
Supplementary Materials for A Flexible Latent Space Model for Multilayer Networks

Xuefei Zhang¹ Songkai Xue¹ Ji Zhu¹

Abstract

This supplementary material contains technical details and additional numerical results of the main article ‘A Flexible Latent Space Model for Multilayer Networks’. Section A provides the proof of Theorem 1 in the main article, which derives the upper bound of the overall maximum likelihood estimators of the model. The proof of Theorem 2, that builds the upper bound of the estimation error of shared latent variables U , is given in Section B. Section C presents the proof of Proposition 1 on identifiability conditions of the model. In Section D we provide additional simulation results under other settings of n , R and k to further support our theoretical discoveries.

A. Proof of Theorem 1

For any parameter $\mathcal{T} \in \mathcal{F}$, where the definition of \mathcal{F} is given in Section 4, the objective function is defined as

$$\begin{aligned}
 l(\mathcal{T}) &= -\log P(A|\mathcal{T}) \\
 &= -\sum_{r=1}^R \sum_{i=1}^n \sum_{j=1}^n \left\{ A_{ij}^{(r)} \Theta_{ij}^{(r)} + \log \left(1 - \sigma(\Theta_{ij}^{(r)}) \right) \right\} \\
 &= -\sum_{r=1}^R \sum_{i=1}^n \sum_{j=1}^n \left\{ A_{ij}^{(r)} \Theta_{ij}^{(r)} - b(\Theta_{ij}^{(r)}) \right\},
 \end{aligned} \tag{S1}$$

where $b(x) = \log(1 + \exp(x))$.

Denote $\mathcal{T}_* \in \mathcal{F}$ be the true parameter value, and $\widehat{\mathcal{T}}$ is obtained from (4), then

$$l(\widehat{\mathcal{T}}) - l(\mathcal{T}_*) \leq 0. \tag{S2}$$

Further, we have

$$\begin{aligned}
 &l(\mathcal{T}_*) - l(\widehat{\mathcal{T}}) \\
 &= \sum_{r=1}^R \sum_{i=1}^n \sum_{j=1}^n \left\{ A_{ij}^{(r)} \left(\widehat{\Theta}_{ij}^{(r)} - \Theta_{*,ij}^{(r)} \right) - \left(b(\widehat{\Theta}_{ij}^{(r)}) - b(\Theta_{*,ij}^{(r)}) \right) \right\} \\
 &= \sum_{r=1}^R \sum_{i=1}^n \sum_{j=1}^n \left(A_{ij}^{(r)} - b'(\Theta_{*,ij}^{(r)}) \right) \left(\widehat{\Theta}_{ij}^{(r)} - \Theta_{*,ij}^{(r)} \right) \\
 &\quad - \sum_{r=1}^R \sum_{i=1}^n \sum_{j=1}^n \left\{ b(\widehat{\Theta}_{ij}^{(r)}) - b(\Theta_{*,ij}^{(r)}) - b'(\Theta_{*,ij}^{(r)}) \left(\widehat{\Theta}_{ij}^{(r)} - \Theta_{*,ij}^{(r)} \right) \right\}.
 \end{aligned} \tag{S3}$$

¹Department of Statistics, University of Michigan, Ann Arbor, MI, USA. Correspondence to: Ji Zhu <jizhu@umich.edu>.

By Taylor's expansion, the last expression in (S3) can be expressed as

$$\begin{aligned} & \sum_{r=1}^R \sum_{i=1}^n \sum_{j=1}^n \left(A_{ij}^{(r)} - b'(\Theta_{*,ij}^{(r)}) \right) \left(\widehat{\Theta}_{ij}^{(r)} - \Theta_{*,ij}^{(r)} \right) - \sum_{r=1}^R \sum_{i=1}^n \sum_{j=1}^n \frac{1}{2} b''(\widetilde{\Theta}_{ij}^{(r)}) \left(\widehat{\Theta}_{ij}^{(r)} - \Theta_{*,ij}^{(r)} \right)^2 \\ & \leq \sum_{r=1}^R \sum_{i=1}^n \sum_{j=1}^n \left(A_{ij}^{(r)} - b'(\Theta_{*,ij}^{(r)}) \right) \left(\widehat{\Theta}_{ij}^{(r)} - \Theta_{*,ij}^{(r)} \right) - \frac{1}{2} \min_{|v| \leq \mu} b''(v) \|\widehat{\mathcal{T}} - \mathcal{T}_*\|_F^2, \end{aligned} \quad (\text{S4})$$

