Supplementary material for ‘Estimating graphical models for count data with applications to single-cell gene network’

Feiyi Xiao *
School of Mathematical Sciences
Peking University
xiaofeiyi1217@pku.edu.cn

Junjie Tang *
School of Mathematical Sciences
Peking University
junjie.tang@pku.edu.cn

Huaying Fang
Beijing Advanced Innovation Center for Imaging Theory and Technology
Capital Normal University
huayingfang@hotmail.com

Ruibin Xi
School of Mathematical Sciences
Center for Statistical Science
Peking University
ruibinxi@math.pku.edu.cn

S1 Theoretical results and proofs for moment estimator

S1.1 A moment based estimator

We utilize the moment estimator of the latent covariance matrix as the initial value for MMLE updating. Let $\boldsymbol{\mu}^* = [\mu_i^*]_{1\le i\le p}$ be the mean vector, $\Sigma^* = \left[\sigma_{ij}^*\right]_{1\le i,j\le p}$ be the covariance matrix and $\alpha_i^* = \exp(\mu_i^* + \sigma_{ii}^*/2)$ for $1 \le i \le p$. From the first two moments of the PLN distribution, we have

$$\begin{aligned} E\left(Y_{ij}/S_i\right) &= \alpha_j^*, \\ E\left(\left(Y_{ij}^2 - Y_{ij}\right)/S_i^2\right) &= (\alpha_j^*)^2 \exp\left(\sigma_{jj}^*\right), \\ E\left(Y_{ij}Y_{ik}/S_i^2\right) &= \alpha_j^*\alpha_k^* \exp\left(\sigma_{jk}^*\right), \end{aligned} \tag{S1}$$

where $1 \le i \le n$ and $1 \le j \ne k \le p$. Let $\tilde{\alpha}_j = n^{-1}\sum_{i=1}^n \left(Y_{ij}/S_i\right)$ for $1 \le j \le p$. Then, a candidate moment estimator $\widetilde{\Sigma}^m = \left[\tilde{\sigma}_{ij}^m\right]_{1\le i,j\le p}$ for the covariance matrix is

$$\tilde{\sigma}_{jk}^m = \begin{cases} \log\left(n^{-1}\sum_{i=1}^n \left[\left\{Y_{ij}\left(Y_{ij}-1\right)\right\}/S_i^2\right]\right) - 2\log\left(\tilde{\alpha}_j\right), & \text{for } 1 \le j = k \le p, \\ \log\left[n^{-1}\sum_{i=1}^n \left\{\left(Y_{ij}Y_{ik}\right)/S_i^2\right\}\right] - \left\{\log\left(\tilde{\alpha}_j\right) + \log\left(\tilde{\alpha}_k\right)\right\}, & \text{for } 1 \le j \ne k \le p. \end{cases} \tag{S2}$$

Also, we can get a moment estimator $\tilde{\boldsymbol{\mu}}^m = [\tilde{\mu}_i^m]_{1\le i\le p}$ with $\tilde{\mu}_i^m = \log(\tilde{\alpha}_i) - \tilde{\sigma}_{ii}/2$ for the mean vector $\boldsymbol{\mu}^*$. Further, similar to the equations (2), (3) in the section 2.3 of the manuscript, we can get the positive definite $\widehat{\Sigma}^m$ from initial value $\widetilde{\Sigma}^m$. Plugging in $\widehat{\Sigma}^m$ to the penalized D-trace loss, we can get a moment based estimator called PLNet-MOM as follows,

$$\widehat{\Theta}^m = \arg\min_{\Theta\succeq 0} \frac{1}{2}\operatorname{tr}\left(\widehat{\Sigma}^m\Theta^2\right) - \operatorname{tr}\left(\Theta\right) + \lambda_n \left\|\Theta\right\|_{1,\text{off}}. \tag{S3}$$

We show that PLNet-MOM $\widehat{\Theta}^m$ is also consistent for Θ^* in the next subsection.

*Equal contribution. Corresponding author: Ruibin Xi (ruibinxi@math.pku.edu.cn)

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

S1.2 Main results

Theorem S1 (Rate of convergence for $\widehat{\Sigma}^m$) *Under the boundedness condition in the manuscript, for any positive integer m and $0 < \epsilon < 6$, there exist constants C_1 depending only on m, such that* $pr\left(\left\|\widehat{\Sigma}^m - \Sigma^*\right\|_\infty > \epsilon\right) < p^2/\left(C_1 n^m \epsilon^{2m}\right)$.

Theorem S2 (Rate of convergence for $\widehat{\Theta}^m$) *Under the irrepresentability condition and the boundedness condition in the mamuscript, for any positive integer m, there exists a constant C_1 that only depends on m, such that for some $\eta > 2$, if*

$$n > \quad C_1^{-1/m} p^{\eta/m} \max\Bigg[12dk_\Gamma,\ 12\gamma^{-1}(k_\Sigma k_\Gamma^2 + k_\Gamma),\ \left\{12\gamma^{-1}\left(k_\Sigma k_\Gamma^3 + k_\Gamma^2\right) + 5dk_\Gamma^2\right\}\theta_{\min}^{-1},$$
$$\min\left\{s^{1/2}, d+1\right\}\left\{12\gamma^{-1}\left(k_\Sigma k_\Gamma^3 + k_\Gamma^2\right) + 5dk_\Gamma^2\right\}\lambda_{\min}^{-1}(\Theta^*), 1/5\Bigg]^2,$$

and

$$\lambda = 12\gamma^{-1}\left(k_\Sigma k_\Gamma^2 + k_\Gamma\right) C_1^{-1/(2m)} p^{\eta/(2m)} n^{-1/2},$$

then with probability $1 - p^{2-\eta}$,

$$\left\|\widehat{\Theta}^m - \Theta^*\right\|_\infty \le \left(12\gamma^{-1}\left(k_\Sigma k_\Gamma^3 + k_\Gamma^2\right) + 5dk_\Gamma^2\right) C_1^{-1/(2m)} p^{\eta/(2m)} n^{-1/2},$$
$$\left\|\widehat{\Theta}^m - \Theta^*\right\|_F \le s^{1/2}\left(12\gamma^{-1}\left(k_\Sigma k_\Gamma^3 + k_\Gamma^2\right) + 5dk_\Gamma^2\right) C_1^{-1/(2m)} p^{\eta/(2m)} n^{-1/2},$$
$$\left\|\widehat{\Theta}^m - \Theta^*\right\|_2 \le \min\left\{s^{1/2}, d+1\right\}\left(12\gamma^{-1}\left(k_\Sigma k_\Gamma^3 + k_\Gamma^2\right) + 5dk_\Gamma^2\right) C_1^{-1/(2m)} p^{\eta/(2m)} n^{-1/2}.$$

Theorem S3 (Sign consistency for $\widehat{\Theta}^m$) *Under all the conditions in Theorem S2, for some $\eta > 2$, choosing the same n and λ in Theorem S2, then with probability $1 - p^{2-\eta}$, $\widehat{\Theta}^m$ recovers all zeros and nonzeros in Θ^*.*

S1.3 Proofs

S1.3.1 Lemmas

For any matrix $\Sigma = [\sigma_{ij}]_{1\le i,j\le d}$, vec$(\Sigma)$ is defined as

$$\mathrm{vec}(\Sigma) = (\sigma_{11}, \sigma_{21}, \ldots, \sigma_{d1}, \sigma_{12}, \sigma_{22}, \ldots, \sigma_{d2}, \ldots, \sigma_{1d}, \sigma_{2d}, \ldots, \sigma_{dd})^T,$$

Lemma S1 *We define*

$$\breve{\Theta}^m = \arg\min_{A=A^T} \frac{1}{2}\mathrm{tr}\left(\widehat{\Sigma}^m A^2\right) - \mathrm{tr}\left(A\right) + \lambda\left\|A\right\|_{1,\mathit{off}}.$$

Then the following propositions hold:

(a) vec $\left(\breve{\Theta}^m\right)_{G^c} = 0$, *if*

$$\|\widehat{\Sigma}^m - \Sigma^*\|_\infty < 1/\left(12dk_\Gamma\right),$$
$$6\|\widehat{\Sigma}^m - \Sigma^*\|_\infty\left(k_\Sigma k_\Gamma^2 + k_\Gamma\right) \le 0.5\gamma\min\left\{\lambda, 1\right\};$$

(b) assuming the conditions in part (a), we also have

$$\left\|\breve{\Theta}^m - \Theta^*\right\|_\infty < \lambda k_\Gamma + \frac{5}{2}d\left(1+\lambda\right)\|\widehat{\Sigma}^m - \Sigma^*\|_\infty k_\Gamma^2.$$

Lemma S1 is the Lemma A1 (b) and (c) in D-trace method [5].

Lemma S2 *Under the boundedness condition in the manuscript, for any positive integer m, there exists a $k_m > 0$ such that*

$$E(Y_{ij}^m) \leq k_m.$$

Lemma S3 *m is a positive integer. Let $\{W_i, 1 \leq i \leq n\}$ be a series of independent random variables with $E(W_i) = 0$ and $E\left(W_i^k\right) \leq u_k$ for all $1 \leq i \leq n, 1 \leq k \leq 2m$. Then, there exists a constant v_m only depending on m, such that*

$$pr\left(\left|n^{-1}\sum_{i=1}^{n} W_i\right| > \epsilon\right) \leq v_m / \left(n^m \epsilon^{2m}\right).$$

Lemma S4 *Under the boundedness condition in the manuscript, for any positive integer m and $0 < \epsilon < 3$, there exists a constant C_0 depending only on m such that for $1 \leq j, k \leq p$,*

$$pr\left(\left|\widetilde{\sigma}_{jk}^m - \sigma_{jk}^*\right| > \epsilon\right) \leq 1/\left(C_0 n^m \epsilon^{2m}\right).$$

S1.3.2 Proof of Lemma S2

Let $C_{k,m}$ be the Stirling numbers of the second kind. From the moment results of Poisson distribution [1], we have

$$E\left(Y_{ij}^m | X_{ij}\right) = \sum_{k=0}^{m} \left(S_i X_{ij}\right)^k C_{k,m} \leq \sum_{k=0}^{m} C^k X_{ij}^k C_{k,m}.$$

From the moment generating function of the normal distribution, we have

$$E\left(X_{ij}^m\right) = \exp\left(\frac{1}{2}m^2\sigma_{jj} + m\mu_j\right) \leq \exp\left(\frac{1}{2}m^2 C + mC\right).$$

Combining the above two inequalities, we have

$$\begin{aligned} E\left(Y_{ij}^m\right) &= E\left(E\left(Y_{ij}^m | X_{ij}\right)\right) \leq E\left(\sum_{k=0}^{m} C^k X_{ij}^k C_{k,m}\right) \\ &= \sum_{k=0}^{m} C^k E\left(X_{ij}^k\right) C_{k,m} \leq \sum_{k=0}^{m} C^k \exp\left(\frac{1}{2}k^2 C + kC\right) C_{k,m}. \end{aligned} \tag{S4}$$

Let $k_m = \sum_{k=0}^{m} C^k \exp\left(\frac{1}{2}k^2 C + kC\right) C_{k,m}$ and then the inequality (S4) leads to Lemma S2.

S1.3.3 Proof of Lemma S3

From Chebyshev inequality, we have

$$pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} W_i\right| > \epsilon\right) = pr\left(\left(\sum_{i=1}^{n} W_i\right)^{2m} > n^{2m}\epsilon^{2m}\right) \leq E\left(\sum_{i=1}^{n} W_i\right)^{2m} / \left(n^{2m}\epsilon^{2m}\right). \tag{S5}$$

Combining the inequality (S5) and the Rosenthal inequality (S6) in [2] as follows,

$$E\left(\sum_{i=1}^{n} W_i\right)^{2m} \leq k_m \max\left[\sum_{i=1}^{n} E\left(W_i\right)^{2m}, \left\{E\left(\sum_{i=1}^{n} W_i^2\right)\right\}^m\right], \tag{S6}$$

while k_m is a constant that only depends on m, we have

$$\begin{aligned} pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} W_i\right| > \epsilon\right) &\leq k_m \max\left[\sum_{i=1}^{n} E\left(W_i\right)^{2m}, \left\{E\left(\sum_{i=1}^{n} W_i^2\right)\right\}^m\right] / \left(n^{2m}\epsilon^{2m}\right) \\ &\leq k_m \max\left\{n u_{2m}, \left(n u_2\right)^m\right\} / \left(n^{2m}\epsilon^{2m}\right) \\ &\leq k_m \max\left\{u_{2m}, u_2^m\right\} / \left(n^m \epsilon^{2m}\right). \end{aligned} \tag{S7}$$

Let $v_m = k_m \max\left\{u_{2m}, u_2^m\right\}$ in the inequality (S7) then Lemma S3 follows.

