k &= \frac{1}{n}\boldsymbol{\epsilon}_j^\top [I - H^X(s_{0j})][I - H^X(s_{0k})]\boldsymbol{\epsilon}_k \\ &= \frac{1}{n}\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_k - \frac{1}{n}\boldsymbol{\epsilon}_j^\top H^X(s_{0k})\boldsymbol{\epsilon}_k - \frac{1}{n}\boldsymbol{\epsilon}_j^\top H^X(s_{0j})\boldsymbol{\epsilon}_k + \frac{1}{n}\boldsymbol{\epsilon}_j^\top H^X(s_{0j})H^X(s_{0k})\boldsymbol{\epsilon}_k. \end{aligned}$$

By Chebyshev's inequality,

$$P(|\frac{1}{n}\boldsymbol{\epsilon}_j^\top H^X(s_{0j})\boldsymbol{\epsilon}_k| > n^{-1/3}) \le n^{1/3} E\frac{1}{n}\boldsymbol{\epsilon}_j^\top H^X(s_{0j})\boldsymbol{\epsilon}_k = n^{1/3} p_{0j}\sigma_{jk}/n \le n^{-2/3}\max_j p_{0j}\max_{j,k}\sigma_j\sigma_k.$$

Similarly, we can show

$$P(|\frac{1}{n}\boldsymbol{\epsilon}_j^\top H^X(s_{0j})\boldsymbol{\epsilon}_k| > n^{-1/3}) \le n^{-2/3}\max_j p_{0j}\max_{j,k}\sigma_j\sigma_k;$$
$$P(|\frac{1}{n}\boldsymbol{\epsilon}_j^\top H^X(s_{0j})H^X(s_{0k})\boldsymbol{\epsilon}_k| > n^{-1/3}) \le n^{-2/3}\max_j p_{0j}\max_{j,k}\sigma_j\sigma_k.$$

A trivial argument yields that

$$P(\frac{1}{n}|\boldsymbol{\epsilon}_j^\top H^X(s_{0k})\boldsymbol{\epsilon}_k + \boldsymbol{\epsilon}_j^\top H^X(s_{0j})\boldsymbol{\epsilon}_k - \boldsymbol{\epsilon}_j^\top H^X(s_{0j})H^X(s_{0k})\boldsymbol{\epsilon}_k| > 3n^{-1/3}) \le cn^{-2/3}\max_j p_{0j},$$

for a generic constant c. By Bonferroni's inequality,

$$P(\max_{ij}\frac{1}{n}|\boldsymbol{\epsilon}_j^\top H^X(s_{0k})\boldsymbol{\epsilon}_k + \boldsymbol{\epsilon}_j^\top H^X(s_{0j})\boldsymbol{\epsilon}_k - \boldsymbol{\epsilon}_j^\top H^X(s_{0j})H^X(s_{0k})\boldsymbol{\epsilon}_k| > 3n^{-1/3}) \le cq^2 n^{-2/3}\max_j p_{0j}.$$

If we take $q = O(n^c)$ with $c \le 1/4$, the above probability converges to zero. Thus, we have

$$\max_{1\le j,k\le q}|\frac{1}{n}\boldsymbol{z}_j^\top \boldsymbol{z}_k - \sigma_{jk}| = \max_{1\le j,k\le q}|\frac{1}{n}\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_k - \sigma_{jk}| + O_p(n^{-1/3}).$$

Hence

$$P(\max_{1\le j,k\le q}|\frac{1}{n}\boldsymbol{z}_j^\top \boldsymbol{z}_k - \sigma_{jk}| > n^{-1/3}c_0) = P(\max_{1\le j,k\le q}|\frac{1}{n}\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_k - \sigma_{jk}| > n^{-1/3}c),$$

for a generic constant c. If c is chosen such that $c \ge \sigma_{\text{MAX}}$ where $\sigma_{\text{MAX}} = \max_{j,k}\sqrt{\text{Var}(\epsilon_j\epsilon_k)}$, the lemma follows from Lemma 1 of Luo and Chen (2014). □

2 Miscellaneous materials

2.1 Some details for simulation study I

Types of design matrix X and s_{0j}'s:

- *Type I.* The rows of X are generated as i.i.d. observations from $N_p(\mathbf{0}, I)$. Each s_{0j} is taken as a random sample of size p_0 from $\{1, \ldots, p\}$.