where $\widetilde{\Theta}_{ij}^{(r)} = \eta_{ij} \widehat{\Theta}_{ij}^{(r)} + (1 - \eta_{ij}) \Theta_{*,ij}^{(r)}$ for some $\eta_{ij} \in (0, 1)$. By (S2), (S3) and (S4), we have

$$\begin{aligned} \|\widehat{\mathcal{T}} - \mathcal{T}_*\|_F^2 & \leq \frac{2}{\min_{|v| < \mu} b''(v)} \sum_{r=1}^R \sum_{i=1}^n \sum_{j=1}^n \left(A_{ij}^{(r)} - b'(\Theta_{*,ij}^{(r)}) \right) \left(\widehat{\Theta}_{ij}^{(r)} - \Theta_{*,ij}^{(r)} \right) \\ & = \frac{2}{\min_{|v| < \mu} b''(v)} \langle Z, \widehat{\mathcal{T}} - \mathcal{T}_* \rangle. \end{aligned} \quad (\text{S5})$$

Here we define $Z = [Z^{(1)}; Z^{(2)}; \dots; Z^{(R)}] \in \mathbb{R}^{n \times n \times R}$ as a three-way tensor with entries $Z_{ij}^{(r)} = A_{ij}^{(r)} - b'(\Theta_{*,ij}^{(r)})$, for $i, j = 1, \dots, n, r = 1, \dots, R$.

For notational simplicity, we decompose each $\mathcal{T} \in \mathcal{F}$ into two parts: $\mathcal{T} = \mathcal{H} + \mathcal{M}$. Here, $\mathcal{H} = [\mathcal{H}^{(1)}; \mathcal{H}^{(2)}; \dots; \mathcal{H}^{(R)}] \in \mathbb{R}^{n \times n \times R}$ with $\mathcal{H}^{(r)} = \alpha^{(r)} \mathbf{1}_n^\top + \mathbf{1}_n \alpha^{(r)\top} \in \mathbb{R}^{n \times n}$ is the term related to node degree heterogeneity parameters, and $\mathcal{M} = [U\Lambda^{(1)}U^\top; U\Lambda^{(2)}U^\top; \dots; U\Lambda^{(R)}U^\top] \in \mathbb{R}^{n \times n \times R}$ is the term related to shared latent representations. Therefore, the quantity $\langle Z, \widehat{\mathcal{T}} - \mathcal{T}_* \rangle$ in (S5) can be decomposed as

$$\begin{aligned} & \langle Z, \widehat{\mathcal{T}} - \mathcal{T}_* \rangle \\ & = \langle Z, \widehat{\mathcal{H}} - \mathcal{H}_* \rangle + \langle Z, \widehat{\mathcal{M}} - \mathcal{M}_* \rangle \\ & = \sum_{r=1}^R \langle Z^{(r)}, \widehat{\alpha}^{(r)} \mathbf{1}_n^\top + \mathbf{1}_n \widehat{\alpha}^{(r)\top} - \alpha_*^{(r)} \mathbf{1}_n^\top - \mathbf{1}_n \alpha_*^{(r)\top} \rangle + \langle Z, \widehat{\mathcal{M}} - \mathcal{M}_* \rangle. \end{aligned} \quad (\text{S6})$$

We bound two summands in (S6) respectively. For any two matrices A and B , we have $|\langle A, B \rangle| \leq \|A\|_2 \|B\|_* \leq \|A\|_2 \sqrt{\text{rank}(B)} \|B\|_F$. The definition of Z implies that entries in Z are independent, mean-zero sub-gaussian random variables with $\mathbb{E}[\exp(tZ_{ij}^{(r)})] \leq \exp(t^2/8)$. Therefore, we can apply Lemma 1 (given in Section A.1) to $Z^{(r)}$ for any given r , and obtain that with probability at least $1 - \exp(-c_1 n)$, $\|Z^{(r)}\|_2 \leq C'_1 \sqrt{n}$, for absolute constants c_1 and C'_1 . Then with probability at least $1 - R \exp(-c_1 n)$, we have $\max_r (\|Z^{(r)}\|_2) \leq C'_1 \sqrt{n}$. Thus, with probability greater than $1 - R \exp(-c_1 n)$, the first term in (S6) can be bounded as