S1.3.4 Proof of Lemma S4

For any $1 \leq j, k \leq p$, notice that

$$\alpha_j = \exp\left(\mu_j^* + \sigma_{jj}^*/2\right) = E\left(Y_{ij}/S_i\right), \tilde{\alpha}_j = n^{-1}\sum_{i=1}^{n} Y_{ij}/S_i,$$

and let

$$u_j = \alpha_j^2 \exp\left(\sigma_{jj}^*\right) = E\left(\left(Y_{ij}^2 - Y_{ij}\right)/S_i^2\right), \tilde{u}_j = n^{-1}\sum_{i=1}^{n}\left(\left(Y_{ij}^2 - Y_{ij}\right)/S_i^2\right),$$

$$w_{jk} = \alpha_j \alpha_k \exp\left(\sigma_{jk}^*\right) = E\left(Y_{ij}Y_{ik}/S_i^2\right), \tilde{w}_{jk} = n^{-1}\sum_{i=1}^{n}\left(Y_{ij}Y_{ik}/S_i^2\right).$$

Since $[Y_{ij}]_{1\leq i\leq n}$, $\left[\left(Y_{ij}^2 - Y_{ij}\right)/S_i^2\right]_{1\leq i\leq n}$, $\left[Y_{ij}Y_{ik}/S_i^2\right]_{1\leq i\leq n}$ are three sets of independent variables, all of which have finite mth moments for any positive integer m by Lemma S2. Then, by Lemma S3, we have

$$pr\left(|\tilde{\alpha}_j - \alpha_j| > \epsilon\right) \leq \frac{v_{1m}}{n^m \epsilon^{2m}}, pr\left(|\tilde{u}_j - u_j| > \epsilon\right) \leq \frac{v_{2m}}{n^m \epsilon^{2m}}, pr\left(|\tilde{w}_{jk} - w_{jk}| > \epsilon\right) \leq \frac{v_{3m}}{n^m \epsilon^{2m}}.$$

Now we can derive the convergence rate of $\tilde{\sigma}_{jk}^m$. Using the boundedness condition in the manuscript, the parameters α_j, u_j, w_{jk} are all in the interval $[\exp(-3C), \exp(4C)]$ for some constant C. Then, for any $\epsilon < \exp(-3C)/2$, we have

$$pr\left(\max\left\{|\tilde{\alpha}_j - \alpha_j|,\ |\tilde{u}_j - u_j|,\ |\tilde{w}_{ij'} - w_{ij'}|\right\} \leq \epsilon\right) > 1 - \left(v_{1m} + v_{2m} + v_{3m}\right)/\left(n^m \epsilon^{2m}\right). \tag{S8}$$

Then with at least probability $1 - \left(v_{1m} + v_{2m} + v_{3m}\right)/\left(n^m \epsilon^{2m}\right)$, we have

$$\max\left\{|\tilde{\alpha}_j - \alpha_j|,\ |\tilde{u}_j - u_j|,\ |\tilde{w}_{ij'} - w_{ij'}|\right\} \leq \epsilon. \tag{S9}$$

According to $\alpha_j, u_j, w_{jk} \geq \exp(-3C)$ and $\epsilon < \exp(-3C)/2$, we can derive from (S9) that

$$\min\left\{\tilde{\alpha}_j, \tilde{u}_j, \tilde{w}_{jk}\right\} > \exp(-3C)/2. \tag{S10}$$

For any $j \neq k$,

$$\sigma_{jj}^* = \log u_j - 2\log \alpha_j,\ \sigma_{jk}^* = \log w_{jk} - \log \alpha_j - \log \alpha_k. \tag{S11}$$

From the Lagrange's mean value theorem, we have, for any $x, y \geq \exp(-3C)/2$,

$$|\log x - \log y| = |x - y|/\xi \leq 2|x - y|/\exp(-3C), \tag{S12}$$

while ξ is a number between x, y. Then combining (S9), (S10) and (S12), we have

$$\max\left\{|\log\tilde{\alpha}_j - \log\alpha_j|, |\log\tilde{u}_j - \log u_j|, |\log\tilde{w}_{jk} - \log w_{jk}|\right\} \leq 2\exp(3C)\epsilon,$$

and thus $\left|\tilde{\sigma}_{jk}^m - \sigma_{jk}^*\right| \leq 6\exp(3C)\epsilon$ using (S11). Then from the probability inequality (S8), for any $\epsilon < \exp(-3C)/2$, we have

$$pr\left(\left|\tilde{\sigma}_{jk}^m - \sigma_{jk}^*\right| \leq 6\exp(3C)\epsilon\right) > 1 - \frac{v_{1m} + v_{2m} + v_{3m}}{n^m \epsilon^{2m}}.$$

So, for any $\eta = 6\exp(3C)\epsilon < 3$ and $C_0 = \{6\exp(3C)\}^{-2m}\left(v_{1m} + v_{2m} + v_{3m}\right)^{-1}$, we have

$$pr\left(\left|\tilde{\sigma}_{jk}^m - \sigma_{jk}^*\right| > \eta\right) \leq \{6\exp(3C)\}^{2m}\frac{v_{1m} + v_{2m} + v_{3m}}{n^m \eta^{2m}} = 1/\left(C_0 n^m \eta^{2m}\right).$$

Then we finish the proof of lemma S4

S1.3.5 Proof of Theorem S1

According to Lemma S4, for any $0 < \epsilon < 3$, $pr\left(\left|\tilde{\sigma}_{jk}^m - \sigma_{jk}^*\right| > \epsilon\right) \le 1/\left(C_0 n^m \epsilon^{2m}\right)$, we have

$$pr\left(\left\|\widetilde{\Sigma}^m - \Sigma^*\right\|_\infty > \epsilon\right) \le p^2/\left(C_0 n^m \epsilon^{2m}\right).$$

Then, from

$$\check{\Sigma}^m = \arg\min_{A \succeq 0} \left\|A - \widetilde{\Sigma}^m\right\|_\infty \Rightarrow \left\|\check{\Sigma}^m - \widetilde{\Sigma}^m\right\|_\infty \le \left\|\Sigma^* - \widetilde{\Sigma}^m\right\|_\infty,$$

we have

$$\begin{aligned}\left\|\widehat{\Sigma}^m - \Sigma^*\right\|_\infty &\le \left\|\widehat{\Sigma}^m - \check{\Sigma}^m\right\|_\infty + \left\|\check{\Sigma}^m - \widetilde{\Sigma}^m\right\|_\infty + \left\|\widetilde{\Sigma}^m - \Sigma^*\right\|_\infty \\ &= 2\left\|\check{\Sigma}^m - \widetilde{\Sigma}^m\right\|_\infty + \left\|\widetilde{\Sigma}^m - \Sigma^*\right\|_\infty \le 3\left\|\widetilde{\Sigma}^m - \Sigma^*\right\|_\infty,\end{aligned}$$

Let $C_1 = 3^{-2m} C_0$ and for any $0 < \epsilon < 6$, we have

$$pr\left(\left\|\widehat{\Sigma} - \Sigma\right\|_\infty > \epsilon\right) \le pr\left(\left\|\widetilde{\Sigma} - \Sigma\right\|_\infty > \epsilon/3\right) \le p^2/\left(C_1 n^m \epsilon^{2m}\right).$$

S1.3.6 Proof of Theorem S2, S3

We define

$$\breve{\Theta}^m = \arg\min_{A = A^T} \frac{1}{2}\mathrm{tr}\left(\widehat{\Sigma}^m A^2\right) - \mathrm{tr}\left(A\right) + \lambda \left\|A\right\|_{1,\mathrm{off}}.$$

Let

$$\begin{aligned}\epsilon = \quad & 1/\max\Bigg[12dk_\Gamma,\ 12\gamma^{-1}(k_\Sigma k_\Gamma^2 + k_\Gamma),\ \left\{12\gamma^{-1}\left(k_\Sigma k_\Gamma^3 + k_\Gamma^2\right) + 5dk_\Gamma^2\right\}\theta_{\min}^{-1}, \\ & \min\left\{s^{1/2}, d+1\right\}\left\{12\gamma^{-1}\left(k_\Sigma k_\Gamma^3 + k_\Gamma^2\right) + 5dk_\Gamma^2\right\}\lambda_{\min}^{-1}(\Theta^*), 1/5\Bigg].\end{aligned}$$

For $\eta > 2$, let $n_f = C_1^{-1/m} p^{\eta/m} \epsilon^{-2}$ and $\epsilon_f = C_1^{-1/(2m)} p^{\eta/(2m)} n^{-1/2}$. According to $n > n_f$, we have

$$\epsilon_f = C_1^{-1/(2m)} p^{\eta/(2m)} n^{-1/2} < C_1^{-1/(2m)} p^{\eta/(2m)} n_f^{-1/2} = \epsilon < 6,$$

while C_1 is the constant in Theorem S1. Then, from Theorem S1, we have

$$pr\left(\|\widehat{\Sigma}^m - \Sigma^*\|_\infty > \epsilon_f\right) < p^2/\left(C_1 n^m \epsilon_f^{2m}\right) = p^{2-\eta},$$

and thus with a probability at least $1 - p^{2-\eta}$,

$$\|\widehat{\Sigma}^m - \Sigma^*\|_\infty \le \epsilon_f < \epsilon \le 1/\max\left\{12dk_\Gamma, 12\gamma^{-1}\left(k_\Sigma k_\Gamma^2 + k_\Gamma\right)\right\}.$$

According to $\lambda = 12\gamma^{-1}\left(k_\Sigma k_\Gamma^2 + k_\Gamma\right)\epsilon_f$, we can get that

$$\begin{aligned}\|\widehat{\Sigma}^m - \Sigma^*\|_\infty < 1/\left(12dk_\Gamma\right), \\ 6\|\widehat{\Sigma}^m - \Sigma^*\|_\infty\left(k_\Sigma k_\Gamma^2 + k_\Gamma\right) \le 0.5\gamma\min\left\{\lambda, 1\right\}.\end{aligned} \tag{S13}$$

Using Lemma S1 (a) with (S13), $\breve{\Theta}^m$ recovers all zeros in Θ^*. That is

$$\mathrm{vec}\left(\breve{\Theta}^m\right)_{G^c} = 0.$$

Using Lemma S1 (b) and according to the fact that $\lambda = 12\gamma^{-1}\left(k_\Sigma k_\Gamma^2 + k_\Gamma\right)\epsilon_f < 12\gamma^{-1}\left(k_\Sigma k_\Gamma^2 + k_\Gamma\right)\epsilon \le 1$ and $\|\widehat{\Sigma}^m - \Sigma^*\|_\infty \le \epsilon_f$, we have

$$\begin{aligned}\left\|\breve{\Theta}^m - \Theta^*\right\|_\infty &< \lambda k_\Gamma + \frac{5}{2}d\left(1 + \lambda\right)\|\widehat{\Sigma}^m - \Sigma^*\|_\infty k_\Gamma^2 \\ &\le \left\{12\gamma^{-1}\left(k_\Sigma k_\Gamma^3 + k_\Gamma^2\right) + 5dk_\Gamma^2\right\}\epsilon_f.\end{aligned} \tag{S14}$$

Then we consider the s nonzeros in Θ^* and vec $\left(\breve{\Theta}^m\right)_{G^c} = 0$, we can easily get

$$\begin{aligned}\left\|\breve{\Theta}^m - \Theta^*\right\|_F &\leq s^{1/2}\left\|\breve{\Theta}^m - \Theta^*\right\|_\infty \\ &< s^{1/2}\left\{12\gamma^{-1}\left(k_\Sigma k_\Gamma^3 + k_\Gamma^2\right) + 5dk_\Gamma^2\right\}\epsilon_f.\end{aligned} \tag{S15}$$

Using $\|A\|_2 \leq \|A\|_F$ and $\|A\|_2 \succeq 0$ while $|A_{jj}| \geq \sum_{k\neq j}|A_{jk}|$ for all $1 \leq j \leq p$

$$\begin{aligned}\left\|\breve{\Theta}^m - \Theta^*\right\|_2 &\leq \min\left\{s^{1/2}, d+1\right\}\left\|\breve{\Theta}^m - \Theta^*\right\|_\infty \\ &< \min\left\{s^{1/2}, d+1\right\}\left\{12\gamma^{-1}\left(k_\Sigma k_\Gamma^3 + k_\Gamma^2\right) + 5dk_\Gamma^2\right\}\epsilon_f.\end{aligned} \tag{S16}$$

From (S14) and combining $\epsilon_f < \epsilon \leq \theta_{\min}/\left(12\gamma^{-1}\left(k_\Sigma k_\Gamma^3 + k_\Gamma^2\right) + 5dk_\Gamma^2\right)$, we have

$$\left\|\breve{\Theta}^m - \Theta^*\right\|_\infty < \theta_{\min},$$

which means that $\breve{\Theta}^m$ also recovers the nonzeros in Θ^*.