- *Type II.* The rows of X are generated as i.i.d. observations from $N_p(\mathbf{0}, \Sigma_X)$, where the (i,j)th entry of Σ_X is $0.5^{|i-j|}$, i.e., Σ_X is a correlation matrix of power decay with $\rho = 0.5$. Each s_{0j} is generated as $(j^*, j^*+1, \ldots, j^*+p_0-1)$ where j^* is chosen at random from $\{1, 2, \ldots, p-p_0+1\}$.

- *Type III.* The rows of X are generated as i.i.d. observations of the random vector $(\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top)^\top$, where $\boldsymbol{X}_1$ is of dimension $p_1 = [\frac{p}{2}]$ and distributed as $N_{p_1}(\mathbf{0}, I)$ and $\boldsymbol{X}_2$ is distributed as $N_{p_2}(\mathbf{0}, \Sigma_{X_2})$, where $p_2 = p - p_1$ and Σ_{X_2} is of power decay with $\rho = 0.5$. Each s_{0j} consists of $p_{01} = [\frac{p_0}{2}]$ random numbers from $\{1, \ldots, p_1\}$ and a random segment of length $p_0 - p_{01}$ from $\{p_1 + 1, \ldots, p\}$.

- *Type IV.* The rows of X are generated as i.i.d. observations of the random vector $(\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top)^\top$, where $\boldsymbol{X}_1$ is distributed as $N_{p_{01}}(\mathbf{0}, \Sigma_{X_1})$, Σ_{X_1} being of power decay with $\rho = 0.5$, and the components of $\boldsymbol{X}_2$ are given by

  $$X_{2j} = t_j + \frac{\sum_{k=1}^{p_{01}} X_k}{p_{01}},$$

  where t_j's are i.i.d. $N(0, 0.08)$. Each s_{0j} consists of $\{1, \ldots, p_{01}\}$ and a random set of size φ_j from $\{p_{01} + 1, \ldots, p\}$, where φ_j is chosen at random from $\{1, \ldots, p_{01}\}$.

Generation of $\boldsymbol{\beta}_j$'s

For each j, $\boldsymbol{\beta}_j$ is generated as follows. Let u follow Bernoulli(0.4), z be a normal random variable with mean 0 and satisfying that $P(|z| \geq 0.1) = 0.25$. The components of $\boldsymbol{\beta}_j$ are first independently generated from the random variable $(-1)^u(4n^{-0.15} + |z|)$ and then scaled such that $\frac{\boldsymbol{\beta}_j^\top \Sigma_X \boldsymbol{\beta}_j}{\boldsymbol{\beta}_j^\top \Sigma_X \boldsymbol{\beta}_j + \sigma_j^2} = h$ for a fixed h, where σ_j^2 is the variance of the Y_j. The number h determines the proportion of the variation of Y_j attributable to the covariates.

2.2 Network graph of the real example

Figure 1 visualizes the network graph of the 20 microRNAs detected by TASCS. There are 22 edges in the graph, that is, the TASCS identified 44 non-zero entries of the precision matrix Ω.

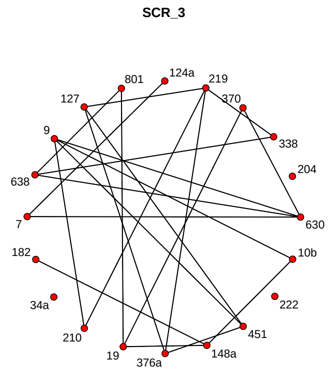

Figure 1: The graphical networks of the twenty selected microRNAs based on sparsity of the estimated precision matrix

It is worth noting that two microRNAs, ”hsa.miR.136” and ”hsa.miR.377”, are included in the previous analyses; however, they are excluded in our study due to their small MAD values and are replaced by another two microRNAs, ”hsa.miR.9” and ”hsa.miR.127”. Except the edges connected to ”hsa.miR.9” and ”hsa.miR.127”, most of the rest edges detected by TASCS are also detected by DML and aMCR. Moreover, the graph in Figure 1 has less edges then those detected by DML and aMCR.

References

Jiang, Y. and Z. Chen (2016). A sequential scaled pairwise selection approach to edge detection in nonparanormal graphical models. *Canadian Journal of Statistics 44*(1), 25–43.

Luo, S. and Z. Chen (2013). Extended BIC for linear regression models with diverging number of relevant features and high or ultra-high feature spaces. *Journal of Statistical Planning and Inference 143*(3), 494–504.

Luo, S. and Z. Chen (2014). Sequential Lasso cum EBIC for feature selection with ultra-high dimensional feature space. *Journal of the American Statistical Association 109*(507), 1229–1240.