"Autoregressive Networks" by B. Jiang, J. Li and Q. Yao

Appendix: Technical proofs

A.1 Proof of Theorem 1

Note all $X_{i,j}^t$ take binary values 0 or 1. Hence

$$\begin{aligned}&P(X_{i,j}^1=1)=P(X_{i,j}^0=1)P(X_{i,j}^1=1|X_{i,j}^0=1)+P(X_{i,j}^0=0)P(X_{i,j}^1=1|X_{i,j}^0=0)\\ =&\pi_{i,j}(1-\beta_{i,j})+(1-\pi_{i,j})\alpha_{i,j}=\frac{\alpha_{i,j}}{\alpha_{i,j}+\beta_{i,j}}(1-\beta_{i,j})+\frac{\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}}\alpha_{i,j}=\frac{\alpha_{i,j}}{\alpha_{i,j}+\beta_{i,j}}=\pi_{i,j}.\end{aligned}$$

Thus $\mathcal{L}(X_{i,j}^1)=\mathcal{L}(X_{i,j}^0)$. Since all $\mathbf{X}^t$ are Erdós-Renyi, $\mathcal{L}(\mathbf{X}^1)=\mathcal{L}(\mathbf{X}^0)$. Condition (2.5) ensures that $\{\mathbf{X}_t\}$ is a homogeneous Markov chain. Hence $\mathcal{L}(\mathbf{X}^t)=\mathcal{L}(\mathbf{X}^0)$ for any $t\geq 1$. This implies the required stationarity.

As $E(X_{i,j}^t)=P(X_{i,j}^t=1)$, and $\mathrm{Var}(X_{i,j}^t)=E(X_{i,j}^t)-\{E(X_{i,j}^t)\}^2$, (2.8) follows from the stationarity, (2.6) and (2.7).

Since the networks are all Erdös-Renyi, (2.9) follows from the Yule-Walker equation (2.10) immediately, noting $\rho_{i,j}(k)=\gamma_{i,j}(k)/\gamma_{i,j}(0)$ and $\rho_{i,j}(0)=1$. To prove (2.10), it follows from (2.1) that for any $k\geq 1$,

$$\begin{aligned}E(X_{i,j}^{t+k}X_{i,j}^t)&=E(X_{i,j}^{t+k-1}X_{i,j}^t)P(\varepsilon_{i,j}^{t+k}=0)+P(\varepsilon_{i,j}^{t+k}=1)EX_{i,j}^t\\&=(1-\alpha_{i,j}-\beta_{i,j})E(X_{i,j}^{t+k-1}X_{i,j}^t)+\alpha_{i,j}^2/(\alpha_{i,j}+\beta_{i,j}).\end{aligned}$$

Thus

$$\begin{aligned}\gamma_{i,j}(k)&=E(X_{i,j}^{t+k}X_{i,j}^t)-(EX_{i,j}^t)^2=E(X_{i,j}^{t+k}X_{i,j}^t)-\frac{\alpha_{i,j}^2}{(\alpha_{i,j}+\beta_{i,j})^2}\\&=(1-\alpha_{i,j}-\beta_{i,j})E(X_{i,j}^{t+k-1}X_{i,j}^t)+\frac{\alpha_{i,j}^2}{\alpha_{i,j}+\beta_{i,j}}(1-\frac{1}{\alpha_{i,j}+\beta_{i,j}})\\&=(1-\alpha_{i,j}-\beta_{i,j})\{E(X_{i,j}^{t+k-1}X_{i,j}^t)-\frac{\alpha_{i,j}^2}{(\alpha_{i,j}+\beta_{i,j})^2}\}=(1-\alpha_{i,j}-\beta_{i,j})\gamma_{i,j}(k-1).\end{aligned}$$

This completes the proof.

A.2 Proof of Theorem 2

We only prove (2.12), as (2.11) follows from (2.12) immediately. To prove (2.12), we only need to show

$$d_{i,j}(k)\equiv P(X_{i,j}^t\neq X_{i,j}^{t+k})=\frac{2\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2}\{1-(1-\alpha_{i,j}-\beta_{i,j})^k\},\quad k=1,2,\cdots. \tag{A.1}$$

We Proceed by induction. It is easy to check that (A.1) holds for $k = 1$. Assuming it also holds for $k \geq 1$, then

$$
\begin{aligned}
&d_{i,j}(k+1) = P(X_{i,j}^t = 0, X_{i,j}^{t+k+1} = 1) + P(X_{i,j}^t = 1, X_{i,j}^{t+k+1} = 0) \\
=& P(X_{i,j}^t = 0, X_{i,j}^{t+k} = 1, X_{i,j}^{t+k+1} = 1) + P(X_{i,j}^t = 0, X_{i,j}^{t+k} = 0, X_{i,j}^{t+k+1} = 1) \\
&+ P(X_{i,j}^t = 1, X_{i,j}^{t+k} = 0, X_{i,j}^{t+k+1} = 0) + P(X_{i,j}^t = 1, X_{i,j}^{t+k} = 1, X_{i,j}^{t+k+1} = 0) \\
=& P(X_{i,j}^t = 0, X_{i,j}^{t+k} = 1)(1 - \beta_{i,j}) + \{P(X_{i,j}^t = 0) - P(X_{i,j}^t = 0, X_{i,j}^{t+k} = 1)\}\alpha_{i,j} \\
&+ P(X_{i,j}^t = 1, X_{i,j}^{t+k} = 0)(1 - \alpha_{i,j}) + \{P(X_{i,j}^t = 1) - P(X_{i,j}^t = 1, X_{i,j}^{t+k} = 0)\}\beta_{i,j} \\
=& \{P(X_{i,j}^t = 0, X_{i,j}^{t+k} = 1) + P(X_{i,j}^t = 1, X_{i,j}^{t+k} = 0)\}(1 - \alpha_{i,j} - \beta_{i,j}) + \frac{2\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j} + \beta_{i,j}} \\
=& d_{i,j}(k)(1 - \alpha_{i,j} - \beta_{i,j}) + \frac{2\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j} + \beta_{i,j}} = \frac{2\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j} + \beta_{i,j})^2}\{1 - (1 - \alpha_{i,j} - \beta_{i,j})^{k+1}\}.
\end{aligned}
$$

Hence (A.1) also holds for $k + 1$. This completes the proof.

A.3 Proof of Theorem 3

Proof of Lemma 1

Proof. Note that for any nonempty elements $A \in \mathcal{F}_0^k, B \in \mathcal{F}_{k+\tau}^\infty$, there exist $A_0 \in \mathcal{F}_0^{k-1}$ and $B_0 \in \mathcal{F}_{k+\tau+1}^\infty$ such that $A = A_0 \times \{0\}, A_0 \times \{1\}$, or $A_0 \times \{0, 1\}$, and $B = B_0 \times \{0\}, B_0 \times \{1\}$, or $B_0 \times \{0, 1\}$. We first consider the case where $B = B_0 \times \{x_k\}$ and $A = A_0 \times \{x_{k+\tau}\}$ where $x_k, x_{k+\tau} = 0$ or 1. Note that

$$
\begin{aligned}
& P(A_0, X_{i,j}^k = x_k, B_0, X_{i,j}^{k+\tau} = x_{k+\tau}) \\
= \;& P(B_0 | X_{i,j}^{k+\tau} = x_{k+\tau}) P(X_{i,j}^{k+\tau} = x_{k+\tau}, A_0, X_{i,j}^k = x_k) \\
= \;& P(B_0, X_{i,j}^{k+\tau} = x_{k+\tau}) P(A_0, X_{i,j}^k = x_k) \cdot \frac{P(X_{i,j}^{k+\tau} = x_{k+\tau} | X_{i,j}^k = x_k)}{P(X_{i,j}^{k+\tau} = x_{k+\tau})} \\
= \;& P(B_0, X_{i,j}^{k+\tau} = x_{k+\tau}) P(A_0, X_{i,j}^k = x_k) \cdot \frac{P(X_{i,j}^{k+\tau} = x_{k+\tau}, X_{i,j}^k = x_k)}{P(X_{i,j}^{k+\tau} = x_{k+\tau}) P(X_{i,j}^k = x_k)}
\end{aligned}
$$

On the other hand, note that

$$
P(X_{i,j}^{k+\tau} = 1, X_{i,j}^k = 1) - P(X_{i,j}^{k+\tau} = 1) P(X_{i,j}^k = 1) = \rho_{i,j}(\tau);
$$

$$
\begin{aligned}
& P(X_{i,j}^{k+\tau} = 1, X_{i,j}^k = 0) - P(X_{i,j}^{k+\tau} = 1) P(X_{i,j}^k = 0) \\
= \;& P(X_{i,j}^{k+\tau} = 1) - P(X_{i,j}^{k+\tau} = 1, X_{i,j}^k = 1) - P(X_{i,j}^{k+\tau} = 1)[1 - P(X_{i,j}^k = 1)] \\
= \;& -\rho_{i,j}(\tau);
\end{aligned}
$$

$$
\begin{aligned}
& P(X_{i,j}^{k+\tau}=0, X_{i,j}^{k}=1) - P(X_{i,j}^{k+\tau}=0)P(X_{i,j}^{k}=1) \\
= \;& P(X_{i,j}^{k}=1) - P(X_{i,j}^{k+\tau}=1, X_{i,j}^{k}=1) - [1-P(X_{i,j}^{k+\tau}=1)]P(X_{i,j}^{k}=1) \\
= \;& -\rho_{i,j}(\tau);
\end{aligned}
$$

$$
\begin{aligned}
& P(X_{i,j}^{k+\tau}=0, X_{i,j}^{k}=0) - P(X_{i,j}^{k+\tau}=0)P(X_{i,j}^{k}=0) \\
= \;& P(X_{i,j}^{k+\tau}=0) - P(X_{i,j}^{k+\tau}=0, X_{i,j}^{k}=1) - P(X_{i,j}^{k+\tau}=0)[1-P(X_{i,j}^{k}=1)] \\
= \;& \rho_{i,j}(\tau).
\end{aligned}
$$

Consequently, we have

$$
\begin{aligned}
& |P(A_0, X_{i,j}^{k}=x_k, B_0, X_{i,j}^{k+\tau}=x_{k+\tau}) - P(A_0, X_{i,j}^{k}=x_k)P(B_0, X_{i,j}^{k+\tau}=x_{k+\tau})| \\
= \;& \left| P(A_0, X_{i,j}^{k}=x_k)P(B_0, X_{i,j}^{k+\tau}=x_{k+\tau}) \left[\frac{P(X_{i,j}^{k+\tau}=x_{k+\tau}, X_{i,j}^{k}=x_k)}{P(X_{i,j}^{k+\tau}=x_{k+\tau})P(X_{i,j}^{k}=x_k)} - 1 \right] \right| \\
\leq \;& \rho_{i,j}(\tau).
\end{aligned}
$$

In the case where $A = A_0 \times \{0,1\}$ and/or $B = B_0 \times \{0,1\}$, since A and B are nonempty, there exist integers $0 < k_1 < k$ and/or $k_2 > k+1$, and correspondingly $A_1 \in \mathcal{F}_0^{k_1-1} \times \{x_{k_1}\}$ and/or $B \in \mathcal{F}_{k_2+\tau+1}^{\infty} \times \{x_{k_2+\tau}\}$ with $x_{k_1}, x_{k_2+\tau} = 0$ or 1, such that $P(A\cap B) - P(A)P(B) = P(A_1 \cap B_1) - P(A_1)P(B_1)$. Following similar arguments above we have $P(A\cap B) - P(A)P(B) \leq \rho_{i,j}(\tau + k_2 - k_1) < \rho_{ij}(\tau)$. We thus proved that $\alpha^{i,j}(\tau) \leq \rho_{i,j}(\tau)$. The lemma follows from Theorem 1. □

We introduce more technical lemmas first.

Lemma 1. *For any $(i,j) \in \mathcal{J}$, denote $Y_{i,j}^t := X_{i,j}^t(1-X_{i,j}^{t-1})$, and let $\mathbf{Y}_t = (Y_{i,j}^t)_{1\leq i,j\leq p}$ be the $p\times p$ matrix at time t. Under the assumptions of Theorem 1, we have $\{\mathbf{Y}_t, t=1,2\ldots\}$ is stationary such that for any $(i,j),(l,m) \in \mathcal{J}$, and $t,s \geq 1$, $t \neq s$,*

$$
EY_{i,j}^t = \frac{\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}}, \quad \mathrm{Var}(Y_{i,j}^t) = \frac{\alpha_{i,j}\beta_{i,j}(\alpha_{i,j}+\beta_{i,j}-\alpha_{i,j}\beta_{i,j})}{(\alpha_{i,j}+\beta_{i,j})^2},
$$

$$
\rho_{Y_{i,j}}(|t-s|) \equiv \mathrm{Corr}(Y_{i,j}^t, Y_{lm}^s) = \begin{cases} -\frac{\alpha_{i,j}\beta_{i,j}(1-\alpha_{i,j}-\beta_{i,j})^{|t-s|-1}}{\alpha_{i,j}+\beta_{i,j}-\alpha_{i,j}\beta_{i,j}} & \text{if} \quad (i,j)=(l,m), \\ 0 & \text{otherwise.} \end{cases}
$$

Proof. Note that $Y_{i,j}^t = X_{i,j}^t(1-X_{i,j}^{t-1}) = (1-X_{i,j}^{t-1})I(\varepsilon_{i,j}^t = 1)$. We thus have:
$E(Y_{i,j}^t) = P(X_{i,j}^{t-1}=0)\alpha_{i,j} = (1-EX_{i,j}^{t-1})\alpha_{i,j} = \frac{\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}}$.
$\mathrm{Var}(Y_{i,j}^t) = E(Y_{i,j}^t)[1-E(Y_{i,j}^t)] = \frac{\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}}\left(1-\frac{\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}}\right) = \frac{\alpha_{i,j}\beta_{i,j}(\alpha_{i,j}+\beta_{i,j}-\alpha_{i,j}\beta_{i,j})}{(\alpha_{i,j}+\beta_{i,j})^2}$.

For $k=1$ we have $E(Y_{i,j}^t Y_{i,j}^{t+1}) = E[(1-X_{i,j}^{t-1})X_{i,j}^t(1-X_{i,j}^t)X_{i,j}^{t+1}] = 0$. For any $k \geq 2$, using the fact that $E(X_{ij}^t X_{ij}^{t+k}) = \frac{\alpha_{ij}}{(\alpha_{ij}+\beta_{ij})^2}\{\beta_{ij}(1-\alpha_{ij}-\beta_{ij})^k + \alpha_{ij}\}$, we have

$$
\begin{aligned}
E(Y_{i,j}^t Y_{i,j}^{t+k}) &= E[X_{i,j}^t(1-X_{i,j}^{t-1})(1-X_{i,j}^{t+k-1})I(\varepsilon_{i,j}^{t+k}=1)] \\
&= \alpha_{i,j}E[X_{i,j}^t(1-X_{i,j}^{t-1})(1-X_{i,j}^{t+k-1})] \\
&= \alpha_{i,j}P(X_{i,j}^{t+k-1}=0|X_{i,j}^t=1)P(X_{i,j}^t=1|X_{i,j}^{t-1}=0)P(X_{i,j}^{t-1}=0) \\
&= \frac{\alpha_{i,j}^2\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}}[1-P(X_{i,j}^{t+k-1}=1|X_{i,j}^t=1)] \\
&= \frac{\alpha_{i,j}^2\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}}\left[1-\frac{E(X_{i,j}^{t+k-1}X_{i,j}^t)}{EX_{i,j}^t}\right] \\
&= \frac{\alpha_{i,j}^2\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}}\left[1-\frac{\beta_{i,j}(1-\alpha_{i,j}-\beta_{i,j})^{k-1}+\alpha_{i,j}}{\alpha_{i,j}+\beta_{i,j}}\right] \\
&= \frac{\alpha_{i,j}^2\beta_{i,j}^2[1-(1-\alpha_{i,j}-\beta_{i,j})^{k-1}]}{(\alpha_{i,j}+\beta_{i,j})^2}.
\end{aligned}
$$

Therefore we have for any $k \geq 1$,

$$
\begin{aligned}
\mathrm{Cov}(Y_{i,j}^t, Y_{i,j}^{t+k}) &= E(Y_{i,j}^t Y_{i,j}^{t+k}) - EY_{i,j}^t EY_{i,j}^{t+k} \\
&= \frac{\alpha_{i,j}^2\beta_{i,j}^2[1-(1-\alpha_{i,j}-\beta_{i,j})^{k-1}]}{(\alpha_{i,j}+\beta_{i,j})^2} - \frac{\alpha_{i,j}^2\beta_{i,j}^2}{(\alpha_{i,j}+\beta_{i,j})^2} \\
&= -\frac{\alpha_{i,j}^2\beta_{i,j}^2(1-\alpha_{i,j}-\beta_{i,j})^{k-1}}{(\alpha_{i,j}+\beta_{i,j})^2}.
\end{aligned}
$$

Consequently, for any $|t-s| = 1, 2, \ldots$, the ACF of the process $\{Y_{i,j}^t, t = 1, 2\ldots\}$ is given as:

$$
\begin{aligned}
\rho_{Y_{i,j}}(|t-s|) &= -\frac{\alpha_{i,j}^2\beta_{i,j}^2(1-\alpha_{i,j}-\beta_{i,j})^{|t-s|-1}}{(\alpha_{i,j}+\beta_{i,j})^2} \cdot \frac{(\alpha_{i,j}+\beta_{i,j})^2}{\alpha_{i,j}\beta_{i,j}(\alpha_{i,j}+\beta_{i,j}-\alpha_{i,j}\beta_{i,j})} \\
&= -\frac{\alpha_{i,j}\beta_{i,j}(1-\alpha_{i,j}-\beta_{i,j})^{|t-s|-1}}{\alpha_{i,j}+\beta_{i,j}-\alpha_{i,j}\beta_{i,j}}.
\end{aligned}
$$

□

Since the mixing property is hereditary, $Y_{i,j}^t$ is also α-mixing. From Lemma 1 and Theorem 1 of Merlevède et al. (2009), we obtain the following concentration inequalities:

Lemma 2. *Let conditions (2.5) and C1 hold. There exist positive constants C_1 and C_2 such that for all $n \geq 4$ and $\varepsilon < \frac{1}{(\log n)(\log\log n)}$,*

$$
P\left(\left|n^{-1}\sum_{t=1}^n X_{i,j}^t - EX_{i,j}^t\right| > \varepsilon\right) \leq \exp\{-C_1 n\varepsilon^2\}, \tag{A.2}
$$

$$
P\left(\left|n^{-1}\sum_{t=1}^n Y_{i,j}^t - EY_{i,j}^t\right| > \varepsilon\right) \leq \exp\{-C_2 n\varepsilon^2\}. \tag{A.3}
$$

Now we are ready to prove Theorem 3.

Proof of Theorem 3

Let $\varepsilon = C\sqrt{\frac{\log p}{n}}$ with $C^2C_1 > 2$ and $C^2C_2 > 2$. Note that under condition (C2) we have $\varepsilon = o\big(\frac{1}{(\log n)(\log\log n)}\big)$. Consequently by Lemma 3, Theorem 1 and Lemma 2, we have

$$P\left(\left|n^{-1}\sum_{t=1}^{n} X_{i,j}^t - \frac{\alpha_{i,j}}{\alpha_{i,j}+\beta_{i,j}}\right| > C\sqrt{\frac{\log p}{n}}\right) \le \exp\{-C^2C_1\log p\}, \tag{A.4}$$

$$P\left(\left|n^{-1}\sum_{t=1}^{n} Y_{i,j}^t - \frac{\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}}\right| > C\sqrt{\frac{\log p}{n}}\right) \le \exp\{-C^2C_2\log p\}. \tag{A.5}$$

Consequently, with probability greater than $1-\exp\{-C^2C_1\log p\}-\exp\{-C^2C_2\log p\}$,

$$\frac{\frac{\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}} - C\sqrt{\frac{\log p}{n}}}{\frac{\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}} + \frac{1}{n} + C\sqrt{\frac{\log p}{n}}} \le \widehat{\alpha}_{i,j} \le \frac{\frac{\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}} + C\sqrt{\frac{\log p}{n}}}{\frac{\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}} - \frac{1}{n} - C\sqrt{\frac{\log p}{n}}}.$$

Note that when n and $\frac{n}{\log p}$ are large enough such that, $\frac{1}{n} \le C\sqrt{\frac{\log p}{n}} \le l/4$, we have

$$\alpha_{i,j} - \frac{\frac{\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}} - C\sqrt{\frac{\log p}{n}}}{\frac{\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}} + \frac{1}{n} + C\sqrt{\frac{\log p}{n}}} \le \frac{2C\alpha_{i,j}\sqrt{\frac{\log p}{n}} + C\sqrt{\frac{\log p}{n}}}{\frac{\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}}} \le 3l^{-1}C\sqrt{\frac{\log p}{n}},$$

and

$$\frac{\frac{\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}} + C\sqrt{\frac{\log p}{n}}}{\frac{\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}} - \frac{1}{n} - C\sqrt{\frac{\log p}{n}}} - \alpha_{i,j} \le \frac{2C\alpha_{i,j}\sqrt{\frac{\log p}{n}} + C\sqrt{\frac{\log p}{n}}}{\frac{\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}} - \frac{l}{2}} \le 6l^{-1}C\sqrt{\frac{\log p}{n}},$$

Therefore we conclude that when when n and $\frac{n}{\log p}$ are large enough,

$$P\left(|\widehat{\alpha}_{i,j} - \alpha_{i,j}| \ge 6l^{-1}C\sqrt{\frac{\log p}{n}}\right) \le \exp\{-C^2C_1\log p\} + \exp\{-C^2C_2\log p\}. \tag{A.6}$$

As a result, we have

$$P\left(\max_{(i,j)\in\mathcal{J}} |\widehat{\alpha}_{i,j} - \alpha_{i,j}| < 6l^{-1}C\sqrt{\frac{\log p}{n}}\right) \ge 1 - p^2\exp\{-C^2C_1\log p\} - p^2\exp\{-C^2C_2\log p\} \to 1.$$

Consequently we have $\max_{(i,j)\in\mathcal{J}} |\widehat{\alpha}_{i,j} - \alpha_{i,j}| = O_p\left(\sqrt{\frac{\log p}{n}}\right)$. Convergence of $\widehat{\beta}_{i,j}$ can be proved similarly.