Finally, we check $\widehat{\Theta}^m = \breve{\Theta}^m$ to finish the proof. We just need to verify $\lambda_{\min}\left(\breve{\Theta}^m\right) > 0$, that can be obtained from $\left\|\breve{\Theta}^m - \Theta^*\right\|_2 < \lambda_{\min}\left(\Theta^*\right)$. So using (S16) and combining

$$\epsilon_f < \epsilon \leq \lambda_{\min}\left(\Theta^*\right)/\left[\min\left\{s^{1/2}, d+1\right\}\left\{12\gamma^{-1}\left(k_\Sigma k_\Gamma^3 + k_\Gamma^2\right) + 5dk_\Gamma^2\right\}\right],$$

we get the conclusion.

Above all, $\widehat{\Theta}^m$ recovers all zeros and nonzeros in Θ^* and meet all the convergence rates for $\breve{\Theta}^m$ in (S14), (S15), (S16) with a probability at least $1 - p^{2-\eta}$, then we finish the proof of Theorem S2 and S3.

S2 Theoretical proofs for maximum marginal likelihood estimator (MMLE)

S2.1 Notation

For any constant $d \leq p$, we say a random vector $\mathbf{y} = (y_1, y_2, ..., y_d)^T$ follows the PLN distribution with parameters $\boldsymbol{\mu}, \Sigma \succeq 0$ and known library size S, if there exists a random vector $\mathbf{x} = (x_1, x_2, ..., x_d)^T$ such that $y_j \sim \mathrm{Poisson}(S\exp(x_j))$ independently for all $j = 1, ..., d$ and $\mathbf{x} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$, and we denote this ditribution by $\mathbf{y} \sim \mathrm{PLN}(S; \boldsymbol{\mu}, \Sigma)$. We also suppose that S is sampled from a bounded support distribution with probability density function $g(S)$ which is independent with $\boldsymbol{\mu}, \Sigma$, then the joint distribution of $\mathbf{y}, S$ can be written as:

$$p(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma) = \int f(\mathbf{x}; \boldsymbol{\mu}, \Sigma)P(\mathbf{y}|\mathbf{x}, S)g(S)d\mathbf{x},$$

where $f(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = \left(1/\sqrt{2\pi}\right)^d\left(\det\left(\Sigma\right)\right)^{-1/2}\exp\left(-(\mathbf{x} - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})/2\right)$ is the probability density function of Gaussian ditribution $\mathrm{N}(\boldsymbol{\mu}, \Sigma)$, and

$$P(\mathbf{y}|\mathbf{x}, S) = \prod_{j=1}^{d}\exp(x_j y_j)S^{y_j}\exp(-S\exp(x_j))/y_j!$$

is the probability function of d indenpendent Poisson distributions $\mathrm{Poisson}(S\exp(x_j))$.

Assuming $\mathbf{y}_i \sim \mathrm{PLN}(S_i; \boldsymbol{\mu}, \Sigma)$, S_i $(1 \leq i \leq n)$ are n independent samples from the bounded distribution $g(S)$. Then we can write the log-likelihood of $y = (\mathbf{y}_1, ..., \mathbf{y}_n)^T$ as follows,

$$\begin{aligned}\log\, L(y, \mathbf{S}; \boldsymbol{\mu}, \Sigma) &= \sum_{i=1}^{n}\log p(\mathbf{y}_i, S_i; \boldsymbol{\mu}, \Sigma) \\ &= \sum_{i=1}^{n}\log\int f(\mathbf{x}; \boldsymbol{\mu}, \Sigma)P\left(\mathbf{y}_i|\mathbf{x}, S_i\right)g(S_i)d\mathbf{x}.\end{aligned} \tag{S17}$$

For any symmetric matrix $\Sigma = [\sigma_{ij}]_{1 \le i,j \le d}$, $\text{vech}(\Sigma)$ is defined as

$$\text{vech}(\Sigma) = (\sigma_{11}, \sigma_{12}, \sigma_{13}, \ldots, \sigma_{1d}, \sigma_{22}, \sigma_{23}, \ldots, \sigma_{2d}, \ldots, \sigma_{(d-1)(d-1)}, \sigma_{(d-1)d}, \sigma_{dd})^T,$$

and $\text{vech}_2(\Sigma)$ is

$$\text{vech}_2(\Sigma) = (\sigma_{11}, 2\sigma_{12}, 2\sigma_{13}, \ldots, 2\sigma_{1d}, \sigma_{22}, 2\sigma_{23}, \ldots, 2\sigma_{2d}, \ldots, \sigma_{(d-1)(d-1)}, 2\sigma_{(d-1)d}, \sigma_{dd})^T.$$

In order to simplify the notations, we define an operator $\mathcal{T}(f)$ that maps functions in $\mathbf{x}$ to functions in $\mathbf{y}$ as

$$\mathcal{T}(h) = \int h(\mathbf{x}) f(\mathbf{x}; \boldsymbol{\mu}, \Sigma) P(\mathbf{y}|\mathbf{x}, S) g(S) d\mathbf{x}.$$

Also, we define $\mathbb{I}(\mathbf{x}) \equiv 1$ as a constant function. Then for any random vector $\mathbf{y} \sim \text{PLN}(S_i; \boldsymbol{\mu}, \Sigma)$, we can define the gradiant (score function) of the log-likelihood as

$$\begin{aligned}\mathcal{S}(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma) &= \frac{\partial \log p(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma)}{\partial(\boldsymbol{\mu}_0^T, \text{vech}(\Sigma)^T)^T} \\ &= \left(\left(\frac{\partial \log p(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma)}{\partial \boldsymbol{\mu}} \right)^T, \left(\frac{\partial \log p(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma)}{\partial \text{vech}(\Sigma)} \right)^T \right)^T,\end{aligned}$$

where

$$\frac{\partial \log p(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma)}{\partial \boldsymbol{\mu}} = \frac{\mathcal{T}(\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}))}{\mathcal{T}(\mathbb{I})}, \tag{S18}$$

and

$$\frac{\partial \log p(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma)}{\partial \text{vech}(\Sigma)} = \frac{\mathcal{T}(\text{vech}_2(\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}))}{2\mathcal{T}(\mathbb{I})} - \frac{1}{2}\text{vech}_2(\Sigma^{-1}). \tag{S19}$$

We also define the second order derivative of the log-likelihood as

$$\mathcal{H}(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma) = \frac{\partial^2 \log p(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma)}{\partial(\boldsymbol{\mu}^T, \text{vech}(\Sigma)^T)^T \partial(\boldsymbol{\mu}^T, \text{vech}(\Sigma)^T)}.$$

Since the form of the $\mathcal{H}(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma)$ is very complicated, it is clear to divide it into several parts to express the form. We define $e_j = (0, ...0, 1, 0..., 0)^T$ is the jth column vector of $d \times d$ identity matrix, then we have the following equations.

For the second order derivative of $\boldsymbol{\mu}$, for any $j = 1, \cdots, d$, we have

$$\begin{aligned}\frac{\partial^2 \log p(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma)}{\partial \mu_j \partial \boldsymbol{\mu}^T} &= \frac{\mathcal{T}(e_j^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} - e_j^T \Sigma^{-1})}{\mathcal{T}(\mathbb{I})} \\ &\quad - \frac{\mathcal{T}(e_j^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}))\mathcal{T}((\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1})}{\mathcal{T}^2(\mathbb{I})}.\end{aligned} \tag{S20}$$

For the second order derivative of Σ, for any $i \neq j \in \{1, \cdots, d\}$, we have

$$\begin{aligned}\frac{\partial^2 \log p(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma)}{\partial \sigma_{ij} \partial \text{vech}(\Sigma)^T} =& - \frac{\mathcal{T}(\text{vech}_2(\Sigma^{-1} e_i (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} e_j)^T)}{\mathcal{T}(\mathbb{I})} \\ & - \frac{\mathcal{T}(\text{vech}_2(\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) e_j^T \Sigma^{-1} e_i^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}))^T)}{\mathcal{T}(\mathbb{I})} \\ & + \frac{\mathcal{T}(e_i^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} e_j \text{vech}_2(\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}))}{2\mathcal{T}(\mathbb{I})} \\ & - \frac{\mathcal{T}(e_i^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} e_j)}{\mathcal{T}(\mathbb{I})} \cdot \frac{\mathcal{T}(\text{vech}_2(\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}))}{2\mathcal{T}(\mathbb{I})} \\ & + \text{vech}_2(\Sigma^{-1} e_i e_j^T \Sigma^{-1}).\end{aligned} \tag{S21}$$

When $i = j$ the derivative is half of the right hand of (S21).

For the interaction term, for any $j = 1, \cdots, d$, we have

$$\begin{aligned}
\frac{\partial^2 \log p(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma)}{\partial \mu_j \partial \mathrm{vech}(\Sigma)^T} = & \frac{\mathcal{T}(\mathrm{vech}_2(e_j^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}))}{2\mathcal{T}(\mathbb{I})} \\
& - \frac{\mathcal{T}(\mathrm{vech}_2(\Sigma^{-1} e_j (\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}))}{\mathcal{T}(\mathbb{I})} \\
& - \frac{\mathcal{T}(e_j^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}))}{\mathcal{T}(\mathbb{I})} \cdot \frac{\mathcal{T}(\mathrm{vech}_2(\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}))}{2\mathcal{T}(\mathbb{I})}.
\end{aligned} \tag{S22}$$

Let $\mathcal{F}$ be the densities of the given PLN family indexed by finite $(d + d(d+1)/2)$ parameters $\theta = (\boldsymbol{\mu}^T, \mathrm{vech}(\Sigma)^T)^T$. Then we can simplify the notation $p(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma)$, $\mathcal{S}(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma)$, $\mathcal{H}(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma)$ using $p(\mathbf{y}, S; \theta)$, $\mathcal{S}(\mathbf{y}, S; \theta)$, $\mathcal{H}(\mathbf{y}, S; \theta)$, respectively. We also define Hellinger distance $d(\cdot, \cdot)$ for two distributions on $\mathcal{F}$ as

$$d(p_1, p_2) = \left[\int \left(p_1^{1/2} - p_2^{1/2}\right)^2 d\nu\right]^{1/2} = \left\|p_1^{1/2} - p_2^{1/2}\right\|_{L_2},$$

where ν is Lebesgue-Stieljes(L-S) measure and p_1, p_2 are two densities.

For any $u > 0$, call a finite set $\{(f_j^L, f_j^U), j = 1, ..., N\}$ a Hellinger u-bracketing of $\mathcal{F}$ if $\left\|(f_j^L)^{1/2} - (f_j^U)^{1/2}\right\|_2 \leq u$ for $j = 1, ..., N$, and for any $p(\mathbf{y}, S; \theta) \in \mathcal{F}$, there is a j such that $f_j^L \leq p(\mathbf{y}, S; \theta) \leq f_j^U$. The bracketing Hellinger metric entropy of $\mathcal{F}$, denoted by the function $H(\cdot, \mathcal{F})$, is defined by $H(u, \mathcal{F})$ = logarithm of the cardinality of the Hellinger u-bracketing of the smallest size.

Under the bounded condition in the manuscript, $\theta^* = ((\boldsymbol{\mu}^*)^T, \mathrm{vech}(\Sigma^*)^T)^T$ is restricted on a bounded set $\mathcal{O}$, while $\mathcal{O} = \{\theta^* | \max_{1 \leq j,k \leq p} \left\{\left|\mu_j^*\right|, \left|\sigma_{jk}^*\right|\right\} \leq M_3, M_4 \leq \lambda_{min}(\Sigma^*) \leq \lambda_{max}(\Sigma^*) \leq M_5\}$. We also define bounded set $\mathcal{O}_d = \{\theta | \max_{1 \leq j,k \leq p} \{|\mu_j|, |\sigma_{jk}|\} \leq M_3, M_4 \leq \lambda_{min}(\Sigma) \leq \lambda_{max}(\Sigma) \leq M_5\}$ for any d-dimensional PLN parameters $\theta = (\boldsymbol{\mu}^T, \mathrm{vech}(\Sigma)^T)^T$, that is $\mathcal{O} = \mathcal{O}_p$.