$$\begin{aligned} & \sum_{r=1}^R \langle Z^{(r)}, \widehat{\alpha}^{(r)} \mathbf{1}_n^\top + \mathbf{1}_n \widehat{\alpha}^{(r)\top} - \alpha_*^{(r)} \mathbf{1}_n^\top - \mathbf{1}_n \alpha_*^{(r)\top} \rangle \\ & \leq 2 \sum_{r=1}^R \|Z^{(r)}\|_2 \|\widehat{\alpha}^{(r)} \mathbf{1}_n^\top + \mathbf{1}_n \widehat{\alpha}^{(r)\top} - \alpha_*^{(r)} \mathbf{1}_n^\top - \mathbf{1}_n \alpha_*^{(r)\top}\|_F \\ & \leq 2C'_1 \sqrt{n} \sum_{r=1}^R \|\widehat{\alpha}^{(r)} \mathbf{1}_n^\top + \mathbf{1}_n \widehat{\alpha}^{(r)\top} - \alpha_*^{(r)} \mathbf{1}_n^\top - \mathbf{1}_n \alpha_*^{(r)\top}\|_F \\ & = 2C'_1 \sqrt{n} \sum_{r=1}^R \|\widehat{\mathcal{H}}^{(r)} - \mathcal{H}_*^{(r)}\|_F \leq 2C'_1 \sqrt{n} \sqrt{R} \|\widehat{\mathcal{H}} - \mathcal{H}_*\|_F. \end{aligned} \quad (\text{S7})$$

The first inequality in (S7) is by the fact that $\text{rank}(\widehat{\alpha}^{(r)} \mathbf{1}_n^\top + \mathbf{1}_n \widehat{\alpha}^{(r)\top} - \alpha_*^{(r)} \mathbf{1}_n^\top - \mathbf{1}_n \alpha_*^{(r)\top}) \leq 4$.

Next we bound the second term in (S6). For any two three-way tensors A and B with same dimensions $n_1 \times n_2 \times n_3$, we have $|\langle A, B \rangle| \leq \|A\|_2 \|B\|_*$. The nuclear norm of B , $\|B\|_*$ is further bounded by $\sqrt{r_1 r_2} \|B\|_F$, where r_1 is the rank of the

matrix that stacks B along its first mode into a matrix of size $n_1 \times (n_2 n_3)$, and similarly, r_2 is the rank of the matrix that stacks B along its second mode into a matrix of size $n_2 \times (n_1 n_3)$ (Wang et al., 2017; Wang & Li, 2018). For any tensor $\mathcal{M} = [U\Lambda^{(1)}U^\top; U\Lambda^{(2)}U^\top; \dots; U\Lambda^{(R)}U^\top] \in \mathbb{R}^{n \times n \times R}$, stacking it along its first mode, we could obtain an $n \times nR$ matrix $\mathcal{M}_1 = [U\Lambda^{(1)}U^\top \ U\Lambda^{(2)}U^\top \ \dots \ U\Lambda^{(R)}U^\top]$. Since

$$\mathcal{M}_1 = [U\Lambda^{(1)}U^\top \ U\Lambda^{(2)}U^\top \ \dots \ U\Lambda^{(R)}U^\top] = U[\Lambda^{(1)}U^\top \ \Lambda^{(2)}U^\top \ \dots \ \Lambda^{(R)}U^\top],$$

so rank of \mathcal{M}_1 is k . Since each layer in \mathcal{M} is symmetric, stacking the tensor along its second mode similarly yields a rank k matrix. This leads to $\|\widehat{\mathcal{M}} - \mathcal{M}\|_* \leq \sqrt{2k \cdot 2k} \|\widehat{\mathcal{M}} - \mathcal{M}\|_F = 2k \|\widehat{\mathcal{M}} - \mathcal{M}\|_F$. Additionally, applying Lemma 1 to Z , we have with probability at least $1 - \exp(-c_2(2n + R))$, $\|Z\|_2 \leq C'_2 \sqrt{2n + R}$ for absolute constants c_2 and C'_2 . Therefore, the second term in (S6) can be bounded as

$$\left\langle Z, \widehat{\mathcal{M}} - \mathcal{M}_* \right\rangle \leq \|Z\|_2 \|\widehat{\mathcal{M}} - \mathcal{M}_*\|_* \leq 2k \|Z\|_2 \|\widehat{\mathcal{M}} - \mathcal{M}_*\|_F \leq 2k C'_2 \sqrt{2n + R} \|\widehat{\mathcal{M}} - \mathcal{M}_*\|_F \quad (\text{S8})$$

with probability at least $1 - \exp(-c_2(2n + R))$.