A.4 Proof of Theorem 4

Note that the log-likelihood function for $(\alpha_{i,j}, \beta_{i,j})$ is:

$$l(\alpha_{i,j},\beta_{i,j}) = \log(\alpha_{i,j})\sum_{t=1}^{n} X_{i,j}^t(1-X_{i,j}^{t-1}) + \log(1-\alpha_{i,j})\sum_{t=1}^{n}(1-X_{i,j}^t)(1-X_{i,j}^{t-1})$$
$$+ \log(\beta_{i,j})\sum_{t=1}^{n}(1-X_{i,j}^t)X_{i,j}^{t-1} + \log(1-\beta_{i,j})\sum_{t=1}^{n} X_{i,j}^t X_{i,j}^{t-1}.$$

Our first observation is that, owing to the independent edge formation assumption, all the $(\widehat{\alpha}_{i,j}, \widehat{\beta}_{i,j}), (i,j) \in \mathcal{J}$ pairs are independent. For each pair $(\alpha_{i,j}, \beta_{i,j})$, the score equations of the log-likelihood function are:

$$\begin{aligned}
\frac{\partial l(\alpha_{i,j}, \beta_{i,j})}{\partial \alpha_{i,j}} &= \frac{1}{\alpha_{i,j}} \sum_{t=1}^{n} X_{i,j}^t (1 - X_{i,j}^{t-1}) - \frac{1}{1-\alpha_{i,j}} \sum_{t=1}^{n} (1 - X_{i,j}^t)(1 - X_{i,j}^{t-1}), \\
&= \left(\frac{1}{\alpha_{i,j}} + \frac{1}{1-\alpha_{i,j}} \right) \sum_{t=1}^{n} Y_{i,j}^t - \frac{1}{1-\alpha_{i,j}} \sum_{t=1}^{n} (1 - X_{i,j}^t) + O(1), \\
\frac{\partial l(\alpha_{i,j}, \beta_{i,j})}{\partial \beta_{i,j}} &= \frac{1}{\beta_{i,j}} \sum_{t=1}^{n} (1 - X_{i,j}^t) X_{i,j}^{t-1} - \frac{1}{1-\beta_{i,j}} \sum_{t=1}^{n} X_{i,j}^t X_{i,j}^{t-1} \\
&= \frac{1}{\beta_{i,j}} \sum_{t=1}^{n} X_{i,j}^{t-1} + \left(\frac{1}{\beta_{i,j}} + \frac{1}{1-\beta_{i,j}} \right) \sum_{t=1}^{n} (Y_{i,j}^t - X_{i,j}^t) \\
&= \left(\frac{1}{\beta_{i,j}} + \frac{1}{1-\beta_{i,j}} \right) \sum_{t=1}^{n} Y_{i,j}^t - \frac{1}{1-\beta_{i,j}} \sum_{t=1}^{n} X_{i,j}^t + O(1).
\end{aligned}$$

Clearly, for any $0 < \alpha_{i,j}, \beta_{i,j}, \alpha_{i,j} + \beta_{i,j} \le 1$, $\left(\frac{1}{\alpha_{i,j}} + \frac{1}{1-\alpha_{i,j}}, \frac{1}{1-\alpha_{i,j}} \right)$ and $\left(\frac{1}{\beta_{i,j}} + \frac{1}{1-\beta_{i,j}}, -\frac{1}{1-\beta_{i,j}} \right)$ are linearly independent. On the other hand, from Lemma 1, Lemma 3 and classical central limit theorems for weakly dependent sequences (Bradley, 2007; Durrett, 2019), we have $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_{i,j}^t$ and $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{i,j}^t$ are asymptotically normally distributed. Consequently, any nontrivial linear combination of $\frac{1}{\sqrt{n}} \frac{\partial l(\alpha_{i,j}, \beta_{i,j})}{\partial \alpha_{i,j}}$, $(i,j) \in J_1$ and $\frac{1}{\sqrt{n}} \frac{\partial l(\alpha_{i,j}, \beta_{i,j})}{\partial \beta_{i,j}}, (i,j) \in J_2$ converges to a normal distribution. By standard arguments for consistency of MLEs, we conclude that $(\sqrt{n}(\widehat{\alpha}_{i,j} - \alpha_{i,j}), \sqrt{n}(\widehat{\beta}_{i,j} - \beta_{i,j}))'$ converges to the normal distribution with mean $\mathbf{0}$ and covariance matrix $I(\alpha_{i,j}, \beta_{i,j})^{-1}$, where $I(\alpha_{i,j}, \beta_{i,j})$ is the Fisher information matrix given as:

$$I(\alpha_{i,j}, \beta_{i,j}) = \frac{1}{n} E \begin{bmatrix} \frac{\sum_{t=1}^{n} X_{i,j}^t (1 - X_{i,j}^{t-1})}{\alpha_{i,j}^2} + \frac{\sum_{t=1}^{n} (1 - X_{i,j}^t)(1 - X_{i,j}^{t-1})}{(1-\alpha_{i,j})^2} & 0 \\ 0 & \frac{\sum_{t=1}^{n} (1 - X_{i,j}^t) X_{i,j}^{t-1}}{\beta_{i,j}^2} + \frac{\sum_{t=1}^{n} X_{i,j}^t X_{i,j}^{t-1}}{(1-\beta_{i,j})^2} \end{bmatrix}.$$

Note that

$$\begin{aligned}
&\frac{1}{n} E \sum_{t=1}^{n} X_{i,j}^t (1 - X_{i,j}^{t-1}) = \frac{1}{n} E \sum_{t=1}^{n} (1 - X_{i,j}^t) X_{i,j}^{t-1} = \frac{\alpha_{i,j} \beta_{i,j}}{\alpha_{i,j} + \beta_{i,j}}, \\
&\frac{1}{n} E \sum_{t=1}^{n} (1 - X_{i,j}^t)(1 - X_{i,j}^{t-1}) = \frac{\beta_{i,j}}{\alpha_{i,j} + \beta_{i,j}} - \frac{\alpha_{i,j} \beta_{i,j}}{\alpha_{i,j} + \beta_{i,j}} = \frac{(1 - \alpha_{i,j}) \beta_{i,j}}{\alpha_{i,j} + \beta_{i,j}}, \\
&\frac{1}{n} E \sum_{t=1}^{n} X_{i,j}^t X_{i,j}^{t-1} = \frac{\alpha_{i,j}(1 - \beta_{i,j})}{\alpha_{i,j} + \beta_{i,j}}.
\end{aligned}$$

We thus have

$$
\begin{aligned}
I(\alpha_{i,j},\beta_{i,j}) &= \begin{bmatrix} \frac{\beta_{i,j}}{\alpha_{i,j}(\alpha_{i,j}+\beta_{i,j})} + \frac{\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})(1-\alpha_{i,j})} & 0 \\ 0 & \frac{\alpha_{i,j}}{\beta_{i,j}(\alpha_{i,j}+\beta_{i,j})} + \frac{\alpha_{i,j}}{(1-\beta_{i,j})(\alpha_{i,j}+\beta_{i,j})} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\beta_{i,j}}{\alpha_{i,j}(\alpha_{i,j}+\beta_{i,j})(1-\alpha_{i,j})} & 0 \\ 0 & \frac{\alpha_{i,j}}{\beta_{i,j}(\alpha_{i,j}+\beta_{i,j})(1-\beta_{i,j})} \end{bmatrix}.
\end{aligned}
$$

Consequently, we have

$$
\begin{bmatrix} \sqrt{n}(\widehat{\alpha}_{i,j}-\alpha_{i,j}) \\ \sqrt{n}(\widehat{\beta}_{i,j}-\beta_{i,j}) \end{bmatrix} \to N\left(\mathbf{0}, \begin{bmatrix} \frac{\alpha_{i,j}(\alpha_{i,j}+\beta_{i,j})(1-\alpha_{i,j})}{\beta_{i,j}} & 0 \\ 0 & \frac{\beta_{i,j}(\alpha_{i,j}+\beta_{i,j})(1-\beta_{i,j})}{\alpha_{i,j}} \end{bmatrix}\right).
$$

This together with the independence among the $(\widehat{\alpha}_{i,j},\widehat{\beta}_{i,j}), (i,j)\in\mathcal{J}$ pairs proves the theorem.

A.5 Proof of Proposition 1

Denote $\mathbf{N} = \text{diag}\{\sqrt{s_1},\ldots,\sqrt{s_q}\}$. Note that

$$
\begin{aligned}
\mathbf{L} &= \mathbf{D}_1^{-1/2}\mathbf{Z}\mathbf{\Omega}_1\mathbf{Z}^\top\mathbf{D}_1^{-1/2} + \mathbf{D}_2^{-1/2}\mathbf{Z}\mathbf{\Omega}_2\mathbf{Z}^\top\mathbf{D}_2^{-1/2} \\
&= \mathbf{Z}\mathbf{D}_1^{-1/2}\mathbf{\Omega}_1\mathbf{D}_1^{-1/2}\mathbf{Z}^\top + \mathbf{Z}\mathbf{D}_2^{-1/2}\mathbf{\Omega}_2\mathbf{D}_2^{-1/2}\mathbf{Z}^\top \\
&= \mathbf{Z}(\widetilde{\mathbf{\Omega}}_1 + \widetilde{\mathbf{\Omega}}_2)\mathbf{Z}^\top \\
&= (\mathbf{Z}\mathbf{N}^{-1})\mathbf{N}\widetilde{\mathbf{\Omega}}\mathbf{N}(\mathbf{Z}\mathbf{N}^{-1})^\top.
\end{aligned}
$$

Note that the columns of $\mathbf{Z}\mathbf{N}^{-1}$ are orthonormal, we thus have $rank(\mathbf{L}) = q$. Let $\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top = \mathbf{N}\widetilde{\mathbf{\Omega}}\mathbf{N}$ be the eigen-decomposition of $\mathbf{N}\widetilde{\mathbf{\Omega}}\mathbf{N}$, we immediately have $\mathbf{L} = (\mathbf{Z}\mathbf{N}^{-1})\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top(\mathbf{Z}\mathbf{N}^{-1})^\top$. Again, since the columns of $\mathbf{Z}\mathbf{N}^{-1}$ are orthonormal, we conclude that $\mathbf{\Gamma}_q = \mathbf{Z}\mathbf{N}^{-1}\mathbf{Q}$, and $\mathbf{U} = \mathbf{N}^{-1}\mathbf{Q}$. On the other hand, note that $\mathbf{U}$ is invertible, we conclude that $\mathbf{z}_{i,\cdot}\mathbf{U} = \mathbf{z}_{j,\cdot}\mathbf{U}$ and $\mathbf{z}_{i,\cdot} = \mathbf{z}_{j,\cdot}$ are equivalent.

A.6 Proof of Theorem 5

The key step is to establish an upper bound for the Frobenius norm $\|\widehat{\mathbf{L}}\widehat{\mathbf{L}} - \mathbf{L}\mathbf{L}\|_F$, and the theorem can be proved by Weyl's inequality and the Davis-Kahan theorem. We first introducing some technical lemmas.

Lemma 3. *Under the assumptions of Theorem 1, we have, there exists a constant $C_l > 0$ such that*

$$
Cov\left(\sum_{t=1}^n Y_{i,j}^t, \sum_{t=1}^n (1-X_{i,j}^{t-1})\right) = -Cov\left(\sum_{t=1}^n Y_{i,j}^t, \sum_{t=1}^n X_{i,j}^{t-1}\right) = \frac{n\alpha_{i,j}\beta_{i,j}(\alpha_{i,j}-\beta_{i,j})}{(\alpha_{i,j}+\beta_{i,j})^3} + C_{i,j},
$$

with $|C_{i,j}| \le C_l$ for any $C_{i,j}, (i,j)\in\mathcal{J}$.

Proof. In the following we shall be using the fact that for any $0 \le x < 1$, $\sum_{h=1}^{n-1} x^{h-1} = \frac{1-x^n}{1-x} = \frac{1}{1-x} + o(1)$, and $\sum_{h=1}^{n-1} hx^{h-1} = \frac{1-x^n-n(1-x)x^{n-1}}{(1-x)^2} = O(1)$. In particular, when $x = 1-\alpha_{i,j}-\beta_{i,j}$, under Condition C1, we have $2l \le 1-x < 1$, the $O(1)$ term in will become bounded uniformly for any $(i,j) \in \mathcal{J}$. In what follows, with some abuse of notation, we shall use $O_l(1)$ to denote a generic constant term with magnitude bounded by a large enough constant C_l that depends on l only.

$$
\begin{aligned}
& Cov\left(\sum_{t=1}^n Y_{i,j}^t, \sum_{t=1}^n (1-X_{i,j}^{t-1})\right) = -Cov\left(\sum_{t=1}^n Y_{i,j}^t, \sum_{t=1}^n X_{i,j}^{t-1}\right) \\
= \quad & -\sum_{t=1}^n\sum_{s=1}^n \left[E(1-X_{i,j}^{t-1})X_{i,j}^t X_{i,j}^{s-1} - \frac{\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}} \cdot \frac{\alpha_{i,j}}{\alpha_{i,j}+\beta_{i,j}}\right] \\
= \quad & -\sum_{t=1}^n\sum_{s=1}^n \left\{\frac{\alpha_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2}\left[\beta_{i,j}(1-\alpha_{i,j}-\beta_{i,j})^{|t-s+1|} + \alpha_{i,j}\right] - \frac{\alpha_{i,j}^2\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2}\right\} \\
& + \sum_{t=1}^n\sum_{s=1}^n E(X_{i,j}^{t-1}X_{i,j}^t X_{i,j}^{s-1}) \\
= \quad & -\sum_{t=1}^n\sum_{s=1}^n \frac{\alpha_{i,j}\beta_{i,j}(1-\alpha_{i,j}-\beta_{i,j})^{|t-s+1|}}{(\alpha_{i,j}+\beta_{i,j})^2} - \frac{n^2\alpha_{i,j}^2(1-\beta_{i,j})}{(\alpha_{i,j}+\beta_{i,j})^2} + (2n-1)E(X_{i,j}^{t-1}X_{i,j}^t) \\
& + \sum_{s<t} E(X_{i,j}^{t-1}X_{i,j}^t X_{i,j}^{s-1}) + \sum_{s>t+1} E(X_{i,j}^{t-1}X_{i,j}^t X_{i,j}^{s-1}). \qquad (A.7)
\end{aligned}
$$

For the first three terms on the right hand side of (A.7), we have

$$
\begin{aligned}
& -\sum_{t=1}^n\sum_{s=1}^n \frac{\alpha_{i,j}\beta_{i,j}(1-\alpha_{i,j}-\beta_{i,j})^{|t-s+1|}}{(\alpha_{i,j}+\beta_{i,j})^2} - \frac{n^2\alpha_{i,j}^2(1-\beta_{i,j})}{(\alpha_{i,j}+\beta_{i,j})^2} + (2n-1)E(X_{i,j}^{t-1}X_{i,j}^t) \\
= \quad & -\frac{\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2}\left[n + \frac{2n(1-\alpha_{i,j}-\beta_{i,j})}{\alpha_{i,j}+\beta_{i,j}}\right] - \frac{n^2\alpha_{i,j}^2(1-\beta_{i,j})}{(\alpha_{i,j}+\beta_{i,j})^2} \\
& + \frac{2n\alpha_{i,j}\left[\beta_{i,j}(1-\alpha_{i,j}-\beta_{i,j}) + \alpha_{i,j}\right]}{(\alpha_{i,j}+\beta_{i,j})^2} + O_l(1) \\
= \quad & \frac{3n\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2} - \frac{2n\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^3} - \frac{2n\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}} + \frac{2n\alpha_{i,j}^2\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2} - \frac{n^2\alpha_{i,j}^2(1-\beta_{i,j})}{(\alpha_{i,j}+\beta_{i,j})^2} + O_l(1).
\end{aligned}
$$

For the last two terms on the right hand side of (A.7), we have

$$
\begin{aligned}
& \sum_{s<t} E(X_{i,j}^{t-1}X_{i,j}^{t}X_{i,j}^{s-1}) + \sum_{s>t+1} E(X_{i,j}^{t-1}X_{i,j}^{t}X_{i,j}^{s-1}) \\
= & \sum_{s<t} P(X_{i,j}^{t}=1|X_{i,j}^{t-1}=1)P(X_{i,j}^{t-1}=1, X_{i,j}^{s-1}=1) \\
& + \sum_{s>t+1} P(X_{i,j}^{s-1}=1|X_{i,j}^{t}=1)P(X_{i,j}^{t}=1, X_{i,j}^{t-1}=1) \\
= & (1-\beta_{i,j})\sum_{s<t} E(X_{i,j}^{t-1}X_{i,j}^{s-1}) + (1-\beta_{i,j})\sum_{s>t+1} E(X_{i,j}^{s-1}X_{i,j}^{t}) \\
= & \frac{(1-\beta_{i,j})\alpha_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2}\sum_{h=1}^{n-1}(n-h)[\beta_{i,j}(1-\alpha_{i,j}-\beta_{i,j})^{h}+\alpha_{i,j}] \\
& + \frac{(1-\beta_{i,j})\alpha_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2}\sum_{h=2}^{n-1}(n-h)[\beta_{i,j}(1-\alpha_{i,j}-\beta_{i,j})^{h-1}+\alpha_{i,j}] \\
= & \frac{(n-1)^2\alpha_{i,j}^2(1-\beta_{i,j})}{(\alpha_{i,j}+\beta_{i,j})^2} + \frac{2n(1-\beta_{i,j})\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^3} + O_l(1).
\end{aligned}
$$

Consequently, we have

$$
\begin{aligned}
& Cov\left(\sum_{t=1}^{n} Y_{i,j}^{t}, \sum_{t=1}^{n}(1-X_{i,j}^{t-1})\right) = -Cov\left(\sum_{t=1}^{n} Y_{i,j}^{t}, \sum_{t=1}^{n} X_{i,j}^{t-1}\right) \\
= & \frac{3n\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2} - \frac{2n\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^3} - \frac{2n\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}} + \frac{2n\alpha_{i,j}^2\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2} - \frac{n^2\alpha_{i,j}^2(1-\beta_{i,j})}{(\alpha_{i,j}+\beta_{i,j})^2} \\
& + \frac{(n-1)^2\alpha_{i,j}^2(1-\beta_{i,j})}{(\alpha_{i,j}+\beta_{i,j})^2} + \frac{2n(1-\beta_{i,j})\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^3} + O_l(1) \\
= & \frac{3n\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2} - \frac{2n\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^3} - \frac{2n\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}} + \frac{2n\alpha_{i,j}^2\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2} \\
& + \frac{2n(1-\beta_{i,j})\alpha_{i,j}\beta_{i,j}(1-\alpha_{i,j}-\beta_{i,j})}{(\alpha_{i,j}+\beta_{i,j})^3} + O_l(1) \\
= & \frac{n\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2} - \frac{2n\alpha_{i,j}\beta_{i,j}^2}{(\alpha_{i,j}+\beta_{i,j})^3} + O_l(1).
\end{aligned}
$$

This proves the lemma. □

Lemma 4. (Bias of $\widehat{\alpha}_{i,j}$ and $\widehat{\beta}_{i,j}$) *Under the assumptions of Theorem 1, we have*

$$
E\widehat{\alpha}_{i,j} - \alpha_{i,j} = \frac{\alpha_{i,j}(\alpha_{i,j}-\beta_{i,j})}{n(\alpha_{i,j}+\beta_{i,j})\beta_{i,j}} + \frac{R_{i,j}^{(1)}}{n}, \quad E\widehat{\beta}_{i,j} - \beta_{i,j} = -\frac{\beta_{i,j}(\alpha_{i,j}-\beta_{i,j})}{n(\alpha_{i,j}+\beta_{i,j})\alpha_{i,j}} + \frac{R_{i,j}^{(2)}}{n},
$$

where $R_{i,j}^{(1)}$ and $R_{i,j}^{(2)}$ are constants such that when n is large enough we have $0 \le R_{i,j}^{(1)}, R_{i,j}^{(2)} \le R_l$ for some constant R_l and all $(i,j) \in \mathcal{J}$.