In section 2.3 of the manuscript, for $1 \leq j \leq p$, we estimate σ_{jj}^* by maximizing the marginal log-likelihood of $Y_{\cdot j}$, and for $1 \leq j \neq k \leq p$, we estimate σ_{jk}^* by maximizing the marginal log-likelihood of $(Y_{\cdot j}, Y_{\cdot k})$. For the PLN model, the marginal distribution is also a PLN distribution, so the marginal distributions of Y_{ij} and (Y_{ij}, Y_{ik}) are both PLN distributions. Then from equation (S17), we can write the marginal log-likelihood of $\mathbf{Y}_{\cdot j}$ as $\log L(\mathbf{Y}_{\cdot j}, \mathbf{S}; \mu, \sigma)$ with parameters $(\mu, \sigma)^T \in \mathcal{O}_1$, and the marginal log-likelihood of $(\mathbf{Y}_{\cdot j}, \mathbf{Y}_{\cdot k})$ as $L((\mathbf{Y}_{\cdot j}, \mathbf{Y}_{\cdot k}), \mathbf{S}; \boldsymbol{\mu}, \Sigma)$ with parameters $(\boldsymbol{\mu}^T, \mathrm{vech}(\Sigma)^T)^T \in \mathcal{O}_2$. Then for any $1 \leq j \neq k \leq p$, we derive

$$\tilde{\sigma}_{jj} = \arg\max_{\sigma} \left[\max_{\mu} \log L(Y_{\cdot j}, \mathbf{S}; \mu, \sigma)\right], \tag{S23}$$

and

$$\tilde{\sigma}_{jk} = \arg\max_{\Sigma_{12}} \left[\max_{\boldsymbol{\mu}, \Sigma_{11}, \Sigma_{22}} \log L((Y_{\cdot j}, Y_{\cdot k}), \mathbf{S}; \boldsymbol{\mu}, \Sigma)\right], \tag{S24}$$

where $(\mu, \sigma)^T$ are restricted on $\mathcal{O}_1$ and $(\boldsymbol{\mu}^T, \mathrm{vech}(\Sigma)^T)^T$ are restricted on $\mathcal{O}_2$. The explicit forms of the first and second order partial derivatives of the marginal log-likelihood function are shown in the equation (S18), (S19), (S20), (S21) and (S22).

S2.2 Proofs

S2.2.1 Lemmas

Lemma S5 *There exist positive constants c_1, c_2, c_3, such that, for any $\epsilon > 0$, if*

$$\int_{s^2/2^8}^{\sqrt{2}s} H^{1/2}\left(u/c_2, \mathcal{F} \cap \left\{\left\|p_2^{1/2} - p_1^{1/2}\right\|_{L_2}^2 \leq 2s^2\right\}\right) du \leq c_3 n^{1/2} s^2 \tag{S25}$$

for all $s \geq \epsilon$, then for the MLE $\hat{\theta}$ of the true parameter θ using n independent samples, we have

$$pr\left(\left\|p^{1/2}(\mathbf{y}, S; \hat{\theta}) - p^{1/2}(\mathbf{y}, S; \theta)\right\|_{L_2} \geq \epsilon\right) \leq 5\exp(-c_1 n\epsilon^2)$$

.

Lemma S5 is the local version of Theorem 1 and 2 in [4]. Most of the following lemmas are verifing the conditions of this Lemma in the PLN model.

Lemma S6 *Suppose that $S \in [M_1, M_2]$, $|\mu| \leq M_3$, and let*

$$p_1(y, S; \mu, \sigma) = (y!)^{-1} \int \exp\left(-\frac{1}{2}(x-\mu)^2/\sigma\right) \exp\left(-S\exp(x)\right) \exp(xy) S^y dx \qquad \text{(S26)}$$

be the 1-dimensional PLN probability density function. Then, for any $\log(y) \geq |\log(M_1)| + |\log(M_2)| + M_3 + 1$, we have a constant **C***, such that,*

$$C/(\sqrt{\pi}y)\exp\left(-\frac{1}{2}(\log(y))^2/\sigma\right) \leq p_1(y, S; \mu, \sigma) \leq$$
$$\left[2\exp\left(\frac{1}{2}\log(y)\left(|\log(M_1)| + |\log(M_2)| + M_3 + 1\right)/\sigma\right) + 1\right] \exp\left(-\frac{1}{2}(\log(y))^2/\sigma\right)/\sqrt{\pi y}.$$

Remark 1 *In fact that for any $(\boldsymbol{\mu}^T, \mathrm{vech}(\Sigma)^T)^T \in \mathcal{O}_d$ and $S \in [M_1, M_2]$, we have two constants C_1, C_2 such that, $C_1 p(\mathbf{y}, S; \boldsymbol{\mu}, M_4 I_d) \leq p(\mathbf{y}, S; \boldsymbol{\mu}, \Sigma) \leq C_2 p(\mathbf{y}, S; \boldsymbol{\mu}, M_5 I_d)$, where I_d is $d \times d$ identity matrix. This is because*

$$\begin{aligned}
&(M_4/M_5)^{d/2} f(\mathbf{x}; \boldsymbol{\mu}, M_4 I_d) \\
=& \left(1/\sqrt{2\pi}\right)^d (M_4/M_5)^{d/2} \left(\det(M_4 I_d)\right)^{-1/2} \exp\left(-(\mathbf{x}-\boldsymbol{\mu})^T (M_4 I_d)^{-1} (\mathbf{x}-\boldsymbol{\mu})/2\right) \\
\leq& \left(1/\sqrt{2\pi}\right)^d \left(\det(\Sigma)\right)^{-1/2} \exp\left(-(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})/2\right) = f(\mathbf{x}; \boldsymbol{\mu}, \Sigma) \\
\leq& \left(1/\sqrt{2\pi}\right)^d (M_5/M_4)^{d/2} \left(\det(M_5 I_d)\right)^{-1/2} \exp\left(-(\mathbf{x}-\boldsymbol{\mu})^T (M_5 I_d)^{-1} (\mathbf{x}-\boldsymbol{\mu})/2\right) \\
=& (M_5/M_4)^{d/2} f(\mathbf{x}; \boldsymbol{\mu}, M_5 I_d).
\end{aligned}$$

$p(\mathbf{y}, S; \boldsymbol{\mu}, M_4 I_d)$ and $p(\mathbf{y}, S; \boldsymbol{\mu}, M_5 I_d)$ can be written as a product of 1-dimensional PLN densities, so using Lemma S6 we can derive that there exists two functions $f_{low}(\cdot)$ and $f_{up}(\cdot)$of $\mathbf{y}$ such that $f_{low}(\mathbf{y}) \leq p(\mathbf{y}, S; \theta) \leq f_{up}(\mathbf{y})$, and $\int f_{low}(\mathbf{y}) K(\mathbf{y}) d\mathbf{y} < \infty, \int f_{up}(\mathbf{y}) K(\mathbf{y}) d\mathbf{y} < \infty$ for any polynomial $K(\mathbf{y})$ of $\mathbf{y}$.

Lemma S7 (Dominating function) *Let*

$$u(\mathbf{y}, S; \theta) = \frac{\mathcal{T}\left((x_1-\mu_1)^4\right)}{\mathcal{T}(\mathbb{I})}, \qquad \text{(S27)}$$

where $\mathbf{y} \in \mathbb{N}^d$ from PLN distribution with parameters $\theta \in \mathcal{O}_d$ and known library size $S \in [M_1, M_2]$, $\mathbf{x}$ is the latent normal random vectors. Then there exists a polynomial function $v(\mathbf{y})$ with a constant C only depending on d and $M_i, i = 1, ..., 5$

$$v(\mathbf{y}) = (4||\mathbf{y}||_2/m)^4 + C,$$

such that $E_\theta(v(\mathbf{y})) < \infty$ and $|u(\mathbf{y}, S; \theta)| \leq v(\mathbf{y})$ for any $p(\mathbf{y}, S; \theta) \in \mathcal{F}$ satisfying conditions.

Remark 2 *Applying the same proof as in Lemma S7, the polynomial $(x_1-\mu_1)^4$ can be replaced by any polynomial with respect to $\mathbf{x}-\boldsymbol{\mu}$, e.g. $x_i - \mu_i$ and $(x_i-\mu_i)(x_j-\mu_j)$. Further, since polynomial set is a multiplicative set, then for any two polynomial functions $\psi_1(\mathbf{x}-\boldsymbol{\mu}), \psi_2(\mathbf{x}-\boldsymbol{\mu})$, $\mathcal{T}\left(\psi_1(\mathbf{x}-\boldsymbol{\mu})\right)\mathcal{T}\left(\psi_2(\mathbf{x}-\boldsymbol{\mu})\right)/\mathcal{T}^2(\mathbb{I})$ also have an polynomial upper bound and have a finite expectation.*

Lemma S7 and Remark 2 are previous results in [3].

Remark 3 *Suppose that* $\mathbf{y}$ *is a random vector from PLN distribution with parameters* $\theta \in \mathcal{O}_d$ *and random* S *from* $g(S)$ *with bounded support* $[M_1, M_2]$*. Then from Remark 2 and the forms of* $\mathcal{S}$ *and* $\mathcal{H}$ *in the equation (S18), (S19), (S20), (S21) and (S22), we have for any* $1 \le i, j \le d + d(d+1)/2$*, there exists two polynomial functions* $K_1(\mathbf{y}), K_2(\mathbf{y})$ *with* $E_\theta[K_1(\mathbf{y})] < \infty, E_\theta[K_2(\mathbf{y})] < \infty$*, such that* $|\mathcal{S}_i(\mathbf{y}, S; \theta)| \le K_1(\mathbf{y})$ *and* $|\mathcal{H}_{i,j}(\mathbf{y}, S; \theta)| \le K_2(\mathbf{y})$*.*

Lemma S8 *Suppose that* $\mathbf{y}$ *is a random vector from PLN distribution with parameters* $\theta \in \mathcal{O}_d$ *and random* S *from* $g(S)$ *with bounded support* $[M_1, M_2]$*. Then* $E_\theta\left(\mathcal{S}(\mathbf{y}, S; \theta)\mathcal{S}(\mathbf{y}, S; \theta)^T\right)$ *is positive defined and continuous for* θ*.*

Lemma S9 *For any* $\theta, \theta' \in \mathcal{O}_d$ *and* $S \in [M_1, M_2]$*, let* $p_1 = p(\mathbf{y}, S; \theta)$*,* $p_2 = p(\mathbf{y}, S; \theta')$*. Then for some constant* $K > 0$ *which is independent of* $\boldsymbol{\mu}, \Sigma$*, there exists a constant* C_1*, such that, for any* $\|\theta - \theta'\|_2 \le K$*, we have* $C_1 \|\theta - \theta'\|_2 \le d(p_1, p_2)$

Remark 4 *According to the existence of the moment estimator, we can easily derive the identifiablity of PLN model, so we have* $d(p_1, p_2) \neq 0$ *when* $p_1 \neq p_2$*. So in the closed bouned domain* $\mathcal{D} = \{(\theta, \theta') | \|\theta - \theta'\|_2 \ge K,\ \theta, \theta' \in \mathcal{O}_d\}$*, the continuous function* $d(p_1, p_2)$ *has the minimum value* $m > 0$*. Since* $\|\theta - \theta'\|_2$ *can be bounded by some constant* M*, then we have* $m \|\theta - \theta'\|_2 / M \le d(p_1, p_2)$ *for* $(\theta, \theta') \in \mathcal{D}$*.So we can remove the constrain* $\|\theta - \theta'\|_2 \le K$ *in Lemma S9 and have* $C_1' \|\theta - \theta'\|_2 \le d(p_1, p_2)$ *for all* $\theta, \theta' \in \mathcal{O}_d$*, where* $C_1' = \min\{C_1, m/M\}$*.*

Lemma S10 *There exists a measurable function* $m(\mathbf{y})$ *and a constant* C_2*, such that* $\int m^2(\mathbf{y}) d\nu = C_2^2 < \infty$*, for any* $\theta_3, \theta_4 \in \mathcal{O}_d$ *and* $S \in [M_1, M_2]$*, let* $p_3 = p(\mathbf{y}, S; \theta_3), p_4 = p(\mathbf{y}, S; \theta_4)$*, we have* $|p_3^{1/2} - p_4^{1/2}| \le m(\mathbf{y}) \|\theta_3 - \theta_4\|_2$*.*

Lemma S11 *Suppose that* $\mathbf{y}_i, i = 1, ..., n$ *are* n *random d-dimensional vectors from the PLN distribution with parameters* $\theta \in \mathcal{O}_d$ *and random variable* S_i *from* $g(S)$ *with bounded support* $[M_1, M_2]$*, let* $\hat{\theta}$ *is the MLE of* θ *restricted on* $\mathcal{O}_d$*. Then we have*

$$pr\left(\left\|\hat{\theta} - \theta\right\|_2 \ge \epsilon\right) \le 5\exp(-Cn\epsilon^2),$$

where C *is a constant independent with parameters.*

S2.3 Proof of Lemma S6

Let $\mu' = \mu + \log(S)$ and $t = x + \log(S)$, then $|\mu'| \le |\log(M_1)| + |\log(M_2)| + M_3$ and

$$p_1(y, S; \mu, \sigma) = (y!)^{-1} \int \exp\left(-\frac{1}{2}(t - \mu^*)^2/\sigma\right) \exp\left(-\exp(t)\right) \exp(ty) dt.$$

The lower bound can be derived from previous result in [3], we only prove the upper bound here. We divide the real line into 5 disjoint intervals and estimate the upper bound of this integration separately. Let $h(t, y) = \exp\left(-\exp(t)\right)\exp(ty)$.