Plugging (S6), (S7) and (S8) into (S5) yields

$$\begin{aligned} \|\widehat{\mathcal{T}} - \mathcal{T}_*\|_F^2 &\leq \frac{2}{\min_{|v| < \mu} b''(v)} \left\langle Z, \widehat{\mathcal{T}} - \mathcal{T}_* \right\rangle \\ &\leq \frac{2}{\min_{|v| < \mu} b''(v)} \left(2C'_1 \sqrt{nR} \|\widehat{\mathcal{H}} - \mathcal{H}_*\|_F + 2k C'_2 \sqrt{2n + R} \|\widehat{\mathcal{M}} - \mathcal{M}_*\|_F \right). \end{aligned} \quad (\text{S9})$$

Lemma 2 in Section A.1 implies that $\|\widehat{\mathcal{H}} - \mathcal{H}_*\|_F \leq \|\widehat{\mathcal{T}} - \mathcal{T}_*\|_F$, so does $\|\widehat{\mathcal{M}} - \mathcal{M}_*\|_F \leq \|\widehat{\mathcal{T}} - \mathcal{T}_*\|_F$. Dividing both sides in (S9) by $\|\widehat{\mathcal{T}} - \mathcal{T}_*\|_F$ leads to

$$\|\widehat{\mathcal{T}} - \mathcal{T}_*\|_F \leq \frac{2}{\min_{|v| < \mu} b''(v)} \left(2C'_1 \sqrt{nR} + 2k C'_2 \sqrt{2n + R} \right). \quad (\text{S10})$$

Taking the square of both sides, we can conclude that with probability at least $1 - R \exp(-c_1 n) - \exp(-c_2(2n + R))$, we have

$$\|\widehat{\mathcal{T}} - \mathcal{T}_*\|_F^2 \leq C_1 nR + C_2 k^2 (2n + R), \quad (\text{S11})$$

where $C_1 = 32(C'_1)^2 / (\min_{|v| < \mu} b''(v))^2$ and $C_2 = 32(C'_2)^2 / (\min_{|v| < \mu} b''(v))^2$.

A.1. Lemmas for Theorem 1

This subsection includes lemmas that are used in the proof of Theorem 1.

Lemma 1 (Theorem 1 in Tomioka & Suzuki (2014)). *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_K}$ be a K -way tensor. Assume each element $\mathcal{X}_{i_1 \dots i_K}$ is independent, zero mean and satisfies $\mathbb{E}[e^{t\mathcal{X}_{i_1 \dots i_K}}] \leq \exp(\sigma^2 t^2 / 2)$. Then there exist constants c and C which only depend on σ^2 and K such that with probability at least $1 - \exp(-c \sum_k n_k)$, the spectral norm of \mathcal{X} is bounded by $\|\mathcal{X}\|_2 \leq C \sqrt{\sum_{k=1}^K n_k}$.*

Lemma 2. *For $\widehat{\mathcal{T}}, \mathcal{T}_* \in \mathcal{F}$, when decomposing $\widehat{\mathcal{T}} = \widehat{\mathcal{H}} + \widehat{\mathcal{M}}$ and $\mathcal{T}_* = \mathcal{H}_* + \mathcal{M}_*$, we have the following identity:*

$$\|\widehat{\mathcal{T}} - \mathcal{T}_*\|_F^2 = \|\widehat{\mathcal{H}} - \mathcal{H}_*\|_F^2 + \|\widehat{\mathcal{M}} - \mathcal{M}_*\|_F^2.$$

