Supplement to "Community detection with dependent connectivity"

Notation

In the following, we denote the membership of node as random variable $z_i, i = 1, \cdots, N$. Then $\boldsymbol{Z} = \{z_1, z_2, \cdots, z_N\}$. Accordingly, we define the true membership of nodes as $z_i^* \in \{1, 2, \cdots, K\}$, $i = 1, \cdots, N$ and $z^* = \{z_1^*, z_2^*, \cdots, z_N^*\}$. We denote $P^*(\cdot) = P(\cdot|\boldsymbol{Z} = z^*)$ as the conditional probability of observed networks given the true nodes' membership z^*. The number of misclassified nodes is denoted as r such that $\|z - z^*\|_0 = r$ for $z \neq z^*$. Define the t-th sample network as $\boldsymbol{Y}^t = (Y_{ij}^t)_{N\times N}$ and t-th sample network standardized by $\hat{\mu}_{aa}$ as $\hat{\boldsymbol{Y}}^{t,a} = (\hat{Y}_{ij}^{t,a})_{N\times N}$ where $\hat{Y}_{ij}^{t,a} = \frac{Y_{ij}^t - \hat{\mu}_{aa}}{\sqrt{\hat{\mu}_{aa}(1-\hat{\mu}_{aa})}}, a = 1, \cdots, K,\ t = 1, \cdots, M$. We further define the s-th column of $\hat{\boldsymbol{Y}}^{t,a}$ as $\hat{Y}_{\cdot s}^{t,a}$. ρ_{ijuv} denotes pairwise correlation between two edges Y_{ij}^t and Y_{uv}^t. Given the empirical estimation $\hat{\rho}_{ijuv} = \rho_{ijuv}$ almost sure as M increase, we assume $\{\rho_{ijuv}\}$ are known in the following proofs.

Denote $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_N)$ as the estimated probability of nodes' memberships. Specifically, let $\alpha_i = (\alpha_{i1}, \cdots, \alpha_{iK})_{1\times K}$ be the probability of nodes i belonging to each community where $\sum_{q=1}^K \alpha_{iq} = 1,\ i = 1, \cdots, N$. For simplicity of notation, if the subscripts indicate the community then $\alpha_q = (\alpha_{1q}, \cdots, \alpha_{Nq})_{1\times N}$ represents the probability of each node belonging to community q, where $q = 1, \cdots, K$. Similarly, $z_q^* = \{z_{1q}^*, z_{2q}^*, \cdots, z_{Nq}^*\}$ is a binary vector indicating nodes whose true membership belongs to community $q, q = 1, \cdots, K$. Let $vec(\cdot)$ stand for the operation of vectorizing a matrix into a column.

The following lemma is introduced as the technical steps in the proofs of Theorem **??**, Theorem **??** and Theorem **??**. The proofs of Lemma 1 is provided in the supplemental material.

lemma 1.1. *Consider function* $f_1(x) = \sqrt{\Big\{x \log \frac{\mu_{z_i z_j}}{\mu_{z_i^* z_j^*}} + (1-x)\log \frac{1-\mu_{z_i z_j}}{1-\mu_{z_i^* z_j^*}}\Big\}_+}$ *and denote*

$$X_t^+ = \{f_1(Y_{12}^t), f_1(Y_{13}^t), \cdots, f_1(Y_{N-1,N}^t)\}$$

where $\{Y_{ij}^t\}_{N\times N}$ *are generated through the stochastic block model in section 3.1 and satisfy condition C1, C2 and C3. Define the covariance matrix of* X_t^+ *as* Σ_1. *Then* X_t^+ *is a subgaussian vector, i.e.,*

$$L = inf\{\alpha \geq 0 : E(\exp(\langle z, X_t^+ - E(X_t^+)\rangle)) \leq \exp\{\alpha^2\langle \Sigma_1 z, \ \ z\rangle\}/2, z \in R^{N(N-1)/2}\} \leq C$$

for some positive constant C.

Proof: recall that X_t^+ is a binary vector. For any random vector z such that $dim(z) = dim(X_t^+)$, consider random vectors $\varepsilon = \Sigma_1^{1/2} z, U_t = \Sigma_1^{-1/2}\{X_t^+ - E(X_t^+)\}$. Therefore,

$$\mathrm{var}(U_t) = \Sigma_1^{-1/2}\Sigma_1\Sigma_1^{-1/2} = I.$$

Given each element in U_t is bounded such that $|(U_t)_i| \leq C_1$ and $E((U_t)_i) = 0$,$1 \leq i \leq \frac{n(n-1)}{2}$, we have

$$\begin{aligned}
&E\{\exp\Big(\langle z, X_t^+ - E(X_t^+)\rangle\Big)\} \\
=&E\Big\{\exp\Big(\langle \Sigma_1^{1/2} z, \Sigma_1^{-1/2}(X_t^+ - E(X_t^+))\rangle\Big)\Big\} = E\Big\{\exp\Big(\langle \epsilon, U_t\rangle\Big)\Big\} \\
=& E\{\prod_{i=1} \exp\Big(\epsilon_i (U_t)_i\Big)\} = E\{E(E(E(V_1)V_2)V_3)\cdots V_{n(n-1)/2}\},
\end{aligned}