Part 1 $(\log(y) - 1, \log(y) + 1]$. Note that $h(t)$ attains the maximum at $t = \log(y)$, and $\log(y) \ge \mu' + 1$ then

$$\begin{aligned}
&\int_{\log(y)-1}^{\log(y)+1} \exp\left(-\frac{1}{2}\left(t - \mu'\right)^2/\sigma\right) h(t, y) dt \\
&\le 2\exp\left(y \log\left(y\right) - y\right) \exp\left(-\frac{1}{2}\left(\log\left(y\right) - 1 - \mu'\right)^2/\sigma\right).
\end{aligned}$$

Part 2 $(\log(y) + 1, 2\log(y)]$. When $t \ge \log(y) + 1$, we note that $h(t)$ attains the maximum at $t = \log(y) + 1$, then we have

$$\begin{aligned}
&\int_{\log(y)+1}^{2\log(y)} \exp\left(-\frac{1}{2}\left(t - \mu'\right)^2/\sigma\right) h(t, y) dt \\
\le &\int_{\log(y)+1}^{2\log(y)} h(t, y) dt \\
\le &\left(\log\left(y\right) - 1\right) \exp\left(y\left(\log\left(y\right) + 1\right) - ey\right).
\end{aligned}$$

Part 3 $(2\log(y), \infty)$. When $t \geq 2\log(y)$, $h(t) \leq 1$. Then we have

$$\begin{aligned}&\int_{2\log(y)}^{\infty} \exp\left(-\frac{1}{2}(t-\mu')^2/\sigma\right) h(t,y)dt \\ \leq &\int_{2\log(y)}^{\infty} \exp\left(-\frac{1}{2}(t-\mu')^2/\sigma\right) dx \leq \sqrt{2\pi\sigma}.\end{aligned}$$

Part 4 $(0, \log(y)-1]$. When $t \leq \log(y)-1$, we note that $h(t,y)$ attains the maximum at $x = \log(y)-1$. Then we have

$$\begin{aligned}&\int_{0}^{\log(y)-1} \exp\left(-\frac{1}{2}(t-\mu')^2/\sigma\right) h(t,y)dt \\ \leq &(\log(y)-1)\exp\left(y(\log(y)-1) - y/e\right).\end{aligned}$$

Part 5 $(-\infty, 0]$. When $t \leq 0$, we note that $h(t,y) \leq 1$. Similar to Part 3, we have

$$\int_{-\infty}^{0} \exp\left(-\frac{1}{2}(t-\mu')^2/\sigma\right) h(t,y)dt \leq \sqrt{2\pi\sigma}.$$

By Stirling's approximation, we have $y! = \exp(y\log(y) - y)\left(\sqrt{2\pi y}\right)(1+o(1))$. Thus, combining Part 1-5, we have

$$\begin{aligned}\sqrt{2\pi y} f(y,S;\mu,\sigma)(1+o(1)) \leq 2\exp\left(-\frac{1}{2}\left(\log(y)-1-\mu'\right)^2/\sigma\right) + \left(\log(y)-1\right)\exp\left((-e+2)y\right). \\ + (\log(y)-1)\exp\left(-y/e\right) + \exp(-y\log(y)+y)\left(2\sqrt{2\pi\sigma}\right).\end{aligned} \tag{S28}$$

Then, it is clear that as $y \to \infty$,

$$\sqrt{\pi y} f(y,S;\mu,\sigma)/\exp\left(-\frac{1}{2}\left(\log(y)\right)^2/\sigma\right) \leq 2\exp\left(\log(y)(\mu'+1)\right) + o(1).$$

Combining with $|\mu^*| \leq |\log(M_1)| + |\log(M_2)| + M_3$, we can derive the upper bound.

S2.3.1 Proof of Lemma S8

Notice that if for any fixed positive S, $E_\theta\left(\mathcal{S}(\mathbf{y},S;\theta)\mathcal{S}(\mathbf{y},S;\theta)^T\right)$ is positive definite, then we take expectation for S, it is still positive definite, so we just need to prove the positive definiteness for fixed S. We just need to prove that if $E_\theta\left[\mathbf{t}^T\left(\mathcal{S}(\mathbf{y},S;\theta)\mathcal{S}(\mathbf{y},S;\theta)^T\right)\mathbf{t}\right] = 0$, then $\mathbf{t} = 0$. Since $\mathbf{y}$ is discrete random vector, so for any $\mathbf{y}$ taken values from non-negative integer, we have $\mathbf{t}^T\mathcal{S}(\mathbf{y},S;\theta) = 0$. According to the definition of $\mathcal{S}(\mathbf{y},S;\theta)$, we split $\mathbf{t}$ into two parts $\mathbf{t} = (\mathbf{t}_1^T, \mathbf{t}_2^T)^T$, where $\mathbf{t}_1$ with length d, and $\mathbf{t}_2$ with length $d(d-1)/2$, then we have,

$$\begin{aligned}&\mathbf{t}_1^T\frac{\partial \log p(\mathbf{y},S;\theta)}{\partial \boldsymbol{\mu}} + \mathbf{t}_2^T\frac{\partial \log p(\mathbf{y},S;\theta)}{\partial \mathrm{vech}(\Sigma)} = 0. \\ \Rightarrow &\frac{\mathcal{T}\left(\mathbf{t}_1^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\mathcal{T}(\mathbb{I})} + \frac{\mathcal{T}\left(\mathbf{t}_2^T\mathrm{vech}_2(\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}-\Sigma^{-1})\right)}{2\mathcal{T}(\mathbb{I})} = 0 \\ \Rightarrow &\int \mathcal{T}\left(\mathbf{t}_1^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) + \mathbf{t}_2^T\mathrm{vech}_2(\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}-\Sigma^{-1})/2\right)v(\mathbf{y})d\mathbf{y} = 0,\end{aligned} \tag{S29}$$

for any integrable function $v(\mathbf{y})$. According the Fubini theorem, we have

$$\begin{aligned}\int \mathcal{T}\left(u(\mathbf{x})\right)v(\mathbf{y})d\mathbf{y} = &\int\int u(\mathbf{x})f(\mathbf{x};\theta)P(\mathbf{y}|\mathbf{x},S)g(S)v(\mathbf{y})d\mathbf{y}d\mathbf{x} \\ = &g(S)E_\theta(E(v(\mathbf{y})|\mathbf{x},S)u(\mathbf{x})).\end{aligned} \tag{S30}$$

Using (S30), we can rewrite (S29) as

$$E_\theta\left(E(v(\mathbf{y})|\mathbf{x},S)\left(\mathbf{t}_1^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) + \mathbf{t}_2^T\mathrm{vech}_2(\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}-\Sigma^{-1})/2\right)\right) = 0.$$

Given positive S, using the moments of poisson distribution, for any non-negative integer vector $\mathbf{n}$, we can choose $v(\mathbf{y})$ such that $E(v(\mathbf{y})|\mathbf{x}, S) = \exp(\mathbf{n}^T\mathbf{x})$. So we have,

$$E_\theta\left(\exp(\mathbf{n}^T\mathbf{x})\left(\mathbf{t}_1^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) + \mathbf{t}_2^T\mathrm{vech}_2(\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1} - \Sigma^{-1})/2)\right)\right) = 0.$$

After calculation using the moment generating function of normal distribution, we have for all non-negative integer vector $\mathbf{n}$,

$$\exp(\mathbf{n}^T\Sigma\mathbf{n}/2 + \mathbf{n}^T\boldsymbol{\mu})(\mathbf{t}_1^T\mathbf{n} + \mathbf{n}^T\mathrm{T}_2\mathbf{n}) = 0,$$

where $\mathrm{vech}_2(\mathrm{T}_2) = \mathbf{t}_2$, that means $\mathbf{t}_1^T\mathbf{n} + \mathbf{n}^T\mathrm{T}_2\mathbf{n} = 0$. However, if $\mathbf{t}_1^T\mathbf{n}_1 + \mathbf{n}_1^T\mathrm{T}_2\mathbf{n}_1 = 0$, $\mathbf{t}_1^T\mathbf{n}_2 + \mathbf{n}_2^T\mathrm{T}_2\mathbf{n}_2 = 0$ and $\mathbf{t}_1^T(\mathbf{n}_1+\mathbf{n}_2) + (\mathbf{n}_1+\mathbf{n}_2)^T\mathrm{T}_2(\mathbf{n}_1+\mathbf{n}_2) = 0$, then we have $\mathbf{n}_1^T\mathrm{T}_2\mathbf{n}_2 = 0$. That means for any non-negative integer vector $\mathbf{n}_1, \mathbf{n}_2$, $\mathbf{n}_1^T\mathrm{T}_2\mathbf{n}_2 = 0$, So T_2 must be zero, and then so does $\mathbf{t}_1$. So we finish the proof of $\mathbf{t} = 0$, and get the positive definiteness for $E_\theta\left(\mathcal{S}(\mathbf{y}, S; \theta)\mathcal{S}(\mathbf{y}, S; \theta)^T\right)$.

As for the continuity, since $\mathcal{S}(\mathbf{y}, S; \theta)$ is continuous for θ and has dominating functions with finite expectation from Remark 3, so $E_\theta\left(\mathcal{S}(\mathbf{y}, S; \theta)\mathcal{S}(\mathbf{y}, S; \theta)^T\right)$ is also continuous for θ. Thus we complete the proof.

S2.3.2 Proof of Lemma S9

First we get the Taylor expansion of $p_2^{1/2}$ on θ,

$$\begin{aligned} p_2^{1/2} &= p_1^{1/2} + p_1^{-1/2}\left(\frac{\partial p(\mathbf{y}, S; \theta)}{\partial\theta}\right)^T(\theta'-\theta)/2 + (\theta'-\theta)^T R(\mathbf{y}, S; \theta^*)(\theta'-\theta)/2 \\ &= p_1^{1/2} + p_1^{1/2}\mathcal{S}(\mathbf{y}, S; \theta)^T(\theta'-\theta)/2 + (\theta'-\theta)^T R(\mathbf{y}, S; \theta^*)(\theta'-\theta)/2, \end{aligned} \tag{S31}$$

where $R(\mathbf{y}, S; \theta^*) = p^{1/2}(\mathbf{y}, S; \theta^*)\left[\mathcal{S}(\mathbf{y}, S; \theta^*)\mathcal{S}(\mathbf{y}, S; \theta^*)^T/4 + \mathcal{H}(\mathbf{y}, S; \theta^*)/2\right]$, θ^* is between θ and θ'. We define $p_* = p(\mathbf{y}, S; \theta^*)$ and $\delta = \theta' - \theta$, then we have,

$$\begin{aligned} d(p_1, p_2) &= \left[\int\left(p_1^{1/2} - p_2^{1/2}\right)^2 d\nu\right]^{1/2} \\ &= \left[\int\left(p_1^{1/2}\mathcal{S}(\mathbf{y}, S; \theta)^T\delta + p_*^{1/2}\delta^T R(\mathbf{y}, S; \theta^*)\delta\right)^2 d\nu\right]^{1/2}/2 \\ &= \left[\int p_1\delta^T\mathcal{S}(\mathbf{y}, S; \theta)\mathcal{S}(\mathbf{y}, S; \theta)^T\delta d\nu + \right. \\ &\quad \int 2p_1^{1/2}p_*^{1/2}\mathcal{S}(\mathbf{y}, S; \theta)^T\delta\delta^T R(\mathbf{y}, S; \theta^*)\delta d\nu + \\ &\quad \left.\int p_*\delta^T R(\mathbf{y}, S; \theta^*)\delta\delta^T R(\mathbf{y}, S; \theta^*)\delta d\nu\right]^{1/2}/2 \\ &=: (\mathrm{I} + \mathrm{II} + \mathrm{III})^{1/2}/2 \end{aligned} \tag{S32}$$

We define the minimum eigenvalue of $E_\theta\left(\mathcal{S}(\mathbf{y}, S; \theta)\mathcal{S}(\mathbf{y}, S; \theta)^T\right)$ is C_{min}, then

$$\mathrm{I} = \delta^T E_\theta\left(\mathcal{S}(\mathbf{y}, S; \theta)\mathcal{S}(\mathbf{y}, S; \theta)^T\right)\delta \geq C_{min}\|\delta\theta\|_2^2.$$