Proof. Since for any $\mathcal{T} \in \mathcal{F}$ we require $JU = U$, then $U^\top \mathbf{1}_n = 0$. Then we have

$$\widehat{\mathcal{M}}^{(r)} \mathbf{1}_n = \mathcal{M}_*^{(r)} \mathbf{1}_n = 0,$$

and

$$\mathbf{1}_n^\top \widehat{\mathcal{M}}^{(r)} = \mathbf{1}_n^\top \mathcal{M}_*^{(r)} = 0$$

for $r = 1, \dots, R$. Therefore,

$$\|\widehat{\Theta}^{(r)} - \Theta_*^{(r)}\|_F^2 = \|\widehat{\mathcal{H}}^{(r)} - \mathcal{H}_*^{(r)}\|_F^2 + \|\widehat{\mathcal{M}}^{(r)} - \mathcal{M}_*^{(r)}\|_F^2. \quad (\text{S12})$$

Summing (S12) over r gives

$$\|\widehat{\mathcal{T}} - \mathcal{T}_*\|_F^2 = \|\widehat{\mathcal{H}} - \mathcal{H}_*\|_F^2 + \|\widehat{\mathcal{M}} - \mathcal{M}_*\|_F^2.$$

□

B. Proofs of Theorem 2 and Corollary 1

Theorem 1 and Lemma 2 imply that with probability at least $1 - R \exp(-c_1 n) - \exp(-c_2(2n + R))$,

$$\|\widehat{\mathcal{M}} - \mathcal{M}_\star\|_F^2 \leq \|\widehat{\mathcal{T}} - \mathcal{T}_\star\|_F^2 \leq C_1 n R + C_2 k^2 (2n + R),$$

or,

$$\frac{1}{R} \sum_{r=1}^R \|\widehat{\mathcal{M}}^{(r)} - \mathcal{M}_\star^{(r)}\|_F^2 \leq C_1 n + (2 + \delta) C_2 k^2 n R^{-1} = C_1 n + \widetilde{C}_2 k^2 n R^{-1} \quad (\text{S13})$$

where $\widetilde{C}_2 = C_2(2 + \delta)$ by the assumption that $R \leq \delta n$. Therefore, there must exist a $r_0 \in \{1, \dots, R\}$, such that

$$\|\widehat{\mathcal{M}}^{(r_0)} - \mathcal{M}_\star^{(r_0)}\|_F^2 \leq C_1 n + \widetilde{C}_2 k^2 n R^{-1}. \quad (\text{S14})$$

The assumptions $\sigma_{\min}(\Lambda_\star^{(r_0)}) \geq \kappa$ and $U_\star^\top U_\star = n I_k$ imply that

$$\sigma_k(\mathcal{M}_\star^{(r_0)}) = \sigma_k(U_\star \Lambda_\star^{(r_0)} U_\star^\top) \geq n \kappa. \quad (\text{S15})$$

We also note that

$$\sigma_{k+1}(\mathcal{M}_\star^{(r_0)}) = 0 \quad (\text{S16})$$

since M_\star is of rank k .

Combining (S13) to (S16), together with Davis-Kahan Theorem (Davis & Kahan, 1970; Yu et al., 2015), we have

$$\min_{O: O^\top O = O O^\top = I_k} \left\{ \|\widehat{U} - U_\star O\|_F^2 \right\} \leq \frac{8n \|\widehat{\mathcal{M}}^{(r_0)} - \mathcal{M}_\star^{(r_0)}\|_F^2}{\{\sigma_k(\mathcal{M}_\star^{(r_0)}) - \sigma_{k+1}(\mathcal{M}_\star^{(r_0)})\}^2} \leq 8 \frac{C_1 n^2 + \widetilde{C}_2 k^2 n^2 R^{-1}}{\kappa^2 n^2} = 8 \kappa^{-2} (C_1 + \widetilde{C}_2 k^2 R^{-1}). \quad (\text{S17})$$

This leads to the results in Theorem 2.

For Corollary 1, note that when $\alpha^{(r)} = 0$ for $r = 1, \dots, R$, we have $\mathcal{T} = \mathcal{M}$. Then all the terms related to the node degree heterogeneity parameters in the calculations in Section A would be dropped and we would obtain

$$\|\widehat{\mathcal{T}} - \mathcal{T}_\star\|_F^2 \leq C k^2 (2n + R),$$

with probability $1 - c \exp(2n + R)$ for constants c and C that only depend on μ . Applying the same procedure in the proof of Theorem 2, we have the results in Corollary 1. Also note that in the proof of Theorem 1, we only utilize the assumption that $JU = U$ to prove Lemma 2. When $\mathcal{T} = \mathcal{M}$, we no longer need Lemma 2 to obtain (S10) from (S9). Therefore, the assumption $JU = U$ can be disregarded in the corollary. When fitting logistic RESCAL model in real data applications, we also do not put such constraints on the estimated parameters.

C. Proof of Proposition 1

This section shows the identifiability conditions of model (1), as proposed in Proposition 1. To prove Proposition 1, we need the following lemma.

Lemma 3. For any $\beta = (\beta_1, \dots, \beta_n)^\top \in \mathbb{R}^n$, if $\beta_1^\top 1_n + 1_n \beta^\top 1_n = 0_n$, then $\beta = 0_n$.

Proof. The condition can be written as

$$n \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^n \beta_i \\ \vdots \\ \sum_{i=1}^n \beta_i \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (\text{S18})$$

which implies $\beta_1 = \dots = \beta_n = -\frac{1}{n} \sum_{i=1}^n \beta_i$. Thus we must have $\beta = 0_n$. \square

Supplementary Materials

By Assumption A1, we have $JU = U$, $JU_{\dagger} = U_{\dagger}$, where $J = I_n - 1_n 1_n^{\top} / n$. Therefore, $U\Lambda^{(r)}U^{\top}1_n = U_{\dagger}\Lambda_{\dagger}^{(r)}U_{\dagger}^{\top}1_n = 0_n$ for $r = 1, \dots, R$. Suppose two sets of parameters yield the same edge connection probabilities, i.e.,

$$\alpha^{(r)}1_n^{\top} + 1_n\alpha^{(r)\top} + U\Lambda^{(r)}U^{\top} = \alpha_{\dagger}^{(r)}1_n^{\top} + 1_n\alpha_{\dagger}^{(r)\top} + U_{\dagger}\Lambda_{\dagger}^{(r)}U_{\dagger}^{\top} \quad (\text{S19})$$

for $r = 1, \dots, R$. Right multiplying 1_n to both sides in (S19) gives

$$\alpha^{(r)}1_n^{\top}1_n + 1_n\alpha^{(r)\top}1_n = \alpha_{\dagger}^{(r)}1_n^{\top}1_n + 1_n\alpha_{\dagger}^{(r)\top}1_n, \quad (\text{S20})$$

or

$$(\alpha^{(r)} - \alpha_{\dagger}^{(r)})1_n^{\top}1_n + 1_n(\alpha^{(r)} - \alpha_{\dagger}^{(r)})^{\top}1_n = \mathbf{0}_n. \quad (\text{S21})$$

Applying Lemma 3, we have

$$\alpha_{\dagger}^{(r)} = \alpha^{(r)}, \quad r = 1, \dots, R. \quad (\text{S22})$$

(S19) and (S22) together imply that

$$U_{\dagger}\Lambda_{\dagger}^{(r)}U_{\dagger}^{\top} = U\Lambda^{(r)}U^{\top},$$

for $r = 1, \dots, R$. This can be further written as

$$\begin{bmatrix} U_{\dagger}\Lambda_{\dagger}^{(1)} \\ \vdots \\ U_{\dagger}\Lambda_{\dagger}^{(R)} \end{bmatrix} U_{\dagger}^{\top} = \begin{bmatrix} U\Lambda^{(1)} \\ \vdots \\ U\Lambda^{(R)} \end{bmatrix} U^{\top}. \quad (\text{S23})$$

Note that $U^{\top}U = nI_k$. Left multiplying U^{\top} to the both sides of (S23) gives

$$\begin{bmatrix} U^{\top}U_{\dagger}\Lambda_{\dagger}^{(1)} \\ \vdots \\ U^{\top}U_{\dagger}\Lambda_{\dagger}^{(R)} \end{bmatrix} U_{\dagger}^{\top} = n \begin{bmatrix} \Lambda^{(1)} \\ \vdots \\ \Lambda^{(R)} \end{bmatrix} U^{\top}. \quad (\text{S24})$$

Further multiplying both sides in (S24) by $[\Lambda^{(1)} \ \Lambda^{(2)}] \dots \Lambda^{(R)} \in \mathbb{R}^{k \times (kR)}$, we have

$$\left\{ \sum_{r=1}^R (U\Lambda^{(r)})^{\top} U_{\dagger}\Lambda_{\dagger}^{(r)} \right\} U_{\dagger}^{\top} = n \left\{ \sum_{r=1}^R (\Lambda^{(r)})^2 \right\} U^{\top}. \quad (\text{S25})$$

$\left\{ \sum_{r=1}^R (\Lambda^{(r)})^2 \right\}$ is a positive semi-definite matrix in $\mathbb{R}^{k \times k}$. Assumption A3 implies that $\left\{ \sum_{r=1}^R (\Lambda^{(r)})^2 \right\}$ is of full rank, and thus invertible. Therefore, (S25) is equivalent to

$$\frac{1}{n} \left\{ \sum_{r=1}^R (\Lambda^{(r)})^2 \right\}^{-1} \left\{ \sum_{r=1}^R (U\Lambda^{(r)})^{\top} U_{\dagger}\Lambda_{\dagger}^{(r)} \right\} U_{\dagger}^{\top} = U^{\top}. \quad (\text{S26})$$

Let $O = \frac{1}{n} \left\{ \sum_{r=1}^R (\Lambda^{(r)})^2 \right\}^{-1} \left\{ \sum_{r=1}^R (U\Lambda^{(r)})^{\top} U_{\dagger}\Lambda_{\dagger}^{(r)} \right\} \in \mathbb{R}^{k \times k}$, then (S26) becomes

$$OU_{\dagger}^{\top} = U^{\top}. \quad (\text{S27})$$

Furthermore, (S27) implies that

$$OU_{\dagger}^{\top}U_{\dagger}O^{\top} = U^{\top}U, \quad O(nI_k)O^{\top} = nI_k.$$

Thus we conclude $U_{\dagger} = UO$ for some O such that $OO^{\top} = O^{\top}O = I_k$.

Lastly, for $r = 1, \dots, R$, we have

$$U_{\dagger}\Lambda_{\dagger}^{(r)}U_{\dagger}^{\top} = U\Lambda^{(r)}U^{\top} = U_{\dagger}O^{\top}\Lambda^{(r)}OU_{\dagger}^{\top},$$

so

$$U_{\dagger}^{\top}U_{\dagger}\Lambda_{\dagger}^{(r)}U_{\dagger}^{\top}U_{\dagger} = U_{\dagger}^{\top}U_{\dagger}O^{\top}\Lambda^{(r)}OU_{\dagger}^{\top}U_{\dagger}, \quad \text{or } (nI_k)\Lambda_{\dagger}^{(r)}(nI_k) = (nI_k)O^{\top}\Lambda^{(r)}O(nI_k),$$

which concludes

$$\Lambda_{\dagger}^{(r)} = O^{\top}\Lambda^{(r)}O.$$

D. Additional Simulation Results

In this section we provide simulation results under the setting that $n = 200$, $R = 50$, $k = 2$ and $n = 400$, $R = 100$, $k = 4$. Figure S1 and Figure S2 show similar patterns of parameter estimation as we have discussed in the main article. For the estimation of the overall connection probabilities $\{\hat{\Theta}^{(r)}\}_{r=1}^{R_0}$, it is bounded below mainly due to the irreducible estimation error induced by the layer-specific parameters $\alpha^{(r)}$ s. As for the estimation of shared latent variables U , after taking the log-log transformation of both relative error of \hat{U} and R_0 , the curves can be fitted well by lines with slopes close to -1 . This again demonstrates that the upper bound given in Theorem 2 is dominated by the term $\tilde{C}_2 R^{-1}$, and the estimation error of U is inversely proportional to the number of layers used for estimation.

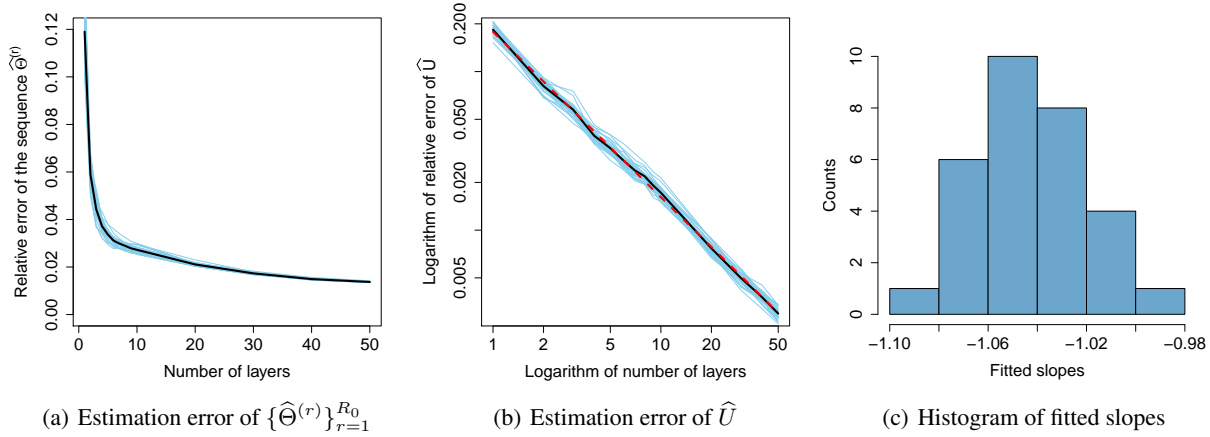


Figure S1. (a) and (b): Estimation error of parameters when $n = 200$, $R = 50$ and $k = 2$. Each light blue curve corresponds to one replication; the black curve corresponds to the average of all replications. The red dashed line corresponds to the line whose intercept and slope equal to the average fitted intercepts and slopes. (c): Histogram of all fitted slopes.

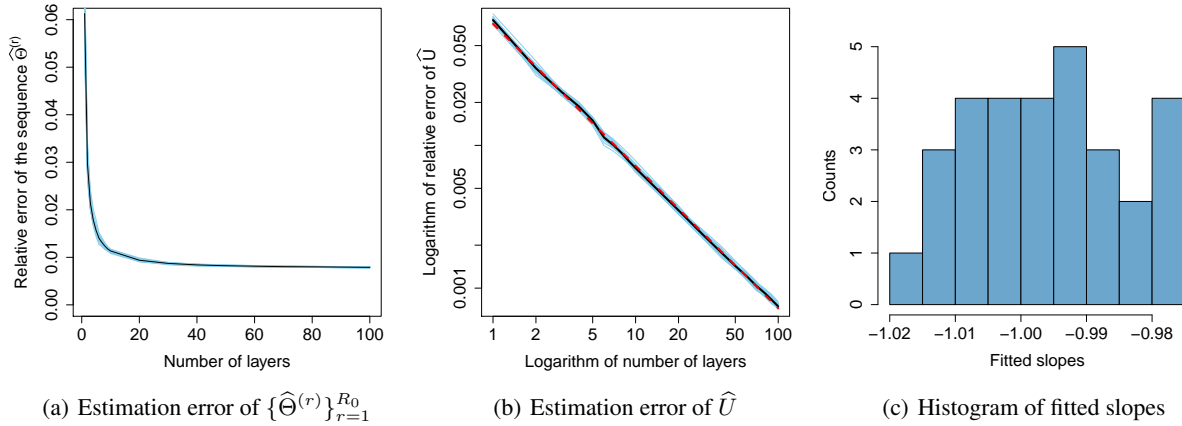


Figure S2. (a) and (b): Estimation error of parameters when $n = 400$, $R = 100$ and $k = 4$. Each light blue curve corresponds to one replication; the black curve corresponds to the average of all replications. The red dashed line corresponds to the line whose intercept and slope equal to the average fitted intercepts and slopes. (c): Histogram of all fitted slopes.

References

Davis, C. and Kahan, W. M. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1): 1–46, 1970.

Supplementary Materials

Tomioka, R. and Suzuki, T. Spectral norm of random tensors. *arXiv preprint arXiv:1407.1870*, 2014.

Wang, M. and Li, L. Learning from binary multiway data: Probabilistic tensor decomposition and its statistical optimality. *arXiv preprint arXiv:1811.05076*, 2018.

Wang, M., Duc, K. D., Fischer, J., and Song, Y. S. Operator norm inequalities between tensor unfoldings on the partition lattice. *Linear algebra and its applications*, 520:44–66, 2017.

Yu, Y., Wang, T., and Samworth, R. J. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2): 315–323, 2015.