Proof. By expanding $\frac{1}{1-n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1}}$ around $\frac{1}{1-\pi_{i,j}}$, we have

$$\begin{aligned} & E\widehat{\alpha}_{i,j} \\ = \ & E\frac{n^{-1}\sum_{t=1}^{n} X_{i,j}^{t}(1-X_{i,j}^{t-1})}{n^{-1}\sum_{t=1}^{n}(1-X_{i,j}^{t-1})} \\ = \ & \frac{1}{n}E\sum_{t=1}^{n} X_{i,j}^{t}(1-X_{i,j}^{t-1})\left[\frac{1}{1-\pi_{i,j}} + \frac{(n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1}-\pi_{i,j})}{(1-\pi_{i,j})^2} + \sum_{k=2}^{\infty}\frac{(n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1}-\pi_{i,j})^k}{(1-\pi_{i,j})^{k+1}}\right]. \end{aligned}$$

Write $R_{i,j}^{(1)} := E\sum_{t=1}^{n} X_{i,j}^{t}(1-X_{i,j}^{t-1})\Big(\sum_{k=2}^{\infty}\frac{(n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1}-\pi_{i,j})^k}{(1-\pi_{i,j})^{k+1}}\Big)$. By Taylor series with Lagrange remainder we have there exist random scalers $r_{i,j}^{t} \in [n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1}, \pi_{i,j}]$ such that

$$R_{i,j}^{(1)} = E\sum_{t=1}^{n} X_{i,j}^{t}(1-X_{i,j}^{t-1})\left(\frac{(n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1}-\pi_{i,j})^2}{(1-r_{i,j}^{t})^3}\right) > 0.$$

On the other hand, note that $|n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1}-\pi_{i,j}| < 1$, we have

$$\begin{aligned} \sum_{k=2}^{\infty}\frac{|n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1}-\pi_{i,j}|^k}{(1-\pi_{i,j})^{k+1}} &\leq \left(n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1}-\pi_{i,j}\right)^2 \sum_{k=2}^{\infty}\frac{1}{(1-\pi_{i,j})^{k+1}} \\ &= \left(n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1}-\pi_{i,j}\right)^2 \frac{1}{(1-\pi_{i,j})^3\pi_{i,j}}. \end{aligned}$$

Therefore,

$$\begin{aligned} R_{i,j}^{(1)} &\leq E\sum_{t=1}^{n}\Big(\sum_{k=2}^{\infty}\frac{|n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1}-\pi_{i,j}|^k}{(1-\pi_{i,j})^{k+1}}\Big) \\ &\leq Var\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{ij}^{t-1}\right)\frac{1}{(1-\pi_{i,j})^3\pi_{i,j}} \\ &= \frac{1}{(1-\pi_{i,j})^3\pi_{i,j}} Var(X_{ij}^{t})\left[1+\frac{2}{n}\sum_{h=1}^{n-1}(n-h)\rho_{ij}(h)\right] \\ &= \frac{1}{(1-\pi_{i,j})^3\pi_{i,j}}\cdot\frac{\alpha_{ij}\beta_{ij}}{(\alpha_{ij}+\beta_{ij})^2}\left[1+\frac{2}{n}\sum_{h=1}^{n-1}(n-h)(1-\alpha_{ij}-\beta_{ij})^h\right] \\ &= \frac{1}{(1-\pi_{i,j})^3\pi_{i,j}}\cdot\frac{\alpha_{ij}\beta_{ij}}{(\alpha_{ij}+\beta_{ij})^2}\left[1+\frac{2(1-\alpha_{ij}-\beta_{ij})}{\alpha_{ij}+\beta_{ij}}+O(n^{-1})\right] \\ &= \frac{1}{(1-\pi_{i,j})^4\pi_{i,j}^2}\cdot\frac{2-\alpha_{ij}-\beta_{ij}}{(\alpha_{ij}+\beta_{ij})}+O(n^{-1}). \end{aligned}$$

Again, since $0 < l \leq \alpha_{i,j}, \beta_{i,j}, \alpha_{i,j}+\beta_{i,j} \leq 1$ holds for all $(i,j) \in \mathcal{J}$, we conclude that there exists a constant R_l such that $R_{i,j}^{(1)} \le R_l$. Together with Lemma 4, we have

$$\begin{aligned}
E\widehat{\alpha}_{i,j} &= E\frac{1}{n}\sum_{t=1}^{n} X_{i,j}^{t}(1-X_{i,j}^{t-1})\left[\frac{1}{1-\pi_{i,j}} + \frac{(n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1} - \pi_{i,j})}{(1-\pi_{i,j})^2}\right] + \frac{R_{i,j}^{(1)}}{n} \\
&= \alpha_{i,j} + \frac{Cov(\sum_{t=1}^{n} Y_{i,j}^{t}, \sum_{t=1}^{n} X_{i,j}^{t})}{n^2(1-\pi_{i,j})^2} + \frac{R_{i,j}^{(1)}}{n} \\
&= \alpha_{i,j} + \frac{\alpha_{i,j}(\alpha_{i,j}-\beta_{i,j})}{n(\alpha_{i,j}+\beta_{i,j})\beta_{i,j}} + \frac{R_{i,j}^{(1)}}{n}.
\end{aligned}$$

Similarly, write $\widetilde{R}_{i,j}^{(2)} := E\sum_{t=1}^{n} X_{i,j}^{t}(1-X_{i,j}^{t-1})\Big(\sum_{k=2}^{\infty} \frac{(n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1} - \pi_{i,j})^k}{(-1)^k \pi_{i,j}^{k+1}}\Big)$. We have,

$$\begin{aligned}
& E\widehat{\beta}_{i,j} \\
=\ & E\frac{n^{-1}\sum_{t=1}^{n}(1-X_{i,j}^{t})X_{i,j}^{t-1}}{n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1}} \\
=\ & E\frac{1}{n}\sum_{t=1}^{n}(1-X_{i,j}^{t})X_{i,j}^{t-1}\left[\frac{1}{\pi_{i,j}} - \frac{(n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1} - \pi_{i,j})}{\pi_{i,j}^2} + \sum_{k=2}^{\infty}\frac{(n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1} - \pi_{i,j})^k}{(-1)^k \pi_{i,j}^{k+1}}\right] \\
=\ & \beta_{i,j} - \frac{Cov(\sum_{t=1}^{n} Y_{i,j}^{t}, \sum_{t=1}^{n} X_{i,j}^{t} - X_{i,j}^{n} + X_{i,j}^{0})}{n^2\pi_{i,j}^2} + \frac{\widetilde{R}_{i,j}^{(2)}}{n} + O(n^{-2}) \\
=\ & \beta_{i,j} - \frac{\beta_{i,j}(\alpha_{i,j}-\beta_{i,j})}{n(\alpha_{i,j}+\beta_{i,j})\alpha_{i,j}} + \frac{\widetilde{R}_{i,j}^{(2)}}{n} + O(n^{-2}). \qquad \text{(A.8)}
\end{aligned}$$

Here in the second last step we have used the fact that $En^{-1}(X_{i,j}^{0} - X_{i,j}^{n})(n^{-1}\sum_{t=1}^{n} X_{i,j}^{t-1} - \pi_{i,j}) = O(n^{-2})$, and in the last step we have used the fact that

$$\begin{aligned}
& n^{-2}E\sum_{t=1}^{n} X_{i,j}^{t}(1-X_{i,j}^{t-1})(X_{i,j}^{n} - X_{i,j}^{0}) \\
=\ & n^{-2}E\left[\sum_{t=1}^{n} X_{i,j}^{t-1}X_{i,j}^{t}X_{i,j}^{0} - \sum_{t=1}^{n} X_{i,j}^{t-1}X_{i,j}^{t}X_{i,j}^{n}\right] + n^{-2}[E(X_{i,j}^{n})^2 - E(X_{i,j}^{n}X_{i,j}^{0})] \\
=\ & n^{-2}\Big[\sum_{t=1}^{n} P(X_{i,j}^{t}=1|X_{i,j}^{t-1}=1)P(X_{i,j}^{t-1}=1|X_{i,j}^{0}=1)P(X_{i,j}^{0}=1) \\
& -\sum_{t=1}^{n} P(X_{i,j}^{n}=1|X_{i,j}^{t}=1)P(X_{i,j}^{t}=1|X_{i,j}^{t-1}=1)P(X_{i,j}^{t-1}=1)\Big] + O(n^{-2}) \\
=\ & O(n^{-2})
\end{aligned}$$

On one hand, similar to $R_{i,j}^{(1)}$, we can show that $\widetilde{R}_{i,j}^{(2)} \simeq O(1)$. Here we use the notation $A \simeq O(B)$ to denote the fact that there exist constants $a, b > 0$ such that $a \le |A/B| \le b$. By writing $R_{i,j}^{(2)} = \widetilde{R}_{i,j}^{(2)} + O(n^{-2})$ in (A.8), we conclude that when n is large enough, there exists a R_l such that $R_{i,j}^{(2)} \le R_l$ for any $(i, j) \in \mathcal{J}$.

□

Lemma 5 implies that the bias of the MLEs is of order $O(n^{-1})$. In addition, since $R_{i,j}^{(1)}$ and $R_{i,j}^{(2)}$ are positive, the bias of $\widehat{\alpha}_{i,j}$ is always positive with exact order $O(n^{-1})$ when $\alpha_{i,j} - \beta_{i,j} > 0$, and the bias of $\widehat{\beta}_{i,j}$ is always positive with exact order $O(n^{-1})$ when $\alpha_{i,j} - \beta_{i,j} < 0$. The bound R_l here also implies that the $O(n^{-1})$ order of the bias holds uniformly for all $(i,j) \in \mathcal{J}$.

Lemma 5. *Let conditions (2.5), C1 and C2 hold. For any constant $B > 0$, there exists a large enough constant $C > 0$ such that*

$$P\left\{\|\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_1 - \mathbf{L}_1\mathbf{L}_1\|_F \geq C\left(\sqrt{\frac{\log(pn)}{np}} + \frac{1}{n} + \frac{1}{p}\right)\right\} \leq 4p\left[(pn)^{-(1+B)} + \exp\{-B\sqrt{p}\}\right], \quad \text{(A.9)}$$

$$P\left\{\|\widehat{\mathbf{L}}_2\widehat{\mathbf{L}}_2 - \mathbf{L}_2\mathbf{L}_2\|_F \geq C\left(\sqrt{\frac{\log(pn)}{np}} + \frac{1}{n} + \frac{1}{p}\right)\right\} \leq 4p\left[(pn)^{-(1+B)} + \exp\{-B\sqrt{p}\}\right], \quad \text{(A.10)}$$

$$P\left\{\|\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_2 - \mathbf{L}_1\mathbf{L}_2\|_F \geq C\left(\sqrt{\frac{\log(pn)}{np}} + \frac{1}{n} + \frac{1}{p}\right)\right\} \leq 4p\left[(pn)^{-(1+B)} + \exp\{-B\sqrt{p}\}\right], \quad \text{(A.11)}$$

$$P\left\{\|\widehat{\mathbf{L}}_2\widehat{\mathbf{L}}_1 - \mathbf{L}_2\mathbf{L}_1\|_F \geq C\left(\sqrt{\frac{\log(pn)}{np}} + \frac{1}{n} + \frac{1}{p}\right)\right\} \leq 4p\left[(pn)^{-(1+B)} + \exp\{-B\sqrt{p}\}\right]. \quad \text{(A.12)}$$

Proof. We only prove (A.9) here as (A.10), (A.11) and (A.12) can be proved similarly. Denote

$$\widetilde{\mathbf{L}}_1 := \mathbf{L}_1 - \text{diag}(\mathbf{L}_1) = \mathbf{D}_1^{-1/2}\left[\mathbf{W}_1 - \text{diag}(\mathbf{W})_1\right]\mathbf{D}_1^{-1/2},$$

and for any $1 \leq i, j \leq p$ we denote the (i,j)th element of $\widetilde{\mathbf{L}}_1\widetilde{\mathbf{L}}_1 - \mathbf{L}_1\mathbf{L}_1$ as $\delta_{i,j}$. Correspondingly, for any $\ell = 1, \ldots, p$, we define $\widetilde{d}_{\ell,1} := d_{\ell,1} - \alpha_{\ell,\ell}$. We first evaluate the error introduced by removing the $\text{diag}(\mathbf{L}_1)$ term. With some abuse of notation, let $\widetilde{\alpha}_{i,j} = \alpha_{i,j}$ for $1 \leq i \neq j \leq p$ and $\widetilde{\alpha}_{i,i} = 0$ for $i = 1, \ldots, p$. We have $\mathbf{W} - \text{diag}(\mathbf{W}) = (\widetilde{\alpha}_{i,j})_{1\leq i,j\leq p}$. Therefore,

$$|\delta_{i,j}| = \left|\sum_{k=1}^{p}\frac{\widetilde{\alpha}_{i,k}\widetilde{\alpha}_{k,j}}{d_{k,1}\sqrt{d_{i,1}d_{j,1}}} - \sum_{k=1}^{p}\frac{\alpha_{i,k}\alpha_{k,j}}{d_{k,1}\sqrt{d_{i,1}d_{j,1}}}\right| \leq \frac{\alpha_{i,i}\alpha_{i,j}}{d_{i,1}\sqrt{d_{i,1}d_{j,1}}} + \frac{\alpha_{i,j}\alpha_{j,j}}{d_{j,1}\sqrt{d_{i,1}d_{j,1}}} \leq \frac{2}{(p-1)^2 l^2}.$$

Consequently, we have

$$\begin{aligned}
\|\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_1 - \mathbf{L}_1\mathbf{L}_1\|_F^2 &= \|(\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_1 - \widetilde{\mathbf{L}}_1\widetilde{\mathbf{L}}_1) + (\widetilde{\mathbf{L}}_1\widetilde{\mathbf{L}}_1 - \mathbf{L}_1\mathbf{L}_1)\|_F^2 \\
&\leq 2\left[\|\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_1 - \widetilde{\mathbf{L}}_1\widetilde{\mathbf{L}}_1\|_F^2 + \|\widetilde{\mathbf{L}}_1\widetilde{\mathbf{L}}_1 - \mathbf{L}_1\mathbf{L}_1\|_F^2\right] \\
&= 2\|\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_1 - \widetilde{\mathbf{L}}_1\widetilde{\mathbf{L}}_1\|_F^2 + \sum_{1\leq i,j\leq p}\delta_{i,j}^2 \\
&\leq 2\|\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_1 - \widetilde{\mathbf{L}}_1\widetilde{\mathbf{L}}_1\|_F^2 + \frac{4p^2}{(p-1)^4 l^4}. \qquad \text{(A.13)}
\end{aligned}$$

Next, we derive the asymptotic bound for $\|\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_1 - \widetilde{\mathbf{L}}_1\widetilde{\mathbf{L}}_1\|_F^2$.

For any $1 \leq i \neq j \leq p$, we denote the (i,j)th element of $\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_1 - \widetilde{\mathbf{L}}_1\widetilde{\mathbf{L}}_1$ as $\Delta_{i,j}$. By definition we have,

$$\Delta_{i,j} = \sum_{\substack{1\leq k\leq p \\ k\neq i,j}}\left(\frac{\widehat{\alpha}_{i,k}\widehat{\alpha}_{k,j}}{\widehat{d}_{k,1}\sqrt{\widehat{d}_{i,1}\widehat{d}_{j,1}}} - \frac{\alpha_{i,k}\alpha_{k,j}}{d_{k,1}\sqrt{d_{i,1}d_{j,1}}}\right),$$

where $\widehat{d}_{\ell,1} = \sum_{k=1}^{p} \widehat{\alpha}_{\ell,k}$ and $d_{\ell,1} = \sum_{k=1}^{p} \alpha_{\ell,k}$ for $l = 1, \ldots, p$. Note that $\widehat{\alpha}_{i,1}, \ldots, \widehat{\alpha}_{i,p}$ are independent. Denote $\sigma_{i,k}^2 := Var(\widehat{\alpha}_{i,k})$, and $\tau_i^2 := \sum_{k=1}^{p} \sigma_{i,k}^2$. Similar to the proofs of Lemma 4 we can show that, when n is large enough, their exists a constant $C_\sigma > (2l)^{-1}$ and $c_\sigma := l(1-l)$ such that $c_\sigma n^{-1} \le \sigma_{i,k}^2 \le C_\sigma n^{-1}$ for any $(i,j) \in \mathcal{J}$. Consequently, $\tau_i^2 \simeq O(n^{-1}p)$. On the other hand, from Lemma 5 we know that there exists a large enough constant $C_\alpha > 0$ such that $|E\widehat{\alpha}_{i,j} - \alpha_{i,j}| \le \frac{C_\alpha}{n}$ for all $(i,j) \in \mathcal{J}$, and consequently, $\frac{|E\widehat{d}_{\ell,1} - d_{\ell,1}|}{p} \le \frac{|E\widehat{d}_{\ell,1} - \widetilde{d}_{\ell,1}|}{p} + \frac{1}{p} < \frac{C_\alpha}{n} + \frac{1}{p}$ for any $l = 1, \ldots, p$. We next break our proofs into three steps:

Step 1. Concentration of $p^{-1}\widehat{d}_{\ell,1}$.

We establish the concentration by taking care of the bias and verifying the moment conditions of the Bernstein's inequality (Lin and Bai, 2011).

From the proof of Theorem 3, similar to (A.6), we have when n is large enough such that $\frac{1}{n} \le C\sqrt{\frac{\log n}{n}} \le l/4$,

$$
\begin{aligned}
& P\left(|\widehat{\alpha}_{\ell,j} - E\widehat{\alpha}_{\ell,j}| \ge (6l^{-1} + C_\alpha)C\sqrt{\frac{\log n}{n}}\right) \\
\le\ & P\left(|\widehat{\alpha}_{\ell,j} - E\widehat{\alpha}_{\ell,j}| \ge 6l^{-1}C\sqrt{\frac{\log n}{n}} + \frac{C_\alpha}{n}\right) \\
\le\ & P\left(|\widehat{\alpha}_{\ell,j} - \alpha_{\ell,j}| \ge 6l^{-1}C\sqrt{\frac{\log n}{n}} + \frac{C_\alpha}{n} - |E\widehat{\alpha}_{\ell,j} - \alpha_{\ell,j}|\right) \\
\le\ & \exp\{-C^2C_1 \log n\} + \exp\{-C^2C_2 \log n\} \\
\le\ & 2\exp\{-C^2C_3 \log n\},
\end{aligned}
$$

where constants C, C_1, C_2 are defined as in Theorem 3 and $C_3 = \min\{C_1, C_2\}$.

For any integer $k > 2$, we denote the event $\left\{|\widehat{\alpha}_{\ell,j} - E\widehat{\alpha}_{\ell,j}| \le (6l^{-1} + C_\alpha)\sqrt{\frac{k\log n}{C_3 n}}\right\}$ as A_k, and denote its complement as A_k^c. When $k < \frac{C_3 l^2 n}{16\log n}$, we have $\sqrt{\frac{k\log n}{C_3 n}} < l/4$. Consequently,

$$
\begin{aligned}
& E|\widehat{\alpha}_{\ell,j} - E\widehat{\alpha}_{\ell,j}|^k \\
=\ & E|\widehat{\alpha}_{\ell,j} - E\widehat{\alpha}_{\ell,j}|^2 |\widehat{\alpha}_{\ell,j} - E\widehat{\alpha}_{\ell,j}|^{k-2} I\{A_k\} + E|\widehat{\alpha}_{\ell,j} - E\widehat{\alpha}_{\ell,j}|^k I\{A_k^c\} \\
\le\ & \sigma_{\ell,j}^2 k^{\frac{k-2}{2}} \left[(6l^{-1} + C_\alpha)\sqrt{\frac{\log n}{C_3 n}}\right]^{k-2} + 2\exp\{-k\log n\}.
\end{aligned}
$$

Note that when $k > 4$, from Stirling's approximation we have $k^{\frac{k-2}{2}} \le e^k k!/(\sqrt{2\pi}k^{3/2}) < e^{k-2}k!/3$. For $k = 3, 4$, we can directly verify that $k^{\frac{k-2}{2}} < e^{k-2}k!/3$. On the other hand, note that $n^{-1} = o\left(\sqrt{\frac{\log n}{n}}\right)$. When n is large enough, we have

$$
2\exp\{-k\log n\} = 2n^{-k} \le n^{-1}c_\sigma k! \left[e(6l^{-1} + C_\alpha)\sqrt{\frac{\log n}{C_3 n}}\right]^{k-2} /6.
$$

Consequently, we have, when n is large enough,

$$E|\widehat{\alpha}_{\ell,j} - E\widehat{\alpha}_{\ell,j}|^k \le \sigma_{\ell,j}^2 k! \left[e(6l^{-1} + C_\alpha)\sqrt{\frac{\log n}{C_3 n}} \right]^{k-2} /2.$$

Next we consider the case where $k > \frac{C_3 l^2 n}{16 \log n}$. Denote $a = e(6l^{-1} + C_\alpha)\sqrt{\frac{\log n}{C_3 n}}$. Clearly when n is large enough, we have $ka > e^2$ hold for any $k > \frac{C_3 l^2 n}{16 \log n} \ge 3$. By Stirling's approximation, we have

$$k! a^{k-2}/2 \ge \sqrt{2\pi} k^{k+\frac{1}{2}} e^{-k} a^{k-2}/2 > (ka)^{k-2} > 1.$$

Consequently, we have

$$E|\widehat{\alpha}_{\ell,j} - E\widehat{\alpha}_{\ell,j}|^k \le E|\widehat{\alpha}_{\ell,j} - E\widehat{\alpha}_{\ell,j}|^2 < \sigma_{\ell,j}^2 k! a^{k-2}/2.$$

Therefore, we conclude that, when n is large enough, the following inequality holds for all integer $k > 2$:

$$E|\widehat{\alpha}_{\ell,j} - E\widehat{\alpha}_{\ell,j}|^k \le \sigma_{\ell,j}^2 k! a^{k-2}/2.$$

This verifies the conditions of the Bernstein's inequality (Bennett, 1962; Lin and Bai, 2011), from which we obtain, for any constant $C_d > 0$:

$$\begin{aligned}
& P\left(\frac{|\widehat{d}_{\ell,1} - d_{\ell,1}|}{p} \ge C_d \sqrt{\frac{\log(pn)}{np}} + \frac{C_\alpha}{n} + \frac{1}{p} \right) \\
\le \ & P\left(\frac{|\widehat{d}_{\ell,1} - E(\widehat{d}_{\ell,1})|}{p} \ge C_d \sqrt{\frac{\log(pn)}{np}} + \frac{C_\alpha}{n} + \frac{1}{p} - \frac{|E(\widehat{d}_{\ell,1}) - d_{\ell,1}|}{p} \right) \\
\le \ & P\left(\frac{|\widehat{d}_{\ell,1} - E(\widehat{d}_{\ell,1})|}{p} \ge C_d \sqrt{\frac{\log(pn)}{np}} \right) \\
\le \ & 2\exp\left\{ -\frac{\sqrt{p} C_d^2 n^{-1} \log(pn)}{2(\sqrt{p} C_\sigma/n + a C_d \sqrt{\log(pn)/n})} \right\} \\
= \ & 2\exp\left\{ -\frac{\sqrt{p} C_d^2 n^{-1} \log(pn)}{2(\sqrt{p} C_\sigma/n + C_d e(6l^{-1} + C_\alpha)\sqrt{\log n/(C_3 n)}\sqrt{\log(pn)/n})} \right\}. \qquad \text{(A.14)}
\end{aligned}$$

When $\sqrt{p} C_\sigma/n > C_d e(6l^{-1} + C_\alpha)\sqrt{\log n/(C_3 n)}\sqrt{\log(pn)/n})$, for any constant $B > 0$, by choosing $C_d > 2\sqrt{(B+1)C_\sigma}$, (A.14) reduces to

$$\begin{aligned}
& P\left(\frac{|\widehat{d}_{\ell,1} - d_{\ell,1}|}{p} \ge C_d \sqrt{\frac{\log(pn)}{np}} + \frac{C_\alpha}{n} + \frac{1}{p} \right) \\
\le \ & 2\exp\left\{ -\frac{\sqrt{p} C_d^2 n^{-1} \log(pn)}{4\sqrt{p} C_\sigma/n} \right\} < 2(pn)^{-(B+1)}. \qquad \text{(A.15)}
\end{aligned}$$

When $\sqrt{p}C_\sigma/n \le C_d e(6l^{-1}+C_\alpha)\sqrt{\log n/(C_3 n)}\sqrt{\log(pn)/n}$, by choosing $C_d = 4Be(6l^{-1}+C_\alpha)/\sqrt{C_3}$, (A.14) reduces to

$$
\begin{aligned}
& P\left(\frac{|\widehat{d}_{\ell,1}-d_{\ell,1}|}{p} \ge C_d\sqrt{\frac{\log(pn)}{np}}+\frac{C_\alpha}{n}+\frac{1}{p}\right) \\
\le\ & 2\exp\left\{-\frac{\sqrt{p}C_d^2 n^{-1}\log(pn)}{4C_d e(6l^{-1}+C_\alpha)\sqrt{\log n/(C_3 n)}\sqrt{\log(pn)/n}}\right\} \\
\le\ & 2\exp\{-B\sqrt{p}\}. \qquad \text{(A.16)}
\end{aligned}
$$

From (A.14), (A.15) and (A.16) we conclude that for any $B>0$, by choosing C_d to be large enough, we have,

$$
\begin{aligned}
& P\left(\max_{l=1,\dots,p}\frac{|\widehat{d}_{\ell,1}-d_{\ell,1}|}{p} \ge C_d\sqrt{\frac{\log(pn)}{np}}+\frac{C_\alpha}{n}+\frac{1}{p}\right) \\
\le\ & 2p\left[(pn)^{-(1+B)}+\exp\{-B\sqrt{p}\}\right]. \qquad \text{(A.17)}
\end{aligned}
$$

Step 2. Concentration of $\Delta_{i,j}$.

Using the fact that $\widehat{\alpha}_{k,k}=0$ for $k=1,\dots,p$, we have,

$$
\Delta_{i,j} = \sum_{k=1}^{p}\left(\frac{\widehat{\alpha}_{i,k}\widehat{\alpha}_{k,j}}{\widehat{d}_{k,1}\sqrt{\widehat{d}_{i,1}\widehat{d}_{j,1}}}-\frac{\widehat{\alpha}_{i,k}\widehat{\alpha}_{k,j}}{d_{k,1}\sqrt{d_{i,1}d_{j,1}}}\right)+\sum_{\substack{1\le k\le p\\ k\ne i,j}}\left(\frac{\widehat{\alpha}_{i,k}\widehat{\alpha}_{k,j}}{d_{k,1}\sqrt{d_{i,1}d_{j,1}}}-\frac{\alpha_{i,k}\alpha_{k,j}}{d_{k,1}\sqrt{d_{i,1}d_{j,1}}}\right).
$$

We next bound the two terms on the right hand side of the above inequality. For the first term, denote $e_k := (\widehat{d}_{k,1}-d_{k,1})/p$. From (A.17) we have there exists a large enough constant C_B such that

$$
P\left\{\max_{k=1,\dots,p}|e_k| \le C_B\left(\sqrt{\frac{\log(pn)}{np}}+\frac{1}{n}+\frac{1}{p}\right)\right\} \ge 1-2p\left[(pn)^{-(1+B)}+\exp\{-B\sqrt{p}\}\right].
$$

Denote the event $\left\{\max_{k=1,\dots,p}|e_k| \le C_B\left(\sqrt{\frac{\log(pn)}{np}}+\frac{1}{n}+\frac{1}{p}\right)\right\}$ as $\mathcal{E}_B$. Under $\mathcal{E}_B$, we have, when n and p are large enough, $\sqrt{p^{-1}d_{k,1}+e_k}=\sqrt{p^{-1}d_{k,1}}+e_k/(2\sqrt{p^{-1}d_{k,1}})+O(e_k^2)$, and hence there exists a large enough constant $C_{l,B}>0$ such that for any $1\le i,j\le p$,

$$
\begin{aligned}
& \left|\frac{\widehat{\alpha}_{i,k}\widehat{\alpha}_{k,j}}{\widehat{d}_{k,1}\sqrt{\widehat{d}_{i,1}\widehat{d}_{j,1}}}-\frac{\widehat{\alpha}_{i,k}\widehat{\alpha}_{k,j}}{d_{k,1}\sqrt{d_{i,1}d_{j,1}}}\right| \\
\le\ & \frac{\left|p^{-1}d_{k,1}\sqrt{p^{-1}d_{i,1}p^{-1}d_{j,1}}-(p^{-1}d_{k,1}+e_k)\sqrt{(p^{-1}d_{i,1}+e_i)(p^{-1}d_{j,1}+e_j)}\right|}{p^2(p^{-1}d_{k,1}+e_k)\sqrt{(p^{-1}d_{i,1}+e_i)(p^{-1}d_{j,1}+e_j)}p^{-1}d_{k,1}\sqrt{p^{-1}d_{i,1}p^{-1}d_{j,1}}} \\
=\ & O(p^{-2}(|e_i|+|e_j|+|e_k|)) \\
\le\ & \frac{C_{l,B}}{p^2}\left(\sqrt{\frac{\log(pn)}{np}}+\frac{1}{n}+\frac{1}{p}\right).
\end{aligned}
$$

Consequently, we have, under $\mathcal{E}_B$,

$$\left|\sum_{k=1}^{p}\left(\frac{\widehat{\alpha}_{i,k}\widehat{\alpha}_{k,j}}{\widehat{d}_{k,1}\sqrt{\widehat{d}_{i,1}\widehat{d}_{j,1}}}-\frac{\widehat{\alpha}_{i,k}\widehat{\alpha}_{k,j}}{d_{k,1}\sqrt{d_{i,1}d_{j,1}}}\right)\right|\leq\frac{C_{l,B}}{p}\left(\sqrt{\frac{\log(pn)}{np}}+\frac{1}{n}+\frac{1}{p}\right). \tag{A.18}$$

For the second term, from the proof of (A.6), we have, there exists a constant $D_B>0$ such that, when n and $\frac{n}{\log(pn)}$ are large enough,

$$P\left(\max_{1\leq i,j\leq p}|\widehat{\alpha}_{i,j}-\alpha_{i,j}|\leq D_B\sqrt{\frac{\log(pn)}{n}}\right)\geq 1-2p^2(pn)^{-(2+B)}>1-2p(pn)^{-(1+B)}.$$

Denote the event $\left\{\max_{1\leq i,j\leq p}|\widehat{\alpha}_{i,j}-\alpha_{i,j}|\leq D_B\sqrt{\frac{\log(pn)}{n}}\right\}$ as $\mathcal{A}_B$. Under $\mathcal{A}_B$, we have, there exists a large enough constant $D_{l,B}>0$ such that when n and p are large enough,

$$\begin{aligned}\left|\sum_{\substack{1\leq k\leq p\\ k\neq i,j}}\left(\frac{\widehat{\alpha}_{i,k}\widehat{\alpha}_{k,j}}{d_{k,1}\sqrt{d_{i,1}d_{j,1}}}-\frac{\alpha_{i,k}\alpha_{k,j}}{d_{k,1}\sqrt{d_{i,1}d_{j,1}}}\right)\right| &\leq \frac{\max_{1\leq i,j\leq p}|\widehat{\alpha}_{i,k}\widehat{\alpha}_{k,j}-\alpha_{i,k}\alpha_{k,j}|}{(p-1)l^2}\\ &\leq \frac{D_{l,B}}{p}\sqrt{\frac{\log(pn)}{n}}. \end{aligned}\tag{A.19}$$

From (A.18) and (A.19) we conclude that, when n and p are large enough,

$$\begin{aligned}&P\left\{\max_{1\leq i,j\leq p}|\Delta_{i,j}|>\frac{C_{l,B}+D_{l,B}}{p}\left(\sqrt{\frac{\log(pn)}{np}}+\frac{1}{n}+\frac{1}{p}\right)\right\}\\ \leq\;& P(\mathcal{E}_B^c)+P(\mathcal{A}_B^c)\\ \leq\;& 2p\left[(pn)^{-(1+B)}+\exp\{-B\sqrt{p}\}\right]+2p(pn)^{-(1+B)}\\ <\;& 4p\left[(pn)^{-(1+B)}+\exp\{-B\sqrt{p}\}\right].\end{aligned}\tag{A.20}$$

Step 3. Proof of (A.9).

Note that $\|\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_1-\widetilde{\mathbf{L}}_1\widetilde{\mathbf{L}}_1\|_F=\sqrt{\sum_{1\leq i,j\leq p}\Delta_{i,j}^2}\leq p\max_{1\leq i,j\leq p}|\Delta_{i,j}|$. Choose $C>C_{l,B}+D_{l,B}$. From (A.13) and (A.20) we immediately have that when n and p are large enough,

$$P\left\{\|\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_1-\mathbf{L}_1\mathbf{L}_1\|_F\geq C\left(\sqrt{\frac{\log(pn)}{np}}+\frac{1}{n}+\frac{1}{p}\right)\right\}\leq 4p\left[(pn)^{-(1+B)}+\exp\{-B\sqrt{p}\}\right].$$

This proves (A.9).

□

Lemma 6. *Let conditions (2.5), C1 and C2 hold. For any constant $B>0$, there exists a large enough constant $C>0$ such that*

$$P\left\{\|\widehat{\mathbf{L}}\widehat{\mathbf{L}}-\mathbf{L}\mathbf{L}\|_F\geq 4C\left(\sqrt{\frac{\log(pn)}{np}}+\frac{1}{n}+\frac{1}{p}\right)\right\}\leq 16p\left[(pn)^{-(1+B)}+\exp\{-B\sqrt{p}\}\right]. \tag{A.21}$$

Proof. Note that from the triangle inequality we have

$$\begin{aligned}
& \|\widehat{\mathbf{L}}\widehat{\mathbf{L}} - \mathbf{L}\mathbf{L}\|_F \\
= \;& \|(\widehat{\mathbf{L}}_1 + \widehat{\mathbf{L}}_2)(\widehat{\mathbf{L}}_1 + \widehat{\mathbf{L}}_2) - (\mathbf{L}_1 + \mathbf{L}_2)(\mathbf{L}_1 + \mathbf{L}_2)\|_F \\
= \;& \|(\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_1 - \mathbf{L}_1\mathbf{L}_1) + (\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_2 - \mathbf{L}_1\mathbf{L}_2) + (\widehat{\mathbf{L}}_2\widehat{\mathbf{L}}_1 - \mathbf{L}_2\mathbf{L}_1) + (\widehat{\mathbf{L}}_2\widehat{\mathbf{L}}_2 - \mathbf{L}_2\mathbf{L}_2)\|_F \\
\leq \;& \|\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_1 - \mathbf{L}_1\mathbf{L}_1\|_F + \|\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_2 - \mathbf{L}_1\mathbf{L}_2\|_F + \|\widehat{\mathbf{L}}_2\widehat{\mathbf{L}}_1 - \mathbf{L}_2\mathbf{L}_1\|_F + \|\widehat{\mathbf{L}}_2\widehat{\mathbf{L}}_2 - \mathbf{L}_2\mathbf{L}_2\|_F.
\end{aligned}$$

Together with Lemma 6 we immediately conclude that (A.21) hold.

□

Proof of Theorem 5

From Weyls inequality and Lemma 7, we have,

$$\max_{i=1,\ldots,p} |\lambda_i^2 - \widehat{\lambda}_i^2| \leq \|\widehat{\mathbf{L}}\widehat{\mathbf{L}} - \mathbf{L}\mathbf{L}\|_2 \leq \|\widehat{\mathbf{L}}\widehat{\mathbf{L}} - \mathbf{L}\mathbf{L}\|_F = O_p\left(\sqrt{\frac{\log(pn)}{np}} + \frac{1}{n} + \frac{1}{p}\right).$$

(3.8) is a direct result of the Davis-Kahan theorem (Rohe et al., 2011; Yu et al., 2015) theorem and Lemma 7.

A.7 Proof of Theorem 6

Recall that $\mathbf{\Gamma}_q = \mathbf{Z}\mathbf{U}$ where $\mathbf{U}$ is defined as in the proof of Proposition 1. For any $1 \leq i \neq j \leq n$ such that $\mathbf{z}_i \neq \mathbf{z}_j$, we need to show that $\|\mathbf{z}_i\mathbf{U}\mathbf{O}_q - \mathbf{z}_j\mathbf{U}_q\mathbf{O}_q\|_2 = \|\mathbf{z}_i\mathbf{U} - \mathbf{z}_j\mathbf{U}\|_2$ is large enough, so that the perturbed version (i.e. the rows of $\widehat{\mathbf{\Gamma}}_q$) is not changing the clustering structure.

Denote the ith row of $\mathbf{\Gamma}_q\mathbf{O}_q$ and $\widehat{\mathbf{\Gamma}}_q$ as $\boldsymbol{\gamma}_i$ and $\widehat{\boldsymbol{\gamma}}_i$, respectively, for $i = 1, \ldots, p$. Notice that from the proof of Proposition 1, we have $\mathbf{U}\mathbf{U}^\top = \mathbf{N}^{-1}\mathbf{Q}\mathbf{Q}^\top\mathbf{N}^{-1} = \mathbf{N}^{-2} = \mathrm{diag}\{s_1^{-1}, \ldots, s_q^{-1}\}$. Consequently, for any $\mathbf{z}_i \neq \mathbf{z}_j$, we have:

$$\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j\|_2 = \|\mathbf{z}_i\mathbf{U}\mathbf{O}_q - \mathbf{z}_j\mathbf{U}_q\mathbf{O}_q\|_2 = \|\mathbf{z}_i\mathbf{U} - \mathbf{z}_j\mathbf{U}\|_2 \geq \sqrt{\frac{2}{s_{\max}}}. \tag{A.22}$$

We first show that $\mathbf{z}_i \neq \mathbf{z}_j$ implies $\widehat{\mathbf{c}}_i \neq \widehat{\mathbf{c}}_j$. Notice that $\mathbf{\Gamma}_q\mathbf{O}_q \in \mathcal{M}_{p,q}$. Denote $\widehat{\mathbf{C}} = (\widehat{\mathbf{c}}_1, \cdots, \widehat{\mathbf{c}}_p)^\top$. By the definition of $\widehat{\mathbf{C}}$ we have

$$\|\mathbf{\Gamma}_q\mathbf{O}_q - \widehat{\mathbf{C}}\|_F^2 \leq \|\widehat{\mathbf{\Gamma}}_q - \widehat{\mathbf{C}}\|_F^2 + \|\widehat{\mathbf{\Gamma}}_q - \mathbf{\Gamma}_q\mathbf{O}_q\|_F^2 \leq 2\|\widehat{\mathbf{\Gamma}}_q - \mathbf{\Gamma}_q\mathbf{O}_q\|_F^2. \tag{A.23}$$

Suppose there exist $i, j \in \{1, \ldots, p\}$ such that $\mathbf{z}_i \neq \mathbf{z}_j$ but $\widehat{\mathbf{c}}_i = \widehat{\mathbf{c}}_j$. We have

$$\|\mathbf{\Gamma}_q\mathbf{O}_q - \widehat{\mathbf{C}}\|_F^2 \geq \|\mathbf{z}_i\mathbf{U}\mathbf{O}_q - \widehat{\mathbf{c}}_i\|_2^2 + \|\mathbf{z}_j\mathbf{U}\mathbf{O}_q - \widehat{\mathbf{c}}_j\|_2^2 \geq \|\mathbf{z}_i\mathbf{U}\mathbf{O}_q - \mathbf{z}_j\mathbf{U}\mathbf{O}_q\|_2^2. \tag{A.24}$$

Combining (A.22), (3.8), (A.23) and (A.24), we have:

$$\sqrt{\frac{2}{s_{\max}}} \leq \|\mathbf{\Gamma}_q\mathbf{O}_q - \widehat{\mathbf{C}}\|_F \leq \sqrt{2}\|\widehat{\mathbf{\Gamma}}_q - \mathbf{\Gamma}_q\mathbf{O}_q\|_F \leq 4\sqrt{2}\lambda_q^{-2}C\left(\sqrt{\frac{\log(pn)}{np}} + \frac{1}{n} + \frac{1}{p}\right).$$

We have reach a contradictory with (3.9). Therefore we conclude that $\widehat{\mathbf{c}}_i \neq \widehat{\mathbf{c}}_j$.

Next we show that if $\mathbf{z}_i = \mathbf{z}_j$ we must have $\widehat{\mathbf{c}}_i = \widehat{\mathbf{c}}_j$. Assume that there exist $1 \leq i \neq j \leq p$ such that $\mathbf{z}_i = \mathbf{z}_j$ and $\widehat{\mathbf{c}}_i \neq \widehat{\mathbf{c}}_j$. Notice that from the previous conclusion (i.e., that different z_i implies different $\widehat{\mathbf{c}}_i$), since there are q distinct rows in $\mathbf{Z}$, there are correspondingly q different rows in $\widehat{\mathbf{C}}$. Consequently for any $\mathbf{z}_i = \mathbf{z}_j$, if $\widehat{\mathbf{c}}_i \neq \widehat{\mathbf{c}}_j$ there must exist a $k \neq i, j$ such that $\mathbf{z}_i = \mathbf{z}_j \neq \mathbf{z}_k$ and $\widehat{\mathbf{c}}_j = \widehat{\mathbf{c}}_k$. Let $\widehat{\mathbf{C}}^*$ be $\widehat{\mathbf{C}}$ with the jth row replaced by $\widehat{\mathbf{c}}_i$. We have

$$
\begin{aligned}
& \|\widehat{\boldsymbol{\Gamma}}_q - \widehat{\mathbf{C}}^*\|_F^2 - \|\widehat{\boldsymbol{\Gamma}}_q - \widehat{\mathbf{C}}\|_F^2 \\
= \;& \|\widehat{\boldsymbol{\gamma}}_j - \widehat{\mathbf{c}}_i\|_2^2 - \|\widehat{\boldsymbol{\gamma}}_j - \widehat{\mathbf{c}}_k\|_2^2 \\
= \;& \|\widehat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j + \boldsymbol{\gamma}_i - \widehat{\mathbf{c}}_i\|_2^2 - \|\widehat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j + \boldsymbol{\gamma}_i - \boldsymbol{\gamma}_k + \boldsymbol{\gamma}_k - \widehat{\mathbf{c}}_k\|_2^2 \\
\leq \;& \|\widehat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j + \boldsymbol{\gamma}_i - \widehat{\mathbf{c}}_i\|_2^2 + \|\widehat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j + \boldsymbol{\gamma}_k - \widehat{\mathbf{c}}_k\|_2^2 - \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_k\|_2^2 \\
\leq \;& \|\widehat{\boldsymbol{\Gamma}}_q - \boldsymbol{\Gamma}_q \mathbf{O}_q\|_F^2 + \|\boldsymbol{\Gamma}_q \mathbf{O}_q - \widehat{\mathbf{C}}\|_F^2 - \frac{2}{s_{\max}} \\
\leq \;& 3\left\{4\lambda_q^{-2} C \left(\sqrt{\frac{\log(pn)}{np}} + \frac{1}{n} + \frac{1}{p}\right)\right\}^2 - \frac{2}{s_{\max}} \\
< \;& 0.
\end{aligned}
$$

Again, we reach a contradiction and so we conclude that if $\mathbf{z}_i = \mathbf{z}_j$ we must have $\widehat{\mathbf{c}}_i = \widehat{\mathbf{c}}_j$.