Since $\theta \in \mathcal{O}_d$ is defined on a compact set, then according to Lemma S8, there exists a positive constant $C_{low} > 0$, such that $C_{low} \leq C_{min}$ for any $\theta \in \mathcal{O}_d$, then we have

$$\mathrm{I} \geq C_{low}\|\delta\theta\|_2^2. \tag{S33}$$

For part II, using Cauchy inequallity, we have

$$\begin{aligned} |\mathrm{II}| &\leq E_\theta\left[\left|\mathcal{S}(\mathbf{y}, S; \theta)^T\delta\delta^T R(\mathbf{y}, S; \theta^*)\delta\right|\right] \\ &\quad + \int p_*\left|\mathcal{S}(\mathbf{y}, S; \theta)^T\delta\delta^T R(\mathbf{y}, S; \theta^*)\delta\right| d\nu. \end{aligned} \tag{S34}$$

Using Remark 3, there exist two polynomial functions $K_3(\mathbf{y}), K_4(\mathbf{y})$, such that

$$\left|\mathcal{S}(\mathbf{y}, S; \theta)^T \delta\right| \leq \|\mathcal{S}(\mathbf{y}, S; \theta)\|_1 \|\delta\|_2 \leq K_3(\mathbf{y}) \|\delta\|_2 \tag{S35}$$

and

$$\left|\delta^T R(\mathbf{y}, S; \theta^*) \delta\right| \leq \|R(\mathbf{y}, S; \theta^*)\|_2 \|\delta\|_2^2 \leq \|R(\mathbf{y}, S; \theta^*)\|_F \|\delta\|_2^2 \leq K_4(\mathbf{y}) \|\delta\|_2^2 . \tag{S36}$$

Using the fact that any PLN distribution have any bounded moments, we can derive that the first part of S34 can be bounded by $C \|\delta\theta\|_2^3$ with a constant C. Then using Remark 1, we know the second part of S34 satisfies

$$\int p_* \left|\mathcal{S}(\mathbf{y}, S; \theta)^T \delta \delta^T R(\mathbf{y}, S; \theta^*) \delta\right| d\nu \leq \int f_{up}(\mathbf{y}) K_3(\mathbf{y}) K_4(\mathbf{y}) \|\delta\|_2^3 d\nu \leq C' \|\delta\|_2^3$$

with a constant C'. Then we can derive

$$|\mathrm{II}| \leq (C + C') \|\delta\theta\|_2^3 . \tag{S37}$$

For $\mathrm{III} = \int p_* \left[\delta^T R(\mathbf{y}, S; \theta^*) \delta \delta^T R(\mathbf{y}, S; \theta^*) \delta\right] d\nu$, using (S36) and similar technique in II, we have

$$\mathrm{III} \leq C'' \|\delta\theta\|_2^4 \tag{S38}$$

with a constant C''.

Since the dominating functions f_{up}, K_3, K_4 are all independent from parameters, then the constants C, C', C'' are all independent from parameters, C_{low} is also independent from parameters, so the local constant K is also independent from parameters θ, θ'.

Combining above three inequalities (S33), (S37), (S38), we finish the proof.

S2.3.3 Proof of Lemma S10

Similar to Section S2.3.2, we have Taylor expansion of $p_4^{1/2}$ on θ_3,

$$p_4^{1/2} = p_3^{1/2} + p_*^{-1/2} \left(\frac{\partial p_*}{\partial \theta^*}\right)^T (\theta_4 - \theta_3)/2, \tag{S39}$$

where θ^* is between θ_3 and θ_4, and $p_* = p(\mathbf{y}, S; \theta^*)$. Let $\delta = \theta_4 - \theta_3$, we have,

$$\left|p_4^{1/2} - p_3^{1/2}\right| = p_*^{1/2} \left|\mathcal{S}(\mathbf{y}, S; \theta^*)^T (\delta)\right| / 2. \tag{S40}$$

Using Lemma 3, similar to (S35), we have polynomial function $K_3(\mathbf{y}_{sub})$, such that

$$\left|\mathcal{S}(\mathbf{y}, S; \theta^*)^T (\delta\theta)\right| \leq K_3(\mathbf{y}) \|\delta\theta\|_2 .$$

According to Remark 1, $p_* \leq f_{up}(\mathbf{y})$. So we have

$$\left|p_4^{1/2} - p_3^{1/2}\right| \leq f_{up}^{1/2}(\mathbf{y}) K_3(\mathbf{y}) \|\delta\|_2 .$$

Remark 1 also shows for any polynomial function K, $\int f_{up}(\mathbf{y}) K(\mathbf{y}) d\mathbf{y} < \infty$, so for polynomial function K_3^2, $\int f_{up}(\mathbf{y}) K_3^2(\mathbf{y}) d\mathbf{y} < \infty$, so let $m = f_{up}^{1/2} K_3$, we finish the proof.

S2.3.4 Proof of Lemma S11

We first check a basic result in our finite dimensional $\mathcal{F}$: There exist constants c_4, c_5, such that,

$$H\left(u, \mathcal{F} \cap \left\{\left\|p_2^{1/2} - p_1^{1/2}\right\|_2^2 \leq 2s^2\right\}\right) \leq c_4 \log(c_5 s/u). \tag{S41}$$

We notice that the parameters which index in $\mathcal{F} \cap \left\{\left\|p_2^{1/2} - p_1^{1/2}\right\|_2^2 \leq 2s^2\right\}$ can be covered by $\mathcal{F}_{\theta,s} = \left\{\theta' \middle| \|\theta' - \theta\|_2^2 \leq 2s^2/C_1^2\right\}$ using Lemma S9 and remark 4. Then it's easy to check that we can cover $\mathcal{F}_{\theta,s}$ using at most $\left(2\sqrt{2}C_2 s/C_1 u\right)^{d+d(d+1)/2}$ balls with radius $u/2C_2$. For any ball $\mathcal{B}$ with radius $u/2C_2$, we define the centre of $\mathcal{B}$ as θ_0, then using Lemma S10, for any $\theta' \in \mathcal{B}$, we have

$$\left|p^{1/2}(\mathbf{y}, S; \theta') - p^{1/2}(\mathbf{y}, S; \theta_0)\right| \le m(\mathbf{y})u/2C_2.$$

So we can choose the minimum and maximum density in each ball as the

$$f_L^{1/2} = [p^{1/2}(\mathbf{y}, S; \theta_0) - m(\mathbf{y})u/2C_2]_+,\ f_U^{1/2} = p^{1/2}(\mathbf{y}, S; \theta_0) + m(\mathbf{y})u/2C_2,$$

where $[f]_+$ takes value f when $f > 0$ and takes value 0 when $f \le 0$. Then we have $\left\|f_U^{1/2} - f_L^{1/2}\right\|_2^2 \le \int m^2(\mathbf{y})d\mathbf{y}u^2/C_2^2 = u^2$, then we can derive (S41).

Next, according to the remark (ii) of Theorem 2 in [4], we can get that there exists a constant c', such that for all $\epsilon \ge c'n^{-1/2}$, (S25) is satisfied, and we will prove it concretely next.

According to (S41), using Cauchy inequality $\left(\int_a^b f^{1/2}dx\right)^2 \le (b-a)\int_a^b f dx$, we just need to prove there exists c', such that for all $s \ge c'n^{-1/2}$,

$$\int_{s^2/2^8}^{\sqrt{2}s} \log(c_2c_5 s/u)du \le ns^3. \tag{S42}$$

After calculating the left hand of (S42), we have

$$\begin{aligned}
\int_{s^2/2^8}^{\sqrt{2}s} \log(c_2c_5 s/u)du &= \log(c_2c_5s)(\sqrt{2}s - s^2/2^8) + [u - u\log u]|_{s^2/2^8}^{\sqrt{2}s} \\
&= \log(c_2c_5s)(\sqrt{2}s - s^2/2^8) + (\sqrt{2}s - s^2/2^8) + (s^2/2^8\log(s^2/2^8) - \sqrt{2}s\log(\sqrt{2}s)) \\
&\le -\sqrt{2}s\log(\sqrt{2}) - s^2/2^8(\log(s) - \log(s^2/2^8)) + (\sqrt{2}s - s^2/2^8) + \log(c_2c_5)\sqrt{2}s \\
&= \sqrt{2}\log(c_2c_5e/\sqrt{2})s - s^2/2^8\log(2^8e/s).
\end{aligned} \tag{S43}$$

Since we only need to consider the s such that $\sqrt{2}s \ge s^2/2^8$, which means $\log(2^8e/s) > 0$, combining (S42) and (S43), we just need $ns^2 \ge \sqrt{2}\log(c_2c_5e/\sqrt{2})$, that equals to $s \ge c'n^{1/2}$, where $c' = \left(\sqrt{2}\log(c_2c_5e/\sqrt{2})\right)^{1/2}$. So we finish the proof of (S42).

Then using Lemma S5, we have

$$pr\left(\left\|p^{1/2}(\mathbf{y}, S; \hat{\theta}) - p^{1/2}(\mathbf{y}, S; \theta)\right\|_{L_2} \ge \epsilon\right) \le 5\exp(-c_1n\epsilon^2),$$

for a constant c_1 and any $\epsilon \ge c'n^{-1/2}$. Notice that for $0 < \epsilon < c'n^{1/2}$, we can have a constant c_0 to satisfy

$$pr\left(\left\|p^{1/2}(\mathbf{y}, S; \hat{\theta}) - p^{1/2}(\mathbf{y}, S; \theta)\right\|_{L_2} \ge \epsilon\right) \le 5\exp(-c_0n\epsilon^2).$$

Choosing constant $c = \min\{c_0, c_1\}$, we have for any $\epsilon > 0$,

$$pr\left(\left\|p^{1/2}(\mathbf{y}, S; \hat{\theta}) - p^{1/2}(\mathbf{y}, S; \theta)\right\|_{L_2} \ge \epsilon\right) \le 5\exp(-cn\epsilon^2).$$

Defining $h(\hat{\theta}, \theta) = \left\|p^{1/2}(\mathbf{y}, S; \hat{\theta}) - p^{1/2}(\mathbf{y}, S; \theta)\right\|_{L_2}$, notice that $\theta, \hat{\theta} \in \mathcal{O}_d$, according to Lemma S9 and remark 4, we have

$$\left\{h(\hat{\theta}, \theta) < \epsilon\right\} \subseteq \left\{\left\|\hat{\theta} - \theta\right\|_2 < \epsilon/C_1\right\},$$

so

$$pr\left(\left\|\hat{\theta} - \theta\right\|_2 < \epsilon/C_1\right) \ge pr\left(h(\hat{\theta}, \theta) < \epsilon\right) \ge 1 - 5\exp(-cn\epsilon^2),$$

and then $pr\left(\left\|\hat{\theta} - \theta\right\|_2 \ge \epsilon'\right) \le 5\exp(-Cn\epsilon'^2)$ for a constant C for all $\epsilon' > 0$. Then we finish our proof.

S2.3.5 Proof of Theorem 1 in the manuscript

Using the bounded condition in the manuscript and the Cauchy interlace theorem, we know that all the parameters of 1-dimensional or 2-dimensional marginal distributions of $PLN(S; \boldsymbol{\mu}^*, \Sigma^*)$ is restricted on the bounded sets $\mathcal{O}_1$ and $\mathcal{O}_2$. So according to the definition of $\widetilde{\Sigma}$ in equation (S23) and (S24), using Lemma S11, we have $pr\left(\left|\widetilde{\sigma}_{ij} - \sigma^*_{ij}\right| \geq \epsilon\right) \leq 5\exp(-Cn\epsilon^2)$ for any $i, j \in 1, ..., p$ and $\epsilon > 0$. Thus we have $pr\left(\left\|\widetilde{\Sigma} - \Sigma^*\right\|_\infty \geq \epsilon\right) \leq 5p^2\exp(-Cn\epsilon^2)$ for any $\epsilon > 0$. Similar to the proof of Lemma S1, we have $pr\left(\left\|\widehat{\Sigma} - \Sigma^*\right\|_\infty \geq \epsilon\right) \leq Ap^2\exp(-Bn\epsilon^2)$ for any $\epsilon > 0$ with constants A, B, thus we finish the proof.

S2.3.6 Proof of Theorem 2, 3 in the manuscript

Similar to the proof of Theorem S2, S3, we define

$$\breve{\Theta} = \arg\min_{A=A^T} \frac{1}{2}\mathrm{tr}\left(\widehat{\Sigma}A^2\right) - \mathrm{tr}\left(A\right) + \lambda\left\|A\right\|_{1,\mathrm{off}}.$$

Let

$$\begin{aligned}\epsilon = \quad & 1/\max\Bigg[12dk_\Gamma,\ 12\gamma^{-1}(k_\Sigma k_\Gamma^2 + k_\Gamma),\ \left\{12\gamma^{-1}\left(k_\Sigma k_\Gamma^3 + k_\Gamma^2\right) + 5dk_\Gamma^2\right\}\theta_{\min}^{-1}, \\ & \min\left\{s^{1/2}, d+1\right\}\left\{12\gamma^{-1}\left(k_\Sigma k_\Gamma^3 + k_\Gamma^2\right) + 5dk_\Gamma^2\right\}\lambda_{\min}^{-1}(\Theta^*)\Bigg].\end{aligned}$$

For $\eta > 2$, let $n_f = B^{-1}(\eta\log p + \log A)\epsilon^{-2}$ and $\epsilon_f = B^{-1/2}(\eta\log p + \log A)^{1/2}n^{-1/2}$. Combining with Theorem S1, the rest of the proof is the same as in Section S1.3.6.

S3 Simulation

S3.1 Details of the simluation settings

We simulate count data from the PLN model. The library sizes are generated from a log-normal distribution N $\left(\log 10, 0.1^2\right)$. The mean vector μ of PLN model is set as $\mu = (-2.35, \ldots, -2.35)^T$ or $\mu = (-3.2, \ldots, -3.2)^T$, where the former corresponds to a low-dropout scenario (about 40 percent of the counts are zeros) and the latter to a high-dropout scenario (about 60 percent of the counts are zeros).

S3.2 Simulation results

S3.2.1 The area under the receiver operating characteristic curve (AUC)

Table S1: Comparisons of PLNet, VPLN, glasso, and PLNet-MOM in terms of the area under the receiver operating characteristic curve (AUC) on simulation results for $n = 500$. The results are averages over 100 replicates with standard deviations in brackets.

Sample size	$n = 500$		$n = 500$		$n = 500$	
Dimension	$p = 100$		$p = 300$		$p = 500$	
Dropout	Low	High	Low	High	Low	High
			Banded graph			
PLNet	**0.97 (0.01)**	**0.89 (0.02)**	**0.96 (0.01)**	**0.87 (0.01)**	0.95 (0.01)	0.86 (0.01)
PLNet-MOM	0.93 (0.01)	0.85 (0.02)	0.92 (0.01)	0.84 (0.01)	0.91 (0.01)	0.83 (0.01)
VPLN	0.96 (0.01)	0.88 (0.01)	0.96 (0.01)	0.87 (0.01)	**0.96 (0.01)**	**0.87 (0.01)**
glasso	0.88 (0.01)	0.61 (0.02)	0.94 (0.01)	0.79 (0.01)	0.95 (0.01)	0.82 (0.01)
			Random graph			
PLNet	0.93 (0.02)	0.85 (0.03)	0.94 (0.01)	0.82 (0.02)	0.94 (0.01)	0.81 (0.02)
PLNet-MOM	0.91 (0.02)	0.83 (0.03)	0.91 (0.01)	0.8 (0.02)	0.91 (0.01)	0.79 (0.02)
VPLN	**0.95 (0.02)**	**0.88 (0.03)**	**0.95 (0.01)**	**0.84 (0.03)**	**0.95 (0.01)**	**0.83 (0.03)**
glasso	0.88 (0.03)	0.68 (0.03)	0.94 (0.01)	0.78 (0.02)	0.94 (0.01)	0.79 (0.02)
			Scale-free Graph			
PLNet	0.94 (0.04)	0.87 (0.03)	**0.95 (0.01)**	0.87 (0.01)	**0.95 (0.01)**	0.86 (0.02)
PLNet-MOM	0.91 (0.04)	0.85 (0.03)	0.91 (0.01)	0.83 (0.02)	0.9 (0.01)	0.82 (0.01)
VPLN	**0.95 (0.01)**	**0.88 (0.03)**	0.95 (0.01)	**0.88 (0.01)**	0.95 (0.01)	**0.88 (0.02)**
glasso	0.91 (0.02)	0.77 (0.07)	0.95 (0.01)	0.86 (0.01)	0.95 (0.01)	0.86 (0.01)
			Blocked graph			
PLNet	0.93 (0.02)	0.84 (0.04)	0.94 (0.02)	0.81 (0.03)	0.93 (0.02)	0.81 (0.02)
PLNet-MOM	0.9 (0.02)	0.82 (0.04)	0.91 (0.02)	0.79 (0.03)	0.9 (0.02)	0.79 (0.02)
VPLN	**0.94 (0.02)**	**0.87 (0.03)**	**0.95 (0.02)**	**0.83 (0.03)**	**0.94 (0.02)**	**0.82 (0.02)**
glasso	0.86 (0.02)	0.68 (0.03)	0.94 (0.02)	0.77 (0.03)	0.93 (0.02)	0.78 (0.02)

Table S2: Comparisons of PLNet, VPLN, glasso, and PLNet-MOM in terms of the area under the receiver operating characteristic curve (AUC) on simulation results for $n = 2000$. The results are averages over 100 replicates with standard deviations in brackets.

Sample size	$n = 2000$		$n = 2000$		$n = 2000$	
Dimension	$p = 100$		$p = 300$		$p = 500$	
Dropout	Low	High	Low	High	Low	High
			Banded graph			
PLNet	**1 (0.01)**	**1 (0.01)**	**1 (0.01)**	**1 (0.01)**	**1 (0.01)**	**1 (0.01)**
PLNet-MOM	1 (0.01)	0.99 (0.01)	1 (0.01)	0.99 (0.01)	1 (0.01)	0.99 (0.01)
VPLN	1 (0.01)	0.99 (0.01)	1 (0.01)	0.99 (0.01)	1 (0.01)	0.99 (0.01)
glasso	0.97 (0.01)	0.53 (0.01)	1 (0.01)	0.87 (0.01)	1 (0.01)	0.93 (0.01)
			Random graph			
PLNet	**1 (0.01)**	**0.98 (0.01)**	**1 (0.01)**	**0.98 (0.01)**	**1 (0.01)**	**0.98 (0.01)**
PLNet-MOM	0.99 (0.01)	0.97 (0.01)	1 (0.01)	0.97 (0.01)	1 (0.01)	0.97 (0.01)
VPLN	0.99 (0.01)	0.98 (0.01)	1 (0.01)	0.98 (0.01)	1 (0.01)	0.98 (0.01)
glasso	0.94 (0.02)	0.65 (0.02)	1 (0.01)	0.87 (0.02)	1 (0.01)	0.92 (0.01)
			Scale-free Graph			
PLNet	**0.98 (0.03)**	**0.98 (0.03)**	**1 (0.01)**	**0.99 (0.01)**	**1 (0.01)**	**0.99 (0.01)**
PLNet-MOM	0.98 (0.03)	0.97 (0.03)	0.99 (0.01)	0.97 (0.01)	0.99 (0.01)	0.97 (0.01)
VPLN	0.97 (0.06)	0.96 (0.04)	0.99 (0.01)	0.98 (0.01)	0.99 (0.01)	0.98 (0.01)
glasso	0.89 (0.1)	0.78 (0.06)	0.99 (0.01)	0.92 (0.01)	0.99 (0.01)	0.95 (0.01)
			Blocked graph			
PLNet	**0.99 (0.01)**	**0.97 (0.02)**	**1 (0.01)**	**0.98 (0.01)**	**1 (0.01)**	**0.98 (0.01)**
PLNet-MOM	0.98 (0.01)	0.95 (0.03)	1 (0.01)	0.97 (0.02)	1 (0.01)	0.97 (0.01)
VPLN	0.99 (0.01)	0.97 (0.01)	1 (0.01)	0.98 (0.01)	1 (0.01)	0.98 (0.01)
glasso	0.91 (0.02)	0.65 (0.03)	0.99 (0.01)	0.86 (0.02)	1 (0.01)	0.91 (0.01)

S3.2.2 The mean predicted networks tunned by BIC criterion

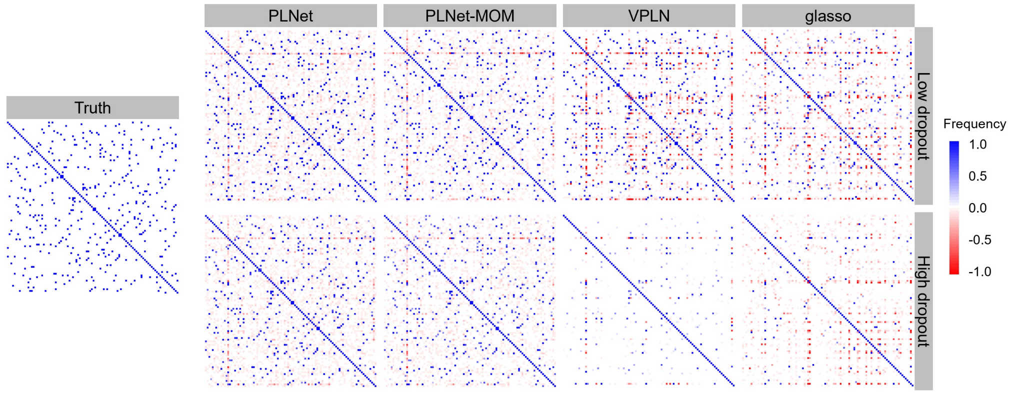


Figure S1: The mean networks predicted by PLNet, VPLN, glasso and PLNet-MOM for the random graph with 100 nodes and n = 2000. False edges are colored in red and true edges are in blue. The left panel is the true network matrix for reference.

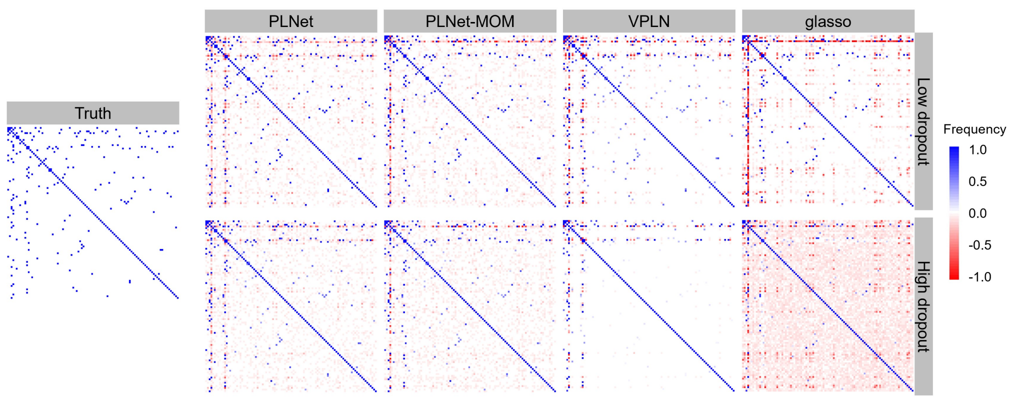


Figure S2: The mean networks predicted by PLNet, VPLN, glasso and PLNet-MOM for the Scale-free graph with 100 nodes and n = 2000. False edges are colored in red and true edges are in blue. The left panel is the true network matrix for reference.

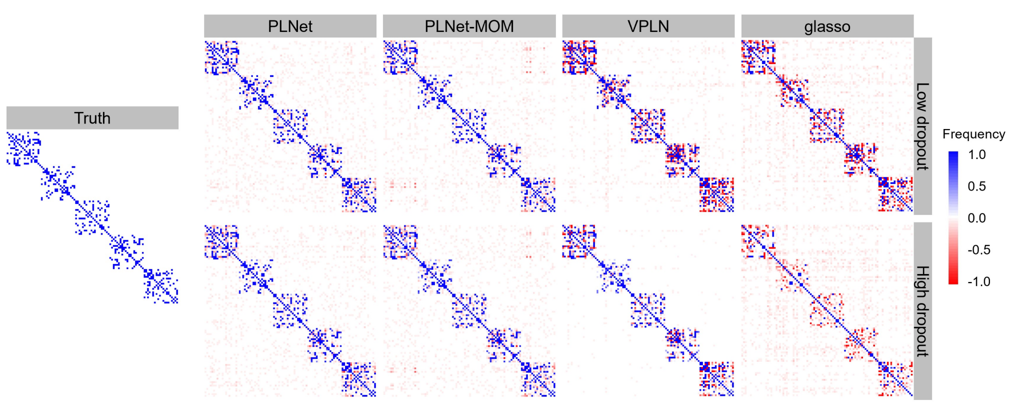


Figure S3: The mean networks predicted by PLNet, VPLN, glasso and PLNet-MOM for the Blocked graph with 100 nodes and n = 2000. False edges are colored in red and true edges are in blue. The left panel is the true network matrix for reference.

S3.2.3 Time cost

Table S3: Comparisons of PLNet, VPLN, glasso, and PLNet-MOM in terms of CPU time (minute) for $n = 500$. The results are averages over 100 replicates with standard deviations in brackets.

Sample size	$n = 500$		$n = 500$		$n = 500$	
Dimension	$p = 100$		$p = 300$		$p = 500$	
Dropout	Low	High	Low	High	Low	High
Banded graph						
PLNet	0.83 (0.06)	0.51 (0.06)	7.69 (1.65)	7.23 (1.03)	53.05 (3.95)	40.65 (3.18)
PLNet-MOM	0.64 (0.04)	0.49 (0.08)	7.75 (0.74)	6.8 (0.58)	40.45 (2.57)	**25.67 (1.78)**
VPLN	1.72 (0.07)	1.32 (0.04)	10.24 (0.33)	9.04 (0.24)	**30.94 (0.67)**	25.92 (0.82)
glasso	**0.04 (0.01)**	**0.04 (0.01)**	**0.65 (0.01)**	**0.93 (0.05)**	225.67 (53.83)	210.47 (47.47)
Random graph						
PLNet	0.74 (0.07)	0.56 (0.06)	9.71 (1.7)	8.31 (1.44)	40 (4.71)	37.88 (4.07)
PLNet-MOM	0.43 (0.04)	0.46 (0.07)	9.26 (1.32)	7.76 (0.87)	37.49 (5.85)	28.94 (5.82)
VPLN	1.73 (0.07)	1.57 (0.03)	12.43 (0.41)	9.58 (0.33)	**28.12 (1.05)**	**27.8 (2.17)**
glasso	**0.03 (0.01)**	**0.04 (0.01)**	**0.48 (0.04)**	**0.68 (0.08)**	217.91 (65.44)	203.17 (43.5)
Scale-free Graph						
PLNet	0.42 (0.06)	0.51 (0.08)	9.79 (1.68)	8.16 (1.25)	41.59 (3.14)	53.23 (3.47)
PLNet-MOM	0.31 (0.09)	0.43 (0.1)	9.48 (1.36)	7.65 (1.16)	38.2 (4.07)	49.73 (6.19)
VPLN	1.5 (0.06)	1.61 (0.16)	12.43 (0.42)	9.27 (0.58)	**27.16 (0.7)**	**29.99 (1.11)**
glasso	**0.03 (0.01)**	**0.07 (0.11)**	**0.44 (0.02)**	**0.53 (0.06)**	210.67 (56.82)	251.6 (64.76)
Blocked graph						
PLNet	0.45 (0.08)	0.42 (0.07)	11.03 (1.66)	8.24 (1.43)	55.13 (3.82)	50.26 (6.61)
PLNet-MOM	0.31 (0.03)	0.32 (0.04)	10.2 (1.4)	8.09 (1.1)	51.9 (6.02)	42.75 (5.61)
VPLN	1.57 (0.08)	1.35 (0.04)	12.01 (0.49)	9.47 (0.38)	**32.28 (1.37)**	**32.78 (2.43)**
glasso	**0.03 (0.01)**	**0.03 (0.01)**	**0.46 (0.04)**	**0.59 (0.09)**	212.82 (59.03)	223.43 (61.7)

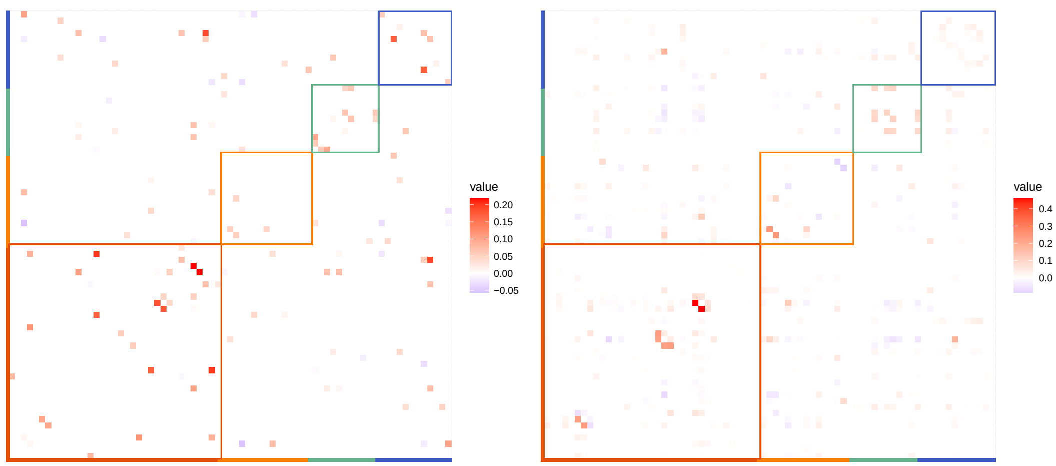
Table S4: Comparisons of PLNet, VPLN, glasso, and PLNet-MOM in terms of CPU time (minute) for $n = 2000$. The results are averages over 100 replicates with standard deviations in brackets.

Sample size	$n = 2000$		$n = 2000$		$n = 2000$	
Dimension	$p = 100$		$p = 300$		$p = 500$	
Dropout	Low	High	Low	High	Low	High
Banded graph						
PLNet	0.51 (0.02)	0.66 (0.06)	15.59 (0.22)	11.84 (1.17)	65.86 (8.41)	46.93 (5.62)
PLNet-MOM	0.61 (0.05)	0.63 (0.05)	11.93 (0.77)	10.16 (0.81)	46.44 (3.66)	41.33 (3.64)
VPLN	7.67 (1.7)	6.26 (0.26)	73.21 (18.63)	56.19 (6.96)	312.78 (137.26)	157.38 (15.68)
glasso	**0.06 (0.01)**	**0.08 (0.01)**	**0.8 (0.02)**	**1.14 (0.06)**	**3.37 (0.29)**	**5.44 (0.42)**
Random graph						
PLNet	0.39 (0.03)	0.44 (0.06)	14.18 (1.03)	10.68 (1.17)	70.48 (5.93)	46.95 (1.87)
PLNet-MOM	0.37 (0.06)	0.37 (0.06)	9.36 (1.19)	9.57 (1.28)	46.92 (8.44)	44.1 (3.85)
VPLN	5.99 (0.31)	5.17 (0.22)	62.57 (7.28)	57.63 (5.83)	166.45 (20)	166.18 (8.07)
glasso	**0.06 (0.01)**	**0.07 (0.01)**	**0.65 (0.04)**	**0.8 (0.07)**	**2.19 (0.21)**	**3.12 (0.4)**
Scale-free Graph						
PLNet	0.37 (0.1)	0.45 (0.06)	10.25 (1.27)	8.79 (1.16)	49.97 (7.6)	42.72 (4.16)
PLNet-MOM	0.41 (0.11)	0.38 (0.08)	7.64 (0.98)	7.74 (0.89)	42.96 (5.89)	45.48 (6.31)
VPLN	12.93 (25.14)	6.39 (0.59)	66.16 (4.88)	58.62 (5.72)	188.59 (22.95)	154.42 (22.21)
glasso	**0.07 (0.06)**	**0.08 (0.06)**	**0.61 (0.06)**	**0.69 (0.03)**	**1.98 (0.12)**	**2.64 (0.21)**
Blocked graph						
PLNet	0.41 (0.04)	0.44 (0.06)	13.48 (1.51)	9.97 (1.1)	68.83 (7.3)	45.94 (3)
PLNet-MOM	0.37 (0.02)	0.36 (0.03)	7.68 (1.22)	8.64 (1.34)	44.01 (7.92)	41.92 (5.41)
VPLN	7.27 (0.27)	6.26 (0.22)	62.42 (8.5)	57.81 (6.53)	176.25 (19.71)	161.61 (14.55)
glasso	**0.06 (0.01)**	**0.06 (0.01)**	**0.61 (0.04)**	**0.71 (0.05)**	**2.24 (0.34)**	**2.92 (0.32)**

Table S3,S4 show the computational time of the four algorithms under the computer configuration of Linux OS, Intel(R) Xeon(R) Gold 6132 CPU @ 2.60GHz and 10G RAM. The glasso method is computationally the most efficient since its optimization problem is much simpler than that of PLNet and VPLN. PLNet is computationally more efficient than VPLN, sometimes by a very large amount. For example, when $n = 2000, p = 100$, for the scale-free graph under the low dropout scenario, the mean computational time of PLNet is 0.37 minutes, which is only about 2.8% of the VPLN's computational time (12.93 minutes). Interestingly, we observe that VPLN generally takes much more time for large sample size cases ($n = 2000, p = 100$) than the small sample size cases ($n = 500, p = 100$). In comparison, the computational efficiency of PLNet is roughly the same for different sample size. VPLN is computationally less efficient since VPLN involves a series of glasso optimizations as well as the variational approximation for each sample.

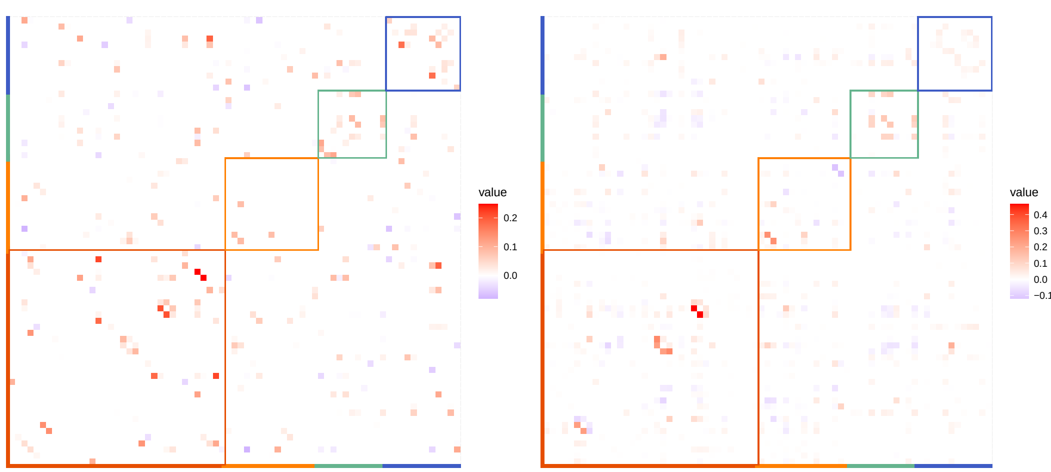
S4 Real data analysis

S4.1 Other densities results



(a) PLNet (b) VPLN

Figure S4: Heat maps of partial correlations between genes in the 4 GO modules given by PLNet (a) and VPLN (b) (density = 3%). Red: Cytokine-mediated signaling pathway (Module M_1); Orange: Neutrophil-mediated immunity (Module M_2); Green: Cellular protein metabolic process (Module M_3); Blue: Proteolysis (Module M_4).



(a) PLNet (b) VPLN

Figure S5: Heat maps of partial correlations between genes in the 4 GO modules given by PLNet (a) and VPLN (b) (density = 7%). Red: Cytokine-mediated signaling pathway (Module M_1); Orange: Neutrophil-mediated immunity (Module M_2); Green: Cellular protein metabolic process (Module M_3); Blue: Proteolysis (Module M_4).

Table S5: The within-between connection ratios of the 4 modules in the networks estimated by PLNet and VPLN tuned such that the network densities are around 3%.

Type	Method	Module 1	Module 2	Module 3	Module 4
Weighted	PLNet	**0.724**	0.273	**0.668**	**0.393**
	VPLN	0.648	**0.413**	0.505	0.276
Unweighted	PLNet	**0.567**	0.211	**0.522**	**0.308**
	VPLN	0.506	**0.234**	0.200	0.299

Table S6: The within-between connection ratios of the 4 modules in the networks estimated by PLNet and VPLN tuned such that the network densities are around 7%.

Type	Method	Module 1	Module 2	Module 3	Module 4
Weighted	PLNet	**0.651**	0.202	**0.520**	**0.369**
	VPLN	0.606	**0.362**	0.433	0.243
Unweighted	PLNet	0.483	0.083	**0.333**	**0.308**
	VPLN	**0.520**	**0.215**	0.177	0.180

References

[1] John Riordan. Moment Recurrence Relations for Binomial, Poisson and Hypergeometric Frequency Distributions. *The Annals of Mathematical Statistics*, 8(2):103–111, 1937.

[2] Haskell P. Rosenthal. On the subspaces of L_p $(p > 2)$ spanned by sequences of independent random variables. *Israel Journal of Mathematics*, 8(3):273–303, 1970.

[3] Junjie Tang, Changhu Wang, Feiyi Xiao, and Ruibin Xi. Single-cell gene regulatory network analysis for mixed cell populations with applications to COVID-19 single cell data. *arXiv e-prints*, art. arXiv:2205.11274, May 2022.

[4] Wing Hung Wong and Xiaotong Shen. Probability Inequalities for Likelihood Ratios and Convergence Rates of Sieve MLES. *The Annals of Statistics*, 23(2):339 – 362, 1995.

[5] Teng Zhang and Hui Zou. Sparse precision matrix estimation via lasso penalized D-trace loss. *Biometrika*, 101(1):103–120, 2014.