A.8 Proof of Theorem 8

Note that from Theorem 6, we have the memberships can be recovered with probability tending to 1, i,e, $P(\widehat{\nu} \neq \nu) \to 0$. On the other hand, given $\widehat{\nu} = \nu$, we have, the log likelihood function of $(\theta_{k,\ell}, \eta_{k,\ell})$, $1 \leq k \leq \ell \leq q$, is

$$
\begin{aligned}
l(\{\theta_{k,\ell}, \eta_{k,\ell}\}; \nu) \;=\; & \sum_{(i,j)\in S_{k,l}} \sum_{t=1}^{n} \Big\{ X_{i,j}^t (1 - X_{i,j}^{t-1}) \log \theta_{k,\ell} + (1 - X_{i,j}^t)(1 - X_{i,j}^{t-1}) \log(1 - \theta_{k,\ell}) \\
& + (1 - X_{i,j}^t) X_{i,j}^{t-1} \log \eta_{k,\ell} + X_{i,j}^t X_{i,j}^{t-1} \log(1 - \eta_{k,\ell}) \Big\}.
\end{aligned}
$$

Using the same arguments as in the proof of Theorem 4, we can conclude that when $\widehat{\nu} = \nu$, $\sqrt{n}\mathbf{N}_{J_1,J_2}^{\frac{1}{2}}(\widehat{\boldsymbol{\Psi}}_{\mathcal{K}_1,\mathcal{K}_2} - \boldsymbol{\Psi}_{\mathcal{K}_1,\mathcal{K}_2}) \to N(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}_{\mathcal{K}_1,\mathcal{K}_2})$. Let $\mathbf{Y} \sim N(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}_{\mathcal{K}_1,\mathcal{K}_2})$. For any $\mathcal{Y} \subset \mathcal{R}^{m_1+m_2}$, let $\boldsymbol{\Phi}(\mathcal{Y}) := P(\mathbf{Y} \in \mathcal{Y})$, we have:

$$
\begin{aligned}
& |P(\sqrt{n}\mathbf{N}_{\mathcal{K}_1,\mathcal{K}_2}^{\frac{1}{2}}(\widehat{\boldsymbol{\Psi}}_{\mathcal{K}_1,\mathcal{K}_2} - \boldsymbol{\Psi}_{\mathcal{K}_1,\mathcal{K}_2}) \in \mathcal{Y}) - \boldsymbol{\Phi}(\mathcal{Y})| \\
\leq \;& P(\widehat{\nu} \neq \nu) + |P(\sqrt{n}\mathbf{N}_{\mathcal{K}_1,\mathcal{K}_2}^{\frac{1}{2}}(\widehat{\boldsymbol{\Psi}}_{\mathcal{K}_1,\mathcal{K}_2} - \boldsymbol{\Psi}_{\mathcal{K}_1,\mathcal{K}_2}) \in \mathcal{Y} | \widehat{\nu} = \nu) - \boldsymbol{\Phi}(\mathcal{Y})| \\
= \;& o(1).
\end{aligned}
$$

This proves the theorem.

A.9 Proof of Theorem 9

Without loss of generality, we consider the case where $\tau \in [n_0, \tau_0]$, as the convergence rate for $\tau \in [\tau_0, n-n_0]$ can be similarly derived. The idea is to break the time interval $[n-n_0, \tau_0]$ into two consecutive parts: $[n_0, \tau_{n,p}]$ and $[\tau_{n,p}, \tau_0]$, where $\tau_{n,p} = \left\lfloor \tau_0 - \kappa n \Delta_F^{-2}\left[\frac{\log(np)}{n} + \sqrt{\frac{\log(np)}{np^2}}\right]\right\rfloor$ for some large enough $\kappa > 0$. Here $\lfloor\cdot\rfloor$ denotes the least integer function. We shall show that when $\tau \in [n-n_0, \tau_{n,p}]$, in which $\widehat{\nu}^{\tau+1,n}$ might be inconsistent in estimating $\nu^{\tau_0+1,n}$, we have $\sup_{\tau\in[n_0,\tau_{n,p}]}[\mathbb{M}_n(\tau)-\mathbb{M}_n(\tau_0)] < 0$ in probability. Hence $\arg\max_{\tau\in[n_0,\tau_0]}\mathbb{M}_n(\tau) = \arg\max_{\tau\in[\tau_{n,p},\tau_0]}\mathbb{M}_n(\tau)$ holds in probability. On the other hand, when $\tau \in [\tau_{n,p}, \tau_0]$, we shall see that the membership maps can be consistently recovered, and hence the convergence rate can be obtained using classical probabilistic arguments. For simplicity, we consider the case where $\nu^{1,\tau_0} = \nu^{\tau_0+1,n} = \nu$ first, and modification of the proofs for the case where $\nu^{1,\tau_0} \neq \nu^{\tau_0+1,n}$ will be provided subsequently.

A.9.1 Change point estimation with $\nu^{1,\tau_0} = \nu^{\tau_0+1,n} = \nu$.

We first consider the case where the membership structures remain unchanged, while the connectivity matrices before/after the change point are different. Specifically, we assume that $\nu^{1,\tau_0} = \nu^{\tau_0+1,n} = \nu$ for some ν, and $(\theta_{1,k,\ell}, \eta_{1,k,\ell}) \neq (\theta_{2,k,\ell}, \eta_{2,k,\ell})$ for some $1 \le k \le l \le q$. For brevity, we shall be using the notations $S_{k,l}$, s_k, $s_{\min}$ and $n_{k,\ell}$ defined as in Section 3, and introduce some new notations as follows:

Define

$$\theta^{\tau}_{2,k,\ell} = \frac{\frac{\tau_0-\tau}{n-\tau}\frac{\theta_{1,k,\ell}\eta_{1,k,\ell}}{\theta_{1,k,\ell}+\eta_{1,k,\ell}} + \frac{n-\tau_0}{n-\tau}\frac{\theta_{2,k,\ell}\eta_{2,k,\ell}}{\theta_{2,k,\ell}+\eta_{2,k,\ell}}}{\frac{\tau_0-\tau}{n-\tau}\frac{\eta_{1,k,\ell}}{\theta_{1,k,\ell}+\eta_{1,k,\ell}} + \frac{n-\tau_0}{n-\tau}\frac{\eta_{2,k,\ell}}{\theta_{2,k,\ell}+\eta_{2,k,\ell}}}, \quad \eta^{\tau}_{2,k,\ell} = \frac{\frac{\tau_0-\tau}{n-\tau}\frac{\theta_{1,k,\ell}\eta_{1,k,\ell}}{\theta_{1,k,\ell}+\eta_{1,k,\ell}} + \frac{n-\tau_0}{n-\tau}\frac{\theta_{2,k,\ell}\eta_{2,k,\ell}}{\theta_{2,k,\ell}+\eta_{2,k,\ell}}}{\frac{\tau_0-\tau}{n-\tau}\frac{\theta_{1,k,\ell}}{\theta_{1,k,\ell}+\eta_{1,k,\ell}} + \frac{n-\tau_0}{n-\tau}\frac{\theta_{2,k,\ell}}{\theta_{2,k,\ell}+\eta_{2,k,\ell}}}.$$

Clearly when $\tau = \tau_0$ we have $\theta^{\tau_0}_{2,k,\ell} = \theta_{2,k,\ell}$ and $\eta^{\tau_0}_{2,k,\ell} = \eta_{2,k,\ell}$.

Correspondingly, we denote the MLEs as

$$\widehat{\theta}^{\tau}_{1,k,\ell} = \sum_{(i,j)\in\widehat{S}^{\tau}_{1,k,\ell}}\sum_{t=1}^{\tau} X^t_{i,j}(1-X^{t-1}_{i,j}) \Big/ \sum_{(i,j)\in\widehat{S}^{\tau}_{1,k,\ell}}\sum_{t=1}^{\tau}(1-X^{t-1}_{i,j}),$$

$$\widehat{\eta}^{\tau}_{1,k,\ell} = \sum_{(i,j)\in\widehat{S}^{\tau}_{1,k,\ell}}\sum_{t=1}^{\tau}(1-X^t_{i,j})X^{t-1}_{i,j} \Big/ \sum_{(i,j)\in\widehat{S}^{\tau}_{1,k,\ell}}\sum_{t=1}^{\tau} X^{t-1}_{i,j},$$

$$\widehat{\theta}^{\tau}_{2,k,\ell} = \sum_{(i,j)\in\widehat{S}^{\tau}_{2,k,\ell}}\sum_{t=\tau+1}^{n} X^t_{i,j}(1-X^{t-1}_{i,j}) \Big/ \sum_{(i,j)\in\widehat{S}^{\tau}_{2,k,\ell}}\sum_{t=\tau+1}^{n}(1-X^{t-1}_{i,j}),$$

$$\widehat{\eta}^{\tau}_{2,k,\ell} = \sum_{(i,j)\in\widehat{S}^{\tau}_{2,k,\ell}}\sum_{t=\tau+1}^{n}(1-X^t_{i,j})X^{t-1}_{i,j} \Big/ \sum_{(i,j)\in\widehat{S}^{\tau}_{2,k,\ell}}\sum_{t=\tau+1}^{n} X^{t-1}_{i,j},$$

where $\widehat{S}^{\tau}_{1,k,\ell}$ and $\widehat{S}^{\tau}_{2,k,\ell}$ are defined in a similar way to $\widehat{S}_{k,\ell}$ (cf. Section 3.2.3), based on the estimated memberships $\widehat{\nu}^{1,\tau}$ and $\widehat{\nu}^{\tau+1,n}$, respectively.

Denote

$$\mathbb{M}_n(\tau) := l(\{\widehat{\theta}^{\tau}_{1,k,\ell}, \widehat{\eta}^{\tau}_{1,k,\ell}\};\ \widehat{\nu}^{1,\tau}) + l(\{\widehat{\theta}^{\tau}_{2,k,\ell}, \widehat{\eta}^{\tau}_{2,k,\ell}\};\ \widehat{\nu}^{\tau+1,n}),$$
$$\mathbb{M}(\tau) := El(\{\theta_{1,k,\ell}, \eta_{1,k,\ell}\}; \nu^{1,\tau}) + El(\{\theta^{\tau}_{2,k,\ell}, \eta^{\tau}_{2,k,\ell}\}; \nu^{\tau+1,n}).$$

We first evaluate several terms in (i)-(v), and all these results will be combined to obtain the error bound in (vi). In particular, (vi) states that as a direct result of (v), we can focus on the small neighborhood of $[\tau_{n,p}, \tau_0]$ when searching for the estimator $\widehat{\tau}$. Further, the inequality (A.43) transforms the error bound for $\tau_0 - \widehat{\tau}$ into the error bounds of the terms that we derived in (i)-(iv).

(i) Evaluating $\mathbb{M}(\tau) - \mathbb{M}(\tau_0)$.

Note that $\tau_0 = \arg\max_{n_0 \le \tau \le n-n_0} \mathbb{M}(\tau)$, and for any $\tau \in [n_0, \tau_0]$,

$$\begin{aligned}
\mathbb{M}(\tau) - \mathbb{M}(\tau_0) &= El(\{\theta_{1,k,\ell}, \eta_{1,k,\ell}\}; \nu^{1,\tau}) + El(\{\theta^{\tau}_{2,k,\ell}, \eta^{\tau}_{2,k,\ell}\}; \nu^{\tau+1,n}) \\
&\quad - El(\{\theta_{1,k,\ell}, \eta_{1,k,\ell}\}; \nu^{1,\tau_0}) - El(\{\theta_{2,k,\ell}, \eta_{2,k,\ell}\}; \nu^{\tau_0+1,n}) \\
&= El(\{\theta^{\tau}_{2,k,\ell}, \eta^{\tau}_{2,k,\ell}\}; \nu^{\tau+1,\tau_0}) - El(\{\theta_{1,k,\ell}, \eta_{1,k,\ell}\}; \nu^{\tau+1,\tau_0}) \\
&\quad + El(\{\theta^{\tau}_{2,k,\ell}, \eta^{\tau}_{2,k,\ell}\}; \nu^{\tau_0+1,n}) - El(\{\theta_{2,k,\ell}, \eta_{2,k,\ell}\}; \nu^{\tau_0+1,n}).
\end{aligned}$$

Recall that

$$\begin{aligned}
l(\{\theta_{k,\ell}, \eta_{k,\ell}\}; \nu) = \sum_{1\le k\le \ell\le q} \sum_{(i,j)\in S_{k,l}} \sum_{t=1}^{n} \Big\{ X^t_{i,j}(1 - X^{t-1}_{i,j}) \log \theta_{k,\ell} \\
+ (1 - X^t_{i,j})(1 - X^{t-1}_{i,j}) \log(1 - \theta_{k,\ell}) + (1 - X^t_{i,j}) X^{t-1}_{i,j} \log \eta_{k,\ell} + X^t_{i,j} X^{t-1}_{i,j} \log(1 - \eta_{k,\ell}) \Big\}.
\end{aligned}$$

By Taylor expansion and the fact that the partial derivative of the expected likelihood evaluated at the true values equals zero we have, there exist $\theta^*_{k,\ell} \in [\theta_{1,k,\ell}, \theta^{\tau}_{2,k,\ell}], \eta^*_{k,\ell} \in [\eta_{1,k,\ell}, \eta^{\tau}_{2,k,\ell}], 1 \le k \le \ell \le q$, such that

$$\begin{aligned}
& El(\{\theta^{\tau}_{2,k,\ell}, \eta^{\tau}_{2,k,\ell}\}; \nu^{\tau+1,\tau_0}) - El(\{\theta_{1,k,\ell}, \eta_{1,k,\ell}\}; \nu^{\tau+1,\tau_0}) \\
= & -\sum_{1\le k\le \ell\le q} s_{k,\ell}(\tau_0 - \tau) \Bigg\{ \frac{\theta_{1,k,\ell}\eta_{1,k,\ell}}{\theta_{1,k,\ell} + \eta_{1,k,\ell}} \Big(\frac{\theta^{\tau}_{2,k,\ell} - \theta_{1,k,\ell}}{\theta^*_{k,\ell}} \Big)^2 + \frac{(1 - \theta_{1,k,\ell})\eta_{1,k,\ell}}{\theta_{1,k,\ell} + \eta_{1,k,\ell}} \Big(\frac{\theta^{\tau}_{2,k,\ell} - \theta_{1,k,\ell}}{1 - \theta^*_{k,\ell}} \Big)^2 \\
& + \frac{\theta_{1,k,\ell}\eta_{1,k,\ell}}{\theta_{1,k,\ell} + \eta_{1,k,\ell}} \Big(\frac{\eta^{\tau}_{2,k,\ell} - \eta_{1,k,\ell}}{\eta^*_{k,\ell}} \Big)^2 + \frac{(1 - \eta_{1,k,\ell})\theta_{1,k,\ell}}{\theta_{1,k,\ell} + \eta_{1,k,\ell}} \Big(\frac{\eta^{\tau}_{2,k,\ell} - \eta_{1,k,\ell}}{1 - \eta^*_{k,\ell}} \Big)^2 \Bigg\} \\
\le & -C_1(\tau_0 - \tau) \sum_{1\le k\le \ell\le q} s_{k,\ell} [(\theta_{1,k,\ell} - \theta_{2,k,\ell})^2 + (\eta_{1,k,\ell} - \eta_{2,k,\ell})^2] \\
\le & -C_1(\tau_0 - \tau) \big[\|\mathbf{W}_{1,1} - \mathbf{W}_{2,1}\|_F^2 + \|\mathbf{W}_{1,2} - \mathbf{W}_{2,2}\|_F^2 \big],
\end{aligned}$$

for some constant $C_1 > 0$. Here in the first step we have used the fact that for any $(i, j) \in S_{k,\ell}$ and $t \le \tau_0$, $EX^t_{i,j}(1 - X^{t-1}_{i,j}) = EX^{t-1}_{i,j}(1 - X^t_{i,j}) = \frac{\theta_{1,k,\ell}\eta_{1,k,\ell}}{\theta_{1,k,\ell}+\eta_{1,k,\ell}}$, $E(1 - X^t_{i,j})(1 - X^{t-1}_{i,j}) = \frac{(1-\theta_{1,k,\ell})\eta_{1,k,\ell}}{\theta_{1,k,\ell}+\eta_{1,k,\ell}}$, and $EX_{i,j}^t X_{i,j}^{t-1} = \frac{(1-\eta_{1,k,\ell})\theta_{1,k,\ell}}{\theta_{1,k,\ell}+\eta_{1,k,\ell}}$. Similarly, there exist $\theta_{k,\ell}^\dagger \in [\theta_{2,k,\ell}, \theta_{2,k,\ell}^\tau], \eta_{k,\ell}^\dagger \in [\eta_{2,k,\ell}, \eta_{2,k,\ell}^\tau], 1 \leq k \leq \ell \leq q$, such that

$$\begin{aligned}
& El(\{\theta_{2,k,\ell}^\tau, \eta_{2,k,\ell}^\tau\}; \nu^{\tau_0+1,n}) - El(\{\theta_{2,k,\ell}, \eta_{2,k,\ell}\}; \nu^{\tau_0+1,n}) \\
= & -\sum_{1\leq k\leq \ell\leq q} s_{k,\ell}(n-\tau_0)\Bigg\{ \frac{\theta_{2,k,\ell}\eta_{2,k,\ell}}{\theta_{2,k,\ell}+\eta_{2,k,\ell}}\Big(\frac{\theta_{2,k,\ell}^\tau - \theta_{2,k,\ell}}{\theta_{k,\ell}^\dagger}\Big)^2 + \frac{(1-\theta_{2,k,\ell})\eta_{2,k,\ell}}{\theta_{2,k,\ell}+\eta_{2,k,\ell}}\Big(\frac{\theta_{2,k,\ell}^\tau - \theta_{2,k,\ell}}{1-\theta_{k,\ell}^\dagger}\Big)^2 \\
& + \frac{\theta_{2,k,\ell}\eta_{2,k,\ell}}{\theta_{2,k,\ell}+\eta_{2,k,\ell}}\Big(\frac{\eta_{2,k,\ell}^\tau - \eta_{2,k,\ell}}{\eta_{k,\ell}^\dagger}\Big)^2 + \frac{(1-\eta_{2,k,\ell})\theta_{2,k,\ell}}{\theta_{2,k,\ell}+\eta_{2,k,\ell}}\Big(\frac{\eta_{2,k,\ell}^\tau - \eta_{2,k,\ell}}{1-\eta_{k,\ell}^\dagger}\Big)^2 \Bigg\} \\
\leq & -C_2'(n-\tau_0)\sum_{1\leq k\leq \ell\leq q} \frac{s_{k,\ell}(\tau_0-\tau)^2}{(n-\tau)^2}[(\theta_{1,k,\ell}-\theta_{2,k,\ell})^2 + (\eta_{1,k,\ell}-\eta_{2,k,\ell})^2] \\
\leq & -\frac{C_2(\tau_0-\tau)^2}{n-\tau}\big[\|\mathbf{W}_{1,1}-\mathbf{W}_{2,1}\|_F^2 + \|\mathbf{W}_{1,2}-\mathbf{W}_{2,2}\|_F^2\big],
\end{aligned}$$

for some constants $C_2', C_2 > 0$. Consequently, we conclude that there exists a constant $C_3 > 0$ such that for any $n_0 \leq \tau \leq \tau_0$, we have

$$\mathbb{M}(\tau) - \mathbb{M}(\tau_0) \leq -C_3(\tau_0-\tau)\big[\|\mathbf{W}_{1,1}-\mathbf{W}_{2,1}\|_F^2 + \|\mathbf{W}_{1,2}-\mathbf{W}_{2,2}\|_F^2\big]. \tag{A.25}$$

(ii) Evaluating $\sup_{\tau\in[\tau_{n,p},\tau_0]} P(\widehat{\nu}(\tau) \neq \nu)$.

Let $\widehat{\nu}(\tau)$ be either $\widehat{\nu}^{1,\tau}$ or $\widehat{\nu}^{\tau+1,n}$. Note that the membership maps of the networks before/after τ remain to be ν. From Theorems 5 and 6, we have, under Conditions C2-C4, for any constant $B > 0$, there exists a large enough constant C_B such that

$$\sup_{\tau\in[\tau_{n,p},\tau_0]} P(\widehat{\nu}(\tau) \neq \nu) \leq C_B(\tau_0-\tau_{n,p})p[(pn)^{-(B+1)} + \exp\{-B\sqrt{p}\}].$$

Note that by choosing B to be large enough, we have $p(\tau_0-\tau_{n,p})(pn)^{-(B+1)} = o\left(\sqrt{\frac{(\tau_0-\tau_{n,p})\log(np)}{n^2 s_{\min}^2}}\right)$. On the other hand, the assumption that $\frac{\log(np)}{\sqrt{p}} \to 0$ in Condition C4 implies $pn\sqrt{\frac{(\tau_0-\tau_{n,p})s_{\min}^2}{\log(np)}} = o(\exp\{B\sqrt{p}\})$ for some large enough constant B. Consequently, we have $(\tau_0-\tau_{n,p})p\exp\{-B\sqrt{p}\} = o\left(\sqrt{\frac{(\tau_0-\tau_{n,p})\log(np)}{n^2 s_{\min}^2}}\right)$, and hence we conclude that $\sup_{\tau\in[\tau_{n,p},\tau_0]} P(\widehat{\nu}(\tau) \neq \nu) = o\left(\sqrt{\frac{(\tau_0-\tau_{n,p})\log(np)}{n^2 s_{\min}^2}}\right)$.

(iii) Evaluating $\sup_{\tau\in[\tau_{n,p},\tau_0]}[\mathbb{M}_n(\tau) - \mathbb{M}(\tau)]$ when $\widehat{\nu}(\tau) = \nu$.

From (ii) we have with probability greater than $1 - o\left(\sqrt{\frac{(\tau_0-\tau_{n,p})\log(np)}{n^2 s_{\min}^2}}\right)$, $\widehat{\nu}(\tau) = \nu$ for all $\tau \in [\tau_{n,p}, \tau_0]$. For simplicity, in this part we assume that $\widehat{S}_{1,k,\ell}^\tau = \widehat{S}_{2,k,\ell}^\tau = S_{k,l}$ (or equivalently $\widehat{\nu}^{1,\tau} = \widehat{\nu}^{\tau+1,n} = \nu$) holds for all $1 \leq k \leq \ell \leq q$ and $\tau_{n,p} \leq \tau \leq \tau_0$ without indicating that this holds in probability.

Denote

$$\begin{aligned}
g_{1,i,j}(\theta, \eta; \tau) = \sum_{t=1}^{\tau} \Big\{ & X_{i,j}^t(1-X_{i,j}^{t-1})\log\theta \\
& + (1-X_{i,j}^t)(1-X_{i,j}^{t-1})\log(1-\theta) + (1-X_{i,j}^t)X_{i,j}^{t-1}\log\eta + X_{i,j}^t X_{i,j}^{t-1}\log(1-\eta)\Big\},
\end{aligned}$$

and

$$g_{2,i,j}(\theta,\eta;\tau) = \sum_{t=\tau+1}^{n} \Big\{ X_{i,j}^t(1-X_{i,j}^{t-1})\log\theta$$
$$+(1-X_{i,j}^t)(1-X_{i,j}^{t-1})\log(1-\theta) + (1-X_{i,j}^t)X_{i,j}^{t-1}\log\eta + X_{i,j}^t X_{i,j}^{t-1}\log(1-\eta)\Big\}.$$

When $\widehat{\nu}=\nu$, we have,

$$\begin{aligned}
& \mathbb{M}_n(\tau) - \mathbb{M}(\tau) \qquad\qquad (A.26)\\
= & \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\widehat{\theta}_{1,k,\ell}^{\tau},\widehat{\eta}_{1,k,\ell}^{\tau};\tau) + \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{2,i,j}(\widehat{\theta}_{2,k,\ell}^{\tau},\widehat{\eta}_{2,k,\ell}^{\tau};\tau)\\
& -E\sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\theta_{1,k,\ell},\eta_{1,k,\ell};\tau) - E\sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{2,i,j}(\theta_{2,k,\ell}^{\tau},\eta_{2,k,\ell}^{\tau};\tau)\\
= & \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\widehat{\theta}_{1,k,\ell}^{\tau},\widehat{\eta}_{1,k,\ell}^{\tau};\tau) + \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{2,i,j}(\widehat{\theta}_{2,k,\ell}^{\tau},\widehat{\eta}_{2,k,\ell}^{\tau};\tau)\\
& -\sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\theta_{1,k,\ell},\eta_{1,k,\ell};\tau) - \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{2,i,j}(\theta_{2,k,\ell}^{\tau},\eta_{2,k,\ell}^{\tau};\tau)\\
& +\sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\theta_{1,k,\ell},\eta_{1,k,\ell};\tau) + \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{2,i,j}(\theta_{2,k,\ell}^{\tau},\eta_{2,k,\ell}^{\tau};\tau)\\
& -E\sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\theta_{1,k,\ell},\eta_{1,k,\ell};\tau) - E\sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{2,i,j}(\theta_{2,k,\ell}^{\tau},\eta_{2,k,\ell}^{\tau};\tau)
\end{aligned}$$

Note that $\{\widehat{\theta}_{1,k,\ell}^{\tau},\widehat{\eta}_{1,k,\ell}^{\tau}\}$ is the maximizer of $\sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\theta_{k,\ell},\eta_{k,\ell};\tau)$. Applying Taylor's expansion we have, there exist random scalers $\theta_{k,\ell}^{-}\in[\widehat{\theta}_{1,k,\ell}^{\tau},\theta_{1,k,\ell}]$, $\eta_{k,\ell}^{-}\in[\widehat{\eta}_{1,k,\ell}^{\tau},\eta_{1,k,\ell}]$ such that

$$\begin{aligned}
& \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\widehat{\theta}_{1,k,\ell}^{\tau},\widehat{\eta}_{1,k,\ell}^{\tau};\tau) - \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\theta_{1,k,\ell},\eta_{1,k,\ell};\tau)\\
\le & \sum_{1\le k\le \ell\le q} s_{k,\ell}\tau\Big\{\Big(\frac{\theta_{1,k,\ell}-\widehat{\theta}_{1,k,\ell}^{\tau}}{\theta_{k,\ell}^{-}}\Big)^2 + \Big(\frac{\theta_{1,k,\ell}-\widehat{\theta}_{1,k,\ell}^{\tau}}{1-\theta_{k,\ell}^{-}}\Big)^2 + \Big(\frac{\eta_{1,k,\ell}-\widehat{\eta}_{1,k,\ell}^{\tau}}{\eta_{k,\ell}^{-}}\Big)^2 + \Big(\frac{\eta_{1,k,\ell}-\widehat{\eta}_{1,k,\ell}^{\tau}}{1-\eta_{k,\ell}^{-}}\Big)^2\Big\}.
\end{aligned}$$

On the other hand, when $\widehat{\nu}=\nu$, similar to Theorem 3 and Theorem 7, we can show that for any $B>0$, there exists a large enough constant C^{-} such that $\max_{1\le k\le \ell\le q,\tau\in[\tau_{n,p},\tau_0]}|\widehat{\theta}_{1,k,\ell}^{\tau}-\theta_{1,k,\ell}|\le C^{-}\sqrt{\frac{\log(np)}{ns_{\min}^2}}$, and $\max_{1\le k\le \ell\le q,\tau\in[\tau_{n,p},\tau_0]}|\widehat{\eta}_{1,k,\ell}^{\tau}-\eta_{1,k,\ell}| = C^{-}\sqrt{\frac{\log(np)}{ns_{\min}^2}}$ hold with probability greater than $1-O((np)^{-B})$. Consequently, we have, when $\widehat{\nu}=\nu$, there exits a large enough constant $C_4>0$ such that

$$\begin{aligned}
& \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\widehat{\theta}_{1,k,\ell}^{\tau},\widehat{\eta}_{1,k,\ell}^{\tau};\tau) - \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\theta_{1,k,\ell},\eta_{1,k,\ell};\tau)\\
\le & C_4\tau\sum_{1\le k\le \ell\le q} s_{k,\ell}\frac{\log(np)}{ns_{\min}^2}\\
\le & \frac{C_4\tau p^2\log(np)}{ns_{\min}^2}. \qquad\qquad (A.27)
\end{aligned}$$

Similarly, we have there exists a large enough constant $C_5 > 0$ such that with probability greater than $1 - O((np)^{-B})$,

$$
\begin{aligned}
& \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{2,i,j}(\widehat{\theta}^{\tau}_{2,k,\ell}, \widehat{\eta}^{\tau}_{2,k,\ell}; \tau) - \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{2,i,j}(\theta^{\tau}_{2,k,\ell}, \eta^{\tau}_{2,k,\ell}; \tau) \\
\le \; & \frac{C_5(n-\tau)p^2\log(np)}{ns^2_{\min}}. \qquad \text{(A.28)}
\end{aligned}
$$

On the other hand, similar to Lemma 3, there exists a constant $C_6 > 0$ such that with probability greater than $1 - O((np)^{-B})$,

$$
\begin{aligned}
& \left| \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\theta_{1,k,\ell}, \eta_{1,k,\ell}; \tau) - E\sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\theta_{1,k,\ell}, \eta_{1,k,\ell}; \tau)\right| \\
\le \; & C_6\tau p^2\sqrt{\frac{\log(np)}{\tau p^2}}, \qquad \text{(A.29)}
\end{aligned}
$$

and

$$
\begin{aligned}
& \left| \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{2,i,j}(\theta^{\tau}_{2,k,\ell}, \eta^{\tau}_{2,k,\ell}; \tau) - E\sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{2,i,j}(\theta^{\tau}_{2,k,\ell}, \eta^{\tau}_{2,k,\ell}; \tau)\right| \\
\le \; & C_6(n-\tau)p^2\sqrt{\frac{\log(np)}{(n-\tau)p^2}}. \qquad \text{(A.30)}
\end{aligned}
$$

Combining (A.26), (A.27), (A.28), (A.29) and (A.30) we conclude that when $\widehat{\nu} = \nu$, there exists a large enough constant $C_0 > 0$ such that with probability greater than $1 - O((np)^{-B})$,

$$
\sup_{\tau\in[\tau_{n,p},\tau_0]} |\mathbb{M}_n(\tau) - \mathbb{M}(\tau)| \le C_0 np^2 \left\{ \frac{\log(np)}{ns^2_{\min}} + \sqrt{\frac{\log(np)}{np^2}} \right\} = O\left(np^2\sqrt{\frac{\log(np)}{ns^2_{\min}}} \right). \qquad \text{(A.31)}
$$

(iv) Evaluating $E\sup_{\tau\in[\tau_{n,p},\tau_0]} |\mathbb{M}_n(\tau) - \mathbb{M}(\tau) - \mathbb{M}_n(\tau_0) + \mathbb{M}(\tau_0)|$.

Notice that when $\widehat{\nu} = \nu$,

$$
\begin{aligned}
& \mathbb{M}_n(\tau) - \mathbb{M}(\tau) - \mathbb{M}_n(\tau_0) + \mathbb{M}(\tau_0) \\
= \; & \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\widehat{\theta}^{\tau}_{1,k,\ell}, \widehat{\eta}^{\tau}_{1,k,\ell}; \tau) + \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{2,i,j}(\widehat{\theta}^{\tau}_{2,k,\ell}, \widehat{\eta}^{\tau}_{1,k,\ell}; \tau) \\
& -E\sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\theta_{1,k,\ell}, \eta_{1,k,\ell}; \tau) - E\sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{2,i,j}(\theta^{\tau}_{2,k,\ell}, \eta^{\tau}_{2,k,\ell}; \tau) \\
& -\sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\widehat{\theta}^{\tau_0}_{1,k,\ell}, \widehat{\eta}^{\tau_0}_{1,k,\ell}; \tau_0) - \sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{2,i,j}(\widehat{\theta}^{\tau_0}_{2,k,\ell}, \widehat{\eta}^{\tau_0}_{2,k,\ell}; \tau_0) \\
& +E\sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{1,i,j}(\theta_{1,k,\ell}, \eta_{1,k,\ell}; \tau_0) + E\sum_{1\le k\le \ell\le q}\sum_{(i,j)\in S_{k,\ell}} g_{2,i,j}(\theta_{2,k,\ell}, \eta_{2,k,\ell}; \tau_0)
\end{aligned}
$$

Note that

$$
\begin{aligned}
& g_{1,i,j}(\widehat{\theta}^{\tau}_{1,k,\ell}, \widehat{\eta}^{\tau}_{1,k,\ell}; \tau) - g_{1,i,j}(\widehat{\theta}^{\tau_0}_{1,k,\ell}, \widehat{\eta}^{\tau_0}_{1,k,\ell}; \tau_0) - E[g_{1,i,j}(\theta_{1,k,\ell}, \eta_{1,k,\ell}; \tau) - g_{1,i,j}(\theta_{1,k,\ell}, \eta_{1,k,\ell}; \tau_0)] \\
= & \sum_{t=1}^{\tau} \left\{ X^t_{i,j}(1 - X^{t-1}_{i,j}) \log \frac{\widehat{\theta}^{\tau}_{1,k,\ell}}{\widehat{\theta}^{\tau_0}_{1,k,\ell}} + (1 - X^t_{i,j})(1 - X^{t-1}_{i,j}) \log \frac{1 - \widehat{\theta}^{\tau}_{1,k,\ell}}{1 - \widehat{\theta}^{\tau_0}_{1,k,\ell}} \right. \\
& \left. + (1 - X^t_{i,j}) X^{t-1}_{i,j} \log \frac{\widehat{\eta}^{\tau}_{1,k,\ell}}{\widehat{\eta}^{\tau_0}_{1,k,\ell}} + X^t_{i,j} X^{t-1}_{i,j} \log \frac{1 - \widehat{\eta}^{\tau}_{1,k,\ell}}{1 - \widehat{\eta}^{\tau_0}_{1,k,\ell}} \right\} - \sum_{t=\tau+1}^{\tau_0} \left\{ X^t_{i,j}(1 - X^{t-1}_{i,j}) \log \widehat{\theta}^{\tau_0}_{1,k,\ell} \right. \\
& \left. + (1 - X^t_{i,j})(1 - X^{t-1}_{i,j}) \log(1 - \widehat{\theta}^{\tau_0}_{1,k,\ell}) + (1 - X^t_{i,j}) X^{t-1}_{i,j} \log \widehat{\eta}^{\tau_0}_{1,k,\ell} + X^t_{i,j} X^{t-1}_{i,j} \log(1 - \widehat{\eta}^{\tau_0}_{1,k,\ell}) \right\} \\
& + E \sum_{t=\tau+1}^{\tau_0} \left\{ X^t_{i,j}(1 - X^{t-1}_{i,j}) \log \theta_{1,k,\ell} + (1 - X^t_{i,j})(1 - X^{t-1}_{i,j}) \log(1 - \theta_{1,k,\ell}) \right. \\
& \left. + (1 - X^t_{i,j}) X^{t-1}_{i,j} \log \eta_{1,k,\ell} + X^t_{i,j} X^{t-1}_{i,j} \log(1 - \eta_{1,k,\ell}) \right\}.
\end{aligned}
$$

When sum over all $(i, j) \in S_{k,\ell}$ and $1 \le k \le \ell \le q$, the last two terms in the above inequality can be bounded similar to (A.27) and (A.29), with τ replaced by $\tau_0 - \tau$. For the first term, with some calculations we have there exists a constant $c_1 > 0$ such that with probability larger than $1 - O(np)^{-B}$),

$$
\begin{aligned}
\sup_{1 \le k \le \ell \le q} \left| \widehat{\theta}^{\tau}_{1,k,\ell} - \widehat{\theta}^{\tau_0}_{1,k,\ell} \right| &\le c_1 \sqrt{\frac{\tau_0 - \tau}{\tau_0}} \sqrt{\frac{\log(np)}{n s^2_{\min}}}, \\
\sup_{1 \le k \le \ell \le q} \left| \widehat{\eta}^{\tau}_{1,k,\ell} - \widehat{\eta}^{\tau_0}_{1,k,\ell} \right| &\le c_1 \sqrt{\frac{\tau_0 - \tau}{\tau_0}} \sqrt{\frac{\log(np)}{n s^2_{\min}}}.
\end{aligned} \tag{A.32}
$$

Brief derivations of (A.32) are provided in Section A.9.3. Consequently, similar to (A.31), we have there exists a large enough constant $c_2 > 0$ such that

$$
\begin{aligned}
& \left| \sum_{1 \le k \le \ell \le q} \sum_{(i,j) \in S_{k,\ell}} \left[g_{1,i,j}(\widehat{\theta}^{\tau}_{1,k,\ell}, \widehat{\eta}^{\tau}_{1,k,\ell}; \tau) - g_{1,i,j}(\widehat{\theta}^{\tau_0}_{1,k,\ell}, \widehat{\eta}^{\tau_0}_{1,k,\ell}; \tau_0) \right] \right. \\
& \left. - E \sum_{1 \le k \le \ell \le q} \sum_{(i,j) \in S_{k,\ell}} \left[g_{1,i,j}(\theta_{1,k,\ell}, \eta_{1,k,\ell}; \tau) - g_{1,i,j}(\theta_{1,k,\ell}, \eta_{1,k,\ell}; \tau_0) \right] \right| \\
\le \; & c_2 p^2 \sqrt{\frac{(\tau_0 - \tau) \log(np)}{s^2_{\min}}}.
\end{aligned} \tag{A.33}
$$

Here in the last step we have used the fact that $\tau_0 \simeq O(n)$, $\sqrt{\frac{\log(np)}{p^2}} \le \sqrt{\frac{\log(np)}{s^2_{\min}}}$, and $\frac{(\tau_0 - \tau)\log(np)}{n s^2_{\min}} =$

$o\Big(\sqrt{\frac{(\tau_0-\tau)\log(np)}{s_{\min}^2}}\Big)$. Similarly, note that,

$$\begin{aligned}
& g_{2,i,j}(\widehat{\theta}^{\tau}_{2,k,\ell},\widehat{\eta}^{\tau}_{2,k,\ell};\tau)-g_{2,i,j}(\widehat{\theta}^{\tau_0}_{2,k,\ell},\widehat{\eta}^{\tau_0}_{2,k,\ell};\tau_0)-E[g_{2,i,j}(\theta^{\tau}_{2,k,\ell},\eta^{\tau}_{2,k,\ell};\tau)-g_{2,i,j}(\theta_{2,k,\ell},\eta_{2,k,\ell};\tau_0)] \\
= & \sum_{t=\tau_0+1}^{n}\Bigg\{X^t_{i,j}(1-X^{t-1}_{i,j})\Bigg[\log\frac{\widehat{\theta}^{\tau}_{2,k,\ell}}{\widehat{\theta}^{\tau_0}_{2,k,\ell}}-\log\frac{\theta^{\tau}_{2,k,\ell}}{\theta_{2,k,\ell}}\Bigg]+(1-X^t_{i,j})(1-X^{t-1}_{i,j})\cdot\Bigg[\log\frac{1-\widehat{\theta}^{\tau}_{2,k,\ell}}{1-\widehat{\theta}^{\tau_0}_{2,k,\ell}} \\
& -\log\frac{1-\theta^{\tau}_{2,k,\ell}}{1-\theta_{2,k,\ell}}\Bigg]+X^t_{i,j}(1-X^{t-1}_{i,j})\Bigg[\log\frac{\widehat{\eta}^{\tau}_{2,k,\ell}}{\widehat{\eta}^{\tau_0}_{2,k,\ell}}-\log\frac{\eta^{\tau}_{2,k,\ell}}{\eta_{2,k,\ell}}\Bigg]+X^t_{i,j}X^{t-1}_{i,j}\Bigg[\log\frac{1-\widehat{\eta}^{\tau}_{2,k,\ell}}{1-\widehat{\eta}^{\tau_0}_{2,k,\ell}} \\
& -\log\frac{1-\eta^{\tau}_{2,k,\ell}}{1-\eta_{2,k,\ell}}\Bigg]\Bigg\}+[g_{2,i,j}(\theta^{\tau}_{2,k,\ell},\eta^{\tau}_{2,k,\ell};\tau_0)-g_{2,i,j}(\theta_{2,k,\ell},\eta_{2,k,\ell};\tau_0)] \\
& -E[g_{2,i,j}(\theta^{\tau}_{2,k,\ell},\eta^{\tau}_{2,k,\ell};\tau_0)-g_{2,i,j}(\theta_{2,k,\ell},\eta_{2,k,\ell};\tau_0)]+\sum_{t=\tau+1}^{\tau_0}\Bigg\{X^t_{i,j}(1-X^{t-1}_{i,j})\log\widehat{\theta}^{\tau}_{2,k,\ell} \\
& +(1-X^t_{i,j})(1-X^{t-1}_{i,j})\log(1-\widehat{\theta}^{\tau}_{2,k,\ell})+(1-X^t_{i,j})X^{t-1}_{i,j}\log\widehat{\eta}^{\tau}_{2,k,\ell}+X^t_{i,j}X^{t-1}_{i,j}\log(1-\widehat{\eta}^{\tau}_{2,k,\ell})\Bigg\} \\
& -E\sum_{t=\tau+1}^{\tau_0}\Big\{X^t_{i,j}(1-X^{t-1}_{i,j})\log\theta^{\tau}_{2,k,\ell}+(1-X^t_{i,j})(1-X^{t-1}_{i,j})\log(1-\theta^{\tau}_{2,k,\ell}) \\
& +(1-X^t_{i,j})X^{t-1}_{i,j}\log\eta^{\tau}_{2,k,\ell}+X^t_{i,j}X^{t-1}_{i,j}\log(1-\eta^{\tau}_{2,k,\ell})\Big\} \\
:= & \; I+II-III+IV-V.
\end{aligned} \tag{A.34}$$

For $II-III$, from Lemma 3 and the fact that $\Big|\theta^{\tau}_{2,k,\ell}-\theta_{2,k,\ell}\Big|\le\frac{c_3(\tau_0-\tau)}{n-\tau}$, and $\Big|\eta^{\tau}_{2,k,\ell}-\eta_{2,k,\ell}\Big|\le\frac{c_3(\tau_0-\tau)}{n-\tau}$ for some large enough constant c_3, we have there exists a large enough constant $c_4>0$ such that with probability greater than $1-O((np)^{-B})$,

$$\Bigg|\sum_{1\le k\le\ell\le q}\sum_{(i,j)\in S_{k,\ell}}(II-III)\Bigg|\le c_4p^2\frac{\tau_0-\tau}{n-\tau}\sqrt{\frac{\log(np)}{\tau_0p^2}}=o\Bigg(p^2\sqrt{\frac{(\tau_0-\tau)\log(np)}{s_{\min}^2}}\Bigg). \tag{A.35}$$

When sum over all $(i,j)\in S_{k,\ell}$ and $1\le k\le\ell\le q$, the $IV-V$ term can be bounded similar to (A.27) and (A.29), with τ replaced by $\tau_0-\tau$, i.e., there exist a constant $c_5>0$ such that with probability greater than $1-O((np)^{-B})$,

$$\begin{aligned}
\Bigg|\sum_{1\le k\le\ell\le q}\sum_{(i,j)\in S_{k,\ell}}(IV-V)\Bigg| & \le c_5p^2\Bigg[\frac{(\tau_0-\tau)\log(np)}{ns_{\min}^2}+\sqrt{\tau_0-\tau}\sqrt{\frac{\log(np)}{p^2}}\Bigg] \\
& = O\Bigg(p^2\sqrt{\frac{(\tau_0-\tau)\log(np)}{s_{\min}^2}}\Bigg).
\end{aligned} \tag{A.36}$$

Lastly, similar to (A.32), we can show that there exists a constant $c_6>0$ such that with probability larger than $1-O(np)^{-B})$,

$$\begin{aligned}
\sup_{1\le k\le\ell\le q}\Bigg|\log\frac{\widehat{\theta}^{\tau}_{2,k,\ell}}{\theta^{\tau}_{2,k,\ell}}-\log\frac{\widehat{\theta}^{\tau_0}_{2,k,\ell}}{\theta_{2,k,\ell}}\Bigg| & \le c_6\sqrt{\frac{\tau_0-\tau}{n}}\sqrt{\frac{\log(np)}{ns_{\min}^2}}, \\
\sup_{1\le k\le\ell\le q}\Bigg|\log\frac{\widehat{\eta}^{\tau}_{2,k,\ell}}{\eta^{\tau}_{2,k,\ell}}-\log\frac{\widehat{\eta}^{\tau_0}_{2,k,\ell}}{\eta_{2,k,\ell}}\Bigg| & \le c_6\sqrt{\frac{\tau_0-\tau}{n}}\sqrt{\frac{\log(np)}{ns_{\min}^2}}.
\end{aligned} \tag{A.37}$$

A brief proof of (A.37) is provided in Section A.9.3. Consequently, we can show that there exists a constant $c_7 > 0$ such that with probability larger than $1 - O(np)^{-B})$,

$$\begin{aligned}
& \Bigg| \sum_{1\le k\le \ell\le q} \sum_{(i,j)\in S_{k,\ell}} \left[g_{2,i,j}(\widehat{\theta}^{\tau}_{2,k,\ell}, \widehat{\eta}^{\tau}_{2,k,\ell}; \tau) - g_{2,i,j}(\widehat{\theta}^{\tau_0}_{2,k,\ell}, \widehat{\eta}^{\tau_0}_{2,k,\ell}; \tau_0) \right] \\
& -E \sum_{1\le k\le \ell\le q} \sum_{(i,j)\in S_{k,\ell}} \left[g_{2,i,j}(\theta^{\tau}_{2,k,\ell}, \eta^{\tau}_{2,k,\ell}; \tau) - g_{2,i,j}(\theta_{2,k,\ell}, \eta_{2,k,\ell}; \tau_0) \right] \Bigg| \\
\le \; & c_7 p^2 \sqrt{\frac{(\tau_0 - \tau)\log(np)}{s^2_{\min}}}. \qquad (A.38)
\end{aligned}$$

Now combining (A.33) and (A.38) and the probability for $\widehat{\nu} \neq \nu$ in (ii), we conclude that there exists a constant $C_0 > 0$ such that

$$\begin{aligned}
& E \sup_{\tau\in[\tau_{n,p},\tau_0]} |\mathbb{M}_n(\tau) - \mathbb{M}(\tau) - \mathbb{M}_n(\tau_0) + \mathbb{M}(\tau_0)| \\
\le \; & C_0 np^2 \left\{ \sqrt{\frac{(\tau_0 - \tau_{n,p})\log(np)}{n^2 s^2_{\min}}} + o\left(\sqrt{\frac{(\tau_0 - \tau_{n,p})\log(np)}{n^2 s^2_{\min}}} \right) \right\} \\
\le \; & 2C_0 p^2 \sqrt{\frac{(\tau_0 - \tau_{n,p})\log(np)}{s^2_{\min}}}. \qquad (A.39)
\end{aligned}$$

(v) Evaluating $\sup_{\tau\in[n_0,\tau_{n,p}]}[\mathbb{M}_n(\tau) - \mathbb{M}_n(\tau_0)]$**.**

In this part we consider the case when $\tau \in [n - n_0, \tau_{n,p}]$. We shall see that $\sup_{\tau\in[n_0,\tau_{n,p}]}[\mathbb{M}_n(\tau) - \mathbb{M}_n(\tau_0)] < 0$ in probability and hence $\arg\max_{\tau\in[n_0,\tau_0]} \mathbb{M}_n(\tau) = \arg\max_{\tau\in[\tau_{n,p},\tau_0]} \mathbb{M}_n(\tau)$ holds in probability. Note that for any $\tau \in [n - n_0, \tau_{n,p}]$,

$$\mathbb{M}_n(\tau) - \mathbb{M}_n(\tau_0) = \mathbb{M}_n(\tau) - \mathbb{M}(\tau) - \mathbb{M}_n(\tau_0) + \mathbb{M}(\tau_0) - [\mathbb{M}(\tau_0) - \mathbb{M}(\tau)]. \qquad (A.40)$$

Given $\widehat{\nu}^{1,\tau}$ and $\widehat{\nu}^{\tau+1,n}$, we define an intermediate term

$$\mathbb{M}^*_n(\tau) := l(\{\theta^-_{\tau,k,\ell}, \eta^-_{\tau,k,\ell}\};\ \widehat{\nu}^{1,\tau}) + l(\{\theta^*_{\tau,k,\ell}, \eta^*_{\tau,k,\ell}\};\ \widehat{\nu}^{\tau+1,n}).$$

where

$$\theta^-_{\tau,k,\ell} = \frac{\sum_{(i,j)\in\widehat{S}^-_{\tau,k,\ell}} \frac{\theta_{1,\nu(i),\nu(j)}\eta_{1,\nu(i),\nu(j)}}{\theta_{1,\nu(i),\nu(j)}+\eta_{1,\nu(i),\nu(j)}}}{\sum_{(i,j)\in\widehat{S}^-_{\tau,k,\ell}} \frac{\eta_{1,\nu(i),\nu(j)}}{\theta_{1,\nu(i),\nu(j)}+\eta_{1,\nu(i),\nu(j)}}}, \quad \eta^-_{\tau,k,\ell} = \frac{\sum_{(i,j)\in\widehat{S}^-_{\tau,k,\ell}} \frac{\theta_{1,\nu(i),\nu(j)}\eta_{1,\nu(i),\nu(j)}}{\theta_{1,\nu(i),\nu(j)}+\eta_{1,\nu(i),\nu(j)}}}{\sum_{(i,j)\in\widehat{S}^-_{\tau,k,\ell}} \frac{\theta_{1,\nu(i),\nu(j)}}{\theta_{1,\nu(i),\nu(j)}+\eta_{1,\nu(i),\nu(j)}}},$$

and

$$\theta^*_{\tau,k,\ell} = \frac{\sum_{(i,j)\in\widehat{S}^+_{\tau,k,\ell}} \left[\frac{(\tau_0-\tau)\theta_{1,\nu(i),\nu(j)}\eta_{1,\nu(i),\nu(j)}}{\theta_{1,\nu(i),\nu(j)}+\eta_{1,\nu(i),\nu(j)}} + \frac{(n-\tau_0)\theta_{2,\nu(i),\nu(j)}\eta_{2,\nu(i),\nu(j)}}{\theta_{2,\nu(i),\nu(j)}+\eta_{2,\nu(i),\nu(j)}} \right]}{\sum_{(i,j)\in\widehat{S}^+_{\tau,k,\ell}} \left[\frac{(\tau_0-\tau)\eta_{1,\nu(i),\nu(j)}}{\theta_{1,\nu(i),\nu(j)}+\eta_{1,\nu(i),\nu(j)}} + \frac{(n-\tau_0)\eta_{2,\nu(i),\nu(j)}}{\theta_{2,\nu(i),\nu(j)}+\eta_{2,\nu(i),\nu(j)}} \right]},$$

$$\eta^*_{\tau,k,\ell} = \frac{\sum_{(i,j)\in\widehat{S}^+_{\tau,k,\ell}} \left[\frac{(\tau_0-\tau)\theta_{1,\nu(i),\nu(j)}\eta_{1,\nu(i),\nu(j)}}{\theta_{1,\nu(i),\nu(j)}+\eta_{1,\nu(i),\nu(j)}} + \frac{(n-\tau_0)\theta_{2,\nu(i),\nu(j)}\eta_{2,\nu(i),\nu(j)}}{\theta_{2,\nu(i),\nu(j)}+\eta_{2,\nu(i),\nu(j)}} \right]}{\sum_{(i,j)\in\widehat{S}^+_{\tau,k,\ell}} \left[\frac{(\tau_0-\tau)\theta_{1,\nu(i),\nu(j)}}{\theta_{1,\nu(i),\nu(j)}+\eta_{1,\nu(i),\nu(j)}} + \frac{(n-\tau_0)\theta_{2,\nu(i),\nu(j)}}{\theta_{2,\nu(i),\nu(j)}+\eta_{2,\nu(i),\nu(j)}} \right]}.$$

We have

$$\mathbb{M}_n(\tau) - \mathbb{M}(\tau) = \mathbb{M}_n(\tau) - E\mathbb{M}_n^*(\tau) + E\mathbb{M}_n^*(\tau) - \mathbb{M}(\tau).$$

Note that the expected log-likelihood $E\sum_{1\le i\le j\le p} g_{1,i,j}(\alpha_{1,i,j}, \beta_{1,i,j}, \tau)$ is maximized at $\alpha_{1,i,j} = \theta_{1,\nu(i),\nu(j)}, \beta_{1,i,j} = \eta_{1,\nu(i),\nu(j)}$, and $E\sum_{1\le i\le j\le p} g_{2,i,j}(\alpha_{2,i,j}, \beta_{2,i,j}, \tau)$ is maximized at $\alpha_{2,i,j} = \theta_{\tau,\nu(i),\nu(j)}$, $\beta_{2,i,j} = \eta_{\tau,\nu(i),\nu(j)}$, we have

$$E\mathbb{M}_n^*(\tau) - \mathbb{M}(\tau) \le 0.$$

On the other hand, notice that given $\widehat{\nu}$, $\{\theta^-_{\tau,k,\ell}, \eta^-_{\tau,k,\ell}\}$ is the maximizer of $El(\{\theta_{k,\ell}, \eta_{k,\ell}\};\ \widehat{\nu}^{1,\tau})$ and $\{\theta^*_{\tau,k,\ell}, \eta^*_{\tau,k,\ell}\}$ is the maximizer of $El(\{\theta_{k,\ell}, \eta_{k,\ell}\};\ \widehat{\nu}^{\tau+1,n})$. Similar to (A.31), there exists a large enough constant $C_7 > 0$ such that with probability greater than $1 - O((np)^{-B})$,

$$\sup_{\tau\in[n_0,\tau_{n,p}]} |\mathbb{M}_n(\tau) - E\mathbb{M}_n^*(\tau)| \le C_7 np^2 \left\{ \frac{\log(np)}{n} + \sqrt{\frac{\log(np)}{np^2}} \right\}.$$

Consequently we have, with probability greater than $1 - O((np)^{-B})$,

$$\sup_{\tau\in[n_0,\tau_{n,p}]} \big[\mathbb{M}_n(\tau) - \mathbb{M}(\tau)\big] \le C_7 np^2 \left\{ \frac{\log(np)}{n} + \sqrt{\frac{\log(np)}{np^2}} \right\}. \tag{A.41}$$

We remark that since the membership structure $\widehat{\nu}^{\tau+1,n}$ can be very different from the original ν, the $s_{\min}$ in (A.31) is simply replaced by the lower bound 1, and hence the upper bound in (A.41) is independent of $\widehat{\nu}^{1;\tau}$ and $\widehat{\nu}^{\tau+1,n}$.

Combining (A.40), (A.41), (A.25), (A.31) (with $\tau = \tau_0$), and choosing $\kappa > 0$ to be large enough, we have with probability greater than $1 - O((np)^{-B})$,

$$\begin{aligned} & \sup_{\tau\in[n_0,\tau_{n,p}]} \big[\mathbb{M}_n(\tau) - \mathbb{M}_n(\tau_0)\big] \\ \le\ & C_7 np^2 \left\{ \frac{\log(np)}{n} + \sqrt{\frac{\log(np)}{np^2}} \right\} + C_0 np^2 \left\{ \frac{\log(np)}{ns_{\min}^2} + \sqrt{\frac{\log(np)}{np^2}} \right\} \\ & -C_3(\tau_0 - \tau_{n,p})\big[\|\mathbf{W}_{1,1} - \mathbf{W}_{2,1}\|_F^2 + \|\mathbf{W}_{1,2} - \mathbf{W}_{2,2}\|_F^2\big] \\ <\ & 0. \end{aligned}$$

Consequently we have,

$$P\left(\arg\max_{\tau\in[n_0,\tau_0]} \big[\mathbb{M}_n(\tau) - \mathbb{M}_n(\tau_0)\big] = \arg\max_{\tau\in[\tau_{n,p},\tau_0]} \big[\mathbb{M}_n(\tau) - \mathbb{M}_n(\tau_0)\big] \right) \ge 1 - O((np)^{-B}). \tag{A.42}$$

(vi) Error bound for $\tau_0 - \widehat{\tau}$.

One of the key steps in the proof of (v) is to compare $M_n(\tau)$, the estimated log-likelihood evaluated under the MLEs at a searching time point τ, with $M(\tau)$, the maximized expected log-likelihood at time τ. The error between $M_n(\tau)$ and $M(\tau)$, which is of order $O\Bigg(np^2 \Big(\frac{\log(np)}{n} +$

$\sqrt{\frac{\log(np)}{np^2}}\Big)\Big)$ reflects the noise level. On the other hand, the signal is captured by $M(\tau_0)-M(\tau) = O(|\tau_0-\tau|p^2\Delta_F^2)$, i.e., the difference between the maximized expected log-likelihood evaluated at the true change-point τ_0 and the maximized expected log-likelihood evaluated at the searching time point τ. Consequently, when $|\tau_0-\tau|p^2\Delta_F^2 > \kappa\left[np^2\Big(\frac{\log(np)}{n}+\sqrt{\frac{\log(np)}{np^2}}\Big)\right]$ for some large enough constant $\kappa > 0$, we are able to claim that $|\tau_0-\widehat{\tau}| \le |\tau_0-\tau| = O_p\Big(n\Delta_F^{-2}\Big[\frac{\log(np)}{n}+\sqrt{\frac{\log(np)}{np^2}}\Big]\Big)$. By further deriving the estimation errors for any τ in the neighborhood of τ_0 with radius $O\Big(\Delta_F^{-2}\Big[\frac{\log(np)}{n}+\sqrt{\frac{\log(np)}{np^2}}\Big]\Big)$, we obtained a better bound based on Markov's inequality (see (A.43) below).

From (A.42) we have for any $0<\epsilon\le\tau_0-\tau_{n,p}$,

$$P(\tau_0-\widehat{\tau}>\epsilon)\le P\Big(\sup_{\tau\in[\tau_{n,p},\tau_0-\epsilon]}\mathbb{M}_n(\tau)-\mathbb{M}_n(\tau_0)\ge 0\Big)+O((np)^{-B}).$$

Note that from (i) and (iv) we have

$$\begin{aligned}
& P\Big(\sup_{\tau\in[\tau_{n,p},\tau_0-\epsilon]}\mathbb{M}_n(\tau)-\mathbb{M}_n(\tau_0)\ge 0\Big) \qquad\qquad (A.43)\\
\le\;& P\Big(\sup_{\tau\in[\tau_{n,p},\tau_0-\epsilon]}\big[(\mathbb{M}_n(\tau)-\mathbb{M}(\tau)-\mathbb{M}_n(\tau_0)+\mathbb{M}(\tau_0))-(\mathbb{M}(\tau_0)-\mathbb{M}(\tau))\big]\ge 0\Big)\\
\le\;& P\Big(\sup_{\tau\in[\tau_{n,p},\tau_0-\epsilon]}|\mathbb{M}_n(\tau)-\mathbb{M}(\tau)-\mathbb{M}_n(\tau_0)+\mathbb{M}(\tau_0)|\ge C_3\epsilon p^2\Delta_F^2\Big)\\
\le\;& \frac{E\sup_{\tau\in[\tau_{n,p},\tau_0-\epsilon]}\big|\mathbb{M}_n(\tau)-\mathbb{M}(\tau)-\mathbb{M}_n(\tau_0)+\mathbb{M}(\tau_0)\big|}{C_3\epsilon p^2\Delta_F^2}\\
\le\;& \frac{2C_0p^2\sqrt{\frac{(\tau_0-\tau_{n,p})\log(np)}{s_{\min}^2}}}{C_3\epsilon p^2\Delta_F^2}.
\end{aligned}$$

We thus conclude that $\tau_0-\widehat{\tau}=O_p\left(\Delta_F^{-2}\sqrt{\frac{(\tau_0-\tau_{n,p})\log(np)}{s_{\min}^2}}\right)$. By the definition of $\tau_{n,p}$ and Condition C5 we have,

$$\Delta_F^{-2}\sqrt{\frac{(\tau_0-\tau_{n,p})\log(np)}{s_{\min}^2}} \;=\; O\left(\frac{\tau_0-\tau_{n,p}}{\Delta_F}\sqrt{\frac{\log(np)}{ns_{\min}^2}}\left[\frac{\log(np)}{n}+\sqrt{\frac{\log(np)}{np^2}}\right]^{-1/2}\right).$$

Consequently, we conclude that

$$\tau_0-\widehat{\tau}=O_p\left((\tau_0-\tau_{n,p})\min\left\{1,\frac{\min\left\{1,(n^{-1}p^2\log(np))^{\frac{1}{4}}\right\}}{\Delta_F s_{\min}}\right\}\right).$$

A.9.2 Change point estimation with $\nu^{1,\tau_0}\neq\nu^{\tau_0+1,n}$.

We modify steps (i)-(v) to the case where $\nu^{1,\tau_0}\neq\nu^{\tau_0+1,n}$.

With some abuse of notations, we put $\mathbf{W}_{1,1} = (\alpha_{1,i,j})_{p\times p}$ with $\alpha_{1,i,j} = \theta_{1,\nu^{1,\tau_0}(i),\nu^{1,\tau_0}(j)}$, $\mathbf{W}_{1,2} = (1-\beta_{1,i,j})_{p\times p}$ with $\beta_{1,i,j} = \eta_{1,\nu^{1,\tau_0}(i),\nu^{1,\tau_0}(j)}$, $\mathbf{W}_{2,1} = (\alpha_{2,i,j})_{p\times p}$ with $\alpha_{2,i,j} = \theta_{1,\nu^{\tau_0+1,n}(i),\nu^{\tau_0+1,n}(j)}$, and $\mathbf{W}_{2,2} = (1-\beta_{2,i,j})_{p\times p}$ with $\beta_{2,i,j} = \eta_{1,\nu^{\tau_0+1,n}(i),\nu^{\tau_0+1,n}(j)}$. Similar to previous proofs we define

$$\begin{aligned}
\mathbb{M}_n(\tau) &:= \sum_{1\le i\le j\le p} g_{1,i,j}(\widehat{\alpha}^{\tau}_{1,i,j}, \widehat{\beta}^{\tau}_{1,i,j}, \tau) + \sum_{1\le i\le j\le p} g_{2,i,j}(\widehat{\alpha}^{\tau}_{2,i,j}, \widehat{\beta}^{\tau}_{2,i,j}, \tau),\\
\mathbb{M}(\tau) &:= E\sum_{1\le i\le j\le p} g_{1,i,j}(\alpha_{1,i,j}, \beta_{1,i,j}, \tau) + E\sum_{1\le i\le j\le p} g_{2,i,j}(\alpha^{\tau}_{2,i,j}, \beta^{\tau}_{2,i,j}, \tau),
\end{aligned}$$

where

$$\begin{aligned}
\alpha^{\tau}_{2,i,j} &= \frac{\frac{\tau_0-\tau}{n-\tau}\frac{\alpha_{1,i,j}\beta_{1,i,j}}{\alpha_{1,i,j}+\beta_{1,i,j}} + \frac{n-\tau_0}{n-\tau}\frac{\alpha_{2,i,j}\beta_{2,i,j}}{\alpha_{2,i,j}+\beta_{2,i,j}}}{\frac{\tau_0-\tau}{n-\tau}\frac{\beta_{1,i,j}}{\alpha_{1,i,j}+\beta_{1,i,j}} + \frac{n-\tau_0}{n-\tau}\frac{\beta_{2,i,j}}{\alpha_{2,i,j}+\beta_{2,i,j}}},\\
\beta^{\tau}_{2,i,j} &= \frac{\frac{\tau_0-\tau}{n-\tau}\frac{\alpha_{1,i,j}\beta_{1,i,j}}{\alpha_{1,i,j}+\beta_{1,i,j}} + \frac{n-\tau_0}{n-\tau}\frac{\alpha_{2,i,j}\beta_{2,i,j}}{\alpha_{2,i,j}+\beta_{2,i,j}}}{\frac{\tau_0-\tau}{n-\tau}\frac{\alpha_{1,i,j}}{\alpha_{1,i,j}+\beta_{1,i,j}} + \frac{n-\tau_0}{n-\tau}\frac{\alpha_{2,i,j}}{\alpha_{2,i,j}+\beta_{2,i,j}}},
\end{aligned}$$

and

$$\begin{aligned}
\widehat{\alpha}^{\tau}_{1,i,j} &= \widehat{\theta}^{\tau}_{1,\widehat{\nu}^{1,\tau}(i),\widehat{\nu}^{1,\tau}(j)}, \quad \widehat{\beta}^{\tau}_{1,i,j} = \widehat{\eta}^{\tau}_{1,\widehat{\nu}^{1,\tau}(i),\widehat{\nu}^{1,\tau}(j)},\\
\widehat{\alpha}^{\tau}_{2,i,j} &= \widehat{\theta}^{\tau}_{2,\widehat{\nu}^{\tau+1,n}(i),\widehat{\nu}^{\tau+1,n}(j)}, \quad \widehat{\beta}^{\tau}_{2,i,j} = \widehat{\eta}^{2}_{\tau,\widehat{\nu}^{\tau+1,n}(i),\widehat{\nu}^{\tau+1,n}(j)}.
\end{aligned}$$

Note that the definition of $M(\tau)$ here is now slightly different from the previous definition in that the $\alpha^{\tau}_{2,i,j}$ and $\beta^{\tau}_{2,i,j}$ will generally be different from $\theta^{\tau}_{2,\nu^{\tau_0+1,n}(i),\nu^{\tau_0+1,n}(j)}$ and $\eta^{\tau}_{2,\nu^{\tau_0+1,n}(i),\nu^{\tau_0+1,n}(j)}$, unless $\nu^{1,\tau_0} = \nu^{\tau_0+1,n}$. We first of all point out the main difference we are facing in the case where $\nu^{1,\tau_0} \neq \nu^{\tau_0+1,n}$. Consider a detection time $\tau \in [\tau_{n,p}, \tau_0]$. In the case where $\widehat{\nu}^{1,\tau} = \widehat{\nu}^{\tau+1,n} = \nu$, we have $\alpha^{\tau}_{2,i,j} = \theta_{\tau,k,\ell}$ for all $(i,j) \in S_{k,\ell}$, and we have $|\widehat{\theta}^{\tau}_{2,k,\ell} - \theta^{\tau}_{2,k,\ell}| = O_p\Big(\sqrt{\frac{\log(np)}{ns^2_{\min}}}\Big)$ for all $1 \le k \le \ell \le q$, or equivalently, $|\widehat{\alpha}^{\tau}_{2,i,j} - \theta^{\tau}_{2,\nu(i),\nu(j)}| = O_p\Big(\sqrt{\frac{\log(np)}{ns^2_{\min}}}\Big)$ for all $1 \le i \le j \le p$. However, when $\widehat{\nu}^{1,\tau} = \nu^{1,\tau_0}$ $\widehat{\nu}^{\tau+1,n} = \nu^{\tau_0+1,n}$ but $\nu^{1,\tau_0} \neq \nu^{\tau_0+1,n}$, the order of the estimation error becomes $O_p\Big(\sqrt{\frac{\log(np)}{ns^2_{\min}}} + \frac{\tau_0-\tau}{n}\Big)$. Here $\frac{\tau_0-\tau}{n}$ is a bias terms brought by the fact that $\widehat{\nu}^{1,\tau} \neq \widehat{\nu}^{\tau+1,n}$. The main issue is that the the following terms from the definition of $\widehat{\theta}^{\tau}_{2,k,\ell}$:

$$\sum_{(i,j)\in \widehat{S}^{\tau}_{2,k,\ell}} \sum_{t=\tau+1}^{\tau_0} X^t_{i,j}(1-X^{t-1}_{i,j}), \qquad \sum_{(i,j)\in \widehat{S}^{\tau}_{2,k,\ell}} \sum_{t=\tau+1}^{\tau_0} (1-X^{t-1}_{i,j}),$$

are no longer unbiased estimators (subject to a normalization) of the following corresponding terms in the definition of $\theta^{\tau}_{2,k,\ell}$:

$$\frac{\theta_{1,k,\ell}\eta_{1,k,\ell}}{\theta_{1,k,\ell}+\eta_{1,k,\ell}}, \quad \frac{\theta_{1,k,\ell},}{\theta_{1,k,\ell}+\eta_{1,k,\ell}}.$$

The proof of (i) does not involve any parameter estimators and hence can be established similarly.

For (ii), note that $|\widehat{\alpha}^{\tau}_{2,i,j}-\alpha_{2,i,j}|\le|\widehat{\alpha}^{\tau}_{2,i,j}-\alpha^{\tau}_{2,i,j}|+O\big(\frac{\tau_0-\tau}{n}\big)$ holds for all $1\le i<j\le p$, where the $O\big(\frac{\tau_0-\tau}{n}\big)$ is independent of i,j. This implies that when estimating the $\alpha_{2,i,j}$, we have introduced a bias term $O\big(\frac{\tau_0-\tau}{n}\big)$ by including the $\tau_0-\tau$ samples before the change point. From the proofs of Lemma 6, and Condition C4, we conclude that (ii) hold for $\widehat{\nu}^{\tau+1,n}$.

For (iii), replacing the order of the error bound for $\widehat{\theta}^{+}_{\tau,k,\ell}$ and $\widehat{\theta}^{+}_{\tau,k,\ell}$ from $\sqrt{\frac{\log(np)}{ns^2_{\min}}}$ to $\sqrt{\frac{\log(np)}{ns^2_{\min}}}+\frac{\tau_0-\tau}{n}$, we have there exists a large enough constant $C_0>0$ such that

$$\begin{aligned}\sup_{\tau\in[\tau_{n,p},\tau_0]}|\mathbb{M}_n(\tau)-\mathbb{M}(\tau)| &\le C_0np^2\left\{\frac{\log(np)}{ns^2_{\min}}+\sqrt{\frac{\log(np)}{np^2}}+\frac{(\tau_0-\tau_{n,p})^2}{n^2}\right\}\\ &= O\left(np^2\left\{\sqrt{\frac{\log(np)}{ns^2_{\min}}}+\frac{(\tau_0-\tau_{n,p})^2}{n^2}\right\}\right).\end{aligned}$$

For (iv), the error bounds related to $g_{1,i,j}(\cdot,\cdot;\cdot)$ remain unchanged. Note that the decomposition (A.34) still holds with $\theta^{\tau}_{2,k,\ell},\eta^{\tau}_{2,k,\ell}$ replaced be $\alpha^{\tau}_{2,i,j},\beta^{\tau}_{2,i,j}$ and $\widehat{\theta}^{\tau}_{2,k,\ell},\widehat{\eta}^{\tau}_{2,k,\ell}$ replaced be $\widehat{\alpha}^{\tau}_{2,i,j},\widehat{\beta}^{\tau}_{2,i,j}$. The bound for (A.35) still holds owing to the fact that $|\alpha^{\tau}_{2,i,j}-\alpha_{2,i,j}|=O\Big(\frac{\tau_0-\tau}{n}\Big)$ and $|\beta^{\tau}_{2,i,j}-\beta_{2,i,j}|=O\Big(\frac{\tau_0-\tau}{n}\Big)$. The bound for (A.36) would become $O\left(p^2\sqrt{\frac{(\tau_0-\tau)\log(np)}{s^2_{\min}}}+\frac{(\tau_0-\tau_{n,p})^2}{n^2}\right)$. Notice that similar to (A.37), we have with probability larger than $1-O((np)^{-B})$,

$$\begin{aligned}\sup_{1\le i\le j\le p}\left|\log\frac{\widehat{\alpha}^{\tau}_{2,i,j}}{\alpha^{\tau}_{2,i,j}}-\log\frac{\widehat{\alpha}^{\tau_0}_{2,k,\ell}}{\alpha_{2,k,\ell}}\right| &= O\left(\sqrt{\frac{\tau_0-\tau}{n}}\sqrt{\frac{\log(np)}{ns^2_{\min}}}+\frac{\tau_0-\tau}{n}\right),\\ \sup_{1\le i\le j\le p}\left|\log\frac{\widehat{\beta}^{\tau}_{2,i,j}}{\beta^{\tau}_{2,i,j}}-\log\frac{\widehat{\beta}^{\tau_0}_{2,k,\ell}}{\beta 2,k,\ell}\right| &= O\left(\sqrt{\frac{\tau_0-\tau}{n}}\sqrt{\frac{\log(np)}{ns^2_{\min}}}+\frac{\tau_0-\tau}{n}\right).\end{aligned}$$

Consequently, we have

$$E\sup_{\tau\in[\tau_{n,p},\tau_0]}|\mathbb{M}_n(\tau)-\mathbb{M}(\tau)-\mathbb{M}_n(\tau_0)+\mathbb{M}(\tau_0)|\le C_0p^2\left\{\sqrt{\frac{(\tau_0-\tau_{n,p})\log(np)}{s^2_{\min}}}+(\tau_0-\tau_{n,p})\right\}.$$

By noticing that $\{\alpha_{1,i,j},\beta_{1,i,j},\alpha^{\tau}_{2,i,j},\beta^{\tau}_{2,i,j}\}$ is the maximizer of $\mathbb{M}(\tau)$, we conclude that (v) also holds. Consequently, for (vi), we have

$$P\left(\sup_{\tau\in[\tau_{n,p},\tau_0-\epsilon]}\mathbb{M}_n(\tau)-\mathbb{M}_n(\tau_0)\ge 0\right)\le\frac{C_0p^2\sqrt{\frac{(\tau_0-\tau_{n,p})\log(np)}{s^2_{\min}}}+C_0p^2(\tau_0-\tau_{n,p})}{C_3\epsilon p^2\Delta_F^2}.$$

Consequently, we conclude that

$$\tau_0-\widehat{\tau}=O_p\left((\tau_0-\tau_{n,p})\min\left\{1,\frac{\min\left\{1,(n^{-1}p^2\log(np))^{\frac{1}{4}}\right\}}{\Delta_F s_{\min}}+\frac{1}{\Delta_F^2}\right\}\right).$$

A.9.3 Proofs of (A.32) and (A.37) when $\widehat{\nu} = \nu$

For (A.32), note that

$$\left|\widehat{\theta}_{1,k,\ell}^{\tau} - \widehat{\theta}_{1,k,\ell}^{\tau_0}\right| = \left| \frac{\sum_{(i,j)\in S_{k,\ell}} \sum_{t=1}^{\tau} X_{i,j}^t (1 - X_{i,j}^{t-1})}{\sum_{(i,j)\in S_{k,\ell}} \sum_{t=1}^{\tau} (1 - X_{i,j}^{t-1})} - \frac{\sum_{(i,j)\in S_{k,\ell}} \sum_{t=1}^{\tau_0} X_{i,j}^t (1 - X_{i,j}^{t-1})}{\sum_{(i,j)\in S_{k,\ell}} \sum_{t=1}^{\tau_0} (1 - X_{i,j}^{t-1})} \right| \quad \text{(A.44)}$$

Similar to Lemma 3, we can show that for any constant $B > 0$, there exists a large enough constant B_1 such that with probability larger than $1 - O((np)^{-(B+2)})$,

$$\left| \frac{1}{\tau n_{k,\ell}} \sum_{(i,j)\in S_{k,\ell}} \sum_{t=1}^{\tau} (1 - X_{i,j}^{t-1}) - \frac{\eta_{1,k,\ell}}{\theta_{1,k,\ell} + \eta_{1,k,\ell}} \right| \le B_1 \sqrt{\frac{\log(np)}{\tau n_{k,\ell}}},$$

$$\left| \frac{1}{\tau n_{k,\ell}} \sum_{(i,j)\in S_{k,\ell}} \sum_{t=1}^{\tau} X_{i,j}^t (1 - X_{i,j}^{t-1}) - \frac{\eta_{1,k,\ell}}{\theta_{1,k,\ell} + \eta_{1,k,\ell}} \right| \le B_1 \sqrt{\frac{\log(np)}{\tau n_{k,\ell}}},$$

and

$$\begin{aligned} \frac{1}{\tau(\tau_0 - \tau) n_{k,\ell}^2} \Bigg| & \Bigg[\sum_{(i,j)\in S_{k,\ell}} \sum_{t=1}^{\tau} X_{i,j}^t (1 - X_{i,j}^{t-1}) \Bigg] \Bigg[\sum_{(i,j)\in S_{k,\ell}} \sum_{t=\tau+1}^{\tau_0} (1 - X_{i,j}^{t-1}) \Bigg] \\ - & \Bigg[\sum_{(i,j)\in S_{k,\ell}} \sum_{t=\tau+1}^{\tau_0} X_{i,j}^t (1 - X_{i,j}^{t-1}) \Bigg] \Bigg[\sum_{(i,j)\in S_{k,\ell}} \sum_{t=1}^{\tau} (1 - X_{i,j}^{t-1}) \Bigg] \Bigg| \le B_1 \sqrt{\frac{\log(np)}{(\tau_0 - \tau) n_{k,\ell}}}. \end{aligned}$$

Plug these into (A.44) we have with probability larger than $1 - O((np)^{-(B+2)})$,

$$\left|\widehat{\theta}_{1,k,\ell}^{\tau} - \widehat{\theta}_{1,k,\ell}^{\tau_0}\right| \le \frac{c_0 \tau (\tau_0 - \tau) n_{k,\ell}^2}{\tau_0 \tau n_{k,\ell}^2} \sqrt{\frac{\log(np)}{(\tau_0 - \tau) n_{k,\ell}}} \le \frac{c_0 \sqrt{\tau_0 - \tau}}{\tau_0} \sqrt{\frac{\log(np)}{n_{k,\ell}}},$$

for some constant $c_0 > 0$. Since $\tau_0 \simeq O(n)$, and $n_{k,\ell} \ge s_{\min}^2$, we conclude that there exists a constant $c_1 > 0$ such that with probability larger than $1 - O(np)^{-B})$,

$$\sup_{1 \le k \le \ell \le q} \left|\widehat{\theta}_{1,k,\ell}^{\tau} - \widehat{\theta}_{1,k,\ell}^{\tau_0}\right| \le c_1 \sqrt{\frac{\tau_0 - \tau}{\tau_0}} \sqrt{\frac{\log(np)}{n s_{\min}^2}}.$$

For (A.37), note that

$$\begin{aligned} & \log \frac{\widehat{\theta}_{2,k,\ell}^{\tau}}{\theta_{2,k,\ell}^{\tau}} - \log \frac{\widehat{\theta}_{2,k,\ell}^{\tau_0}}{\theta_{2,k,\ell}} \\ = \; & \log \frac{\frac{1}{n_{k,\ell}(n-\tau)} \sum_{(i,j)\in S_{k,\ell}} \sum_{t=\tau+1}^{n} X_{i,j}^t (1 - X_{i,j}^{t-1})}{\frac{\tau_0 - \tau}{n - \tau} \frac{\theta_{1,k,\ell}\eta_{1,k,\ell}}{\theta_{1,k,\ell} + \eta_{1,k,\ell}} + \frac{n - \tau_0}{n - \tau} \frac{\eta_{2,k,\ell}\eta_{2,k,\ell}}{\theta_{2,k,\ell} + \eta_{2,k,\ell}}} - \log \frac{\sum_{(i,j)\in S_{k,\ell}} \sum_{t=\tau_0+1}^{n} X_{i,j}^t (1 - X_{i,j}^{t-1})}{n_{k,\ell}(n - \tau_0) \cdot \frac{\eta_{2,k,\ell}\eta_{2,k,\ell}}{\theta_{2,k,\ell} + \eta_{2,k,\ell}}} \\ & - \log \frac{\frac{1}{n_{k,\ell}(n-\tau)} \sum_{(i,j)\in S_{k,\ell}} \sum_{t=\tau+1}^{n} (1 - X_{i,j}^{t-1})}{\frac{\tau_0 - \tau}{n - \tau} \frac{\eta_{1,k,\ell}}{\theta_{1,k,\ell} + \eta_{1,k,\ell}} + \frac{n - \tau_0}{n - \tau} \frac{\eta_{2,k,\ell}}{\theta_{2,k,\ell} + \eta_{2,k,\ell}}} + \log \frac{\sum_{(i,j)\in S_{k,\ell}} \sum_{t=\tau_0+1}^{n} (1 - X_{i,j}^{t-1})}{n_{k,\ell}(n - \tau_0) \cdot \frac{\eta_{2,k,\ell}}{\theta_{2,k,\ell} + \eta_{2,k,\ell}}}. \end{aligned}$$

It suffices to establish a bound for

$$\Delta_{\tau_0,\tau} := \frac{\frac{1}{n_{k,\ell}(n-\tau)} \sum_{(i,j)\in S_{k,\ell}} \sum_{t=\tau+1}^{n} X_{i,j}^t (1 - X_{i,j}^{t-1})}{\frac{\tau_0 - \tau}{n - \tau} \frac{\theta_{1,k,\ell}\eta_{1,k,\ell}}{\theta_{1,k,\ell} + \eta_{1,k,\ell}} + \frac{n - \tau_0}{n - \tau} \frac{\eta_{2,k,\ell}\eta_{2,k,\ell}}{\theta_{2,k,\ell} + \eta_{2,k,\ell}}} - \frac{\sum_{(i,j)\in S_{k,\ell}} \sum_{t=\tau_0+1}^{n} X_{i,j}^t (1 - X_{i,j}^{t-1})}{n_{k,\ell}(n - \tau_0) \cdot \frac{\eta_{2,k,\ell}\eta_{2,k,\ell}}{\theta_{2,k,\ell} + \eta_{2,k,\ell}}}.$$

Note that for any $B > 0$, there exists a large enough constant B_2 such that with probability greater than $1 - O((np)^{-(B+2)})$,

$$
\begin{aligned}
|\Delta_{\tau_0,\tau}| \leq & \left| \frac{\frac{1}{n_{k,\ell}(n-\tau)} \sum_{(i,j)\in S_{k,\ell}} \sum_{t=\tau+1}^{\tau_0} \left[X_{i,j}^t (1 - X_{i,j}^{t-1}) - \frac{\theta_{1,k,\ell}\eta_{1,k,\ell}}{\theta_{1,k,\ell}+\eta_{1,k,\ell}} \right]}{\frac{\tau_0-\tau}{n-\tau} \frac{\theta_{1,k,\ell}\eta_{1,k,\ell}}{\theta_{1,k,\ell}+\eta_{1,k,\ell}} + \frac{n-\tau_0}{n-\tau} \frac{\eta_{2,k,\ell}\eta_{2,k,\ell}}{\theta_{2,k,\ell}+\eta_{2,k,\ell}}} \right| \\
& + \left| \left(\frac{\frac{1}{n_{k,\ell}(n-\tau)}}{\frac{\tau_0-\tau}{n-\tau} \frac{\theta_{1,k,\ell}\eta_{1,k,\ell}}{\theta_{1,k,\ell}+\eta_{1,k,\ell}} + \frac{n-\tau_0}{n-\tau} \frac{\eta_{2,k,\ell}\eta_{2,k,\ell}}{\theta_{2,k,\ell}+\eta_{2,k,\ell}}} - \frac{\frac{1}{n_{k,\ell}(n-\tau_0)}}{\frac{\eta_{2,k,\ell}\eta_{2,k,\ell}}{\theta_{2,k,\ell}+\eta_{2,k,\ell}}} \right) \sum_{(i,j)\in S_{k,\ell}} \sum_{t=\tau_0+1}^{n} \left[X_{i,j}^t (1 - X_{i,j}^{t-1}) \right. \right. \\
& \left. \left. - \frac{\theta_{2,k,\ell}\eta_{2,k,\ell}}{\theta_{2,k,\ell}+\eta_{2,k,\ell}} \right] \right| \\
\leq & B_2 \frac{\tau_0-\tau}{n-\tau} \sqrt{\frac{\log(np)}{(\tau_0-\tau)n_{k,\ell}}} + B_2 \frac{\tau_0-\tau}{n-\tau} \sqrt{\frac{\log(np)}{(n-\tau_0)n_{k,\ell}}}.
\end{aligned}
$$

(A.37) then follows by noticing that $\frac{\tau_0-\tau}{n-\tau} \sqrt{\frac{\log(np)}{(n-\tau_0)n_{k,\ell}}} = o\Big(\frac{\tau_0-\tau}{n-\tau} \sqrt{\frac{\log(np)}{(\tau_0-\tau)n_{k,\ell}}} \Big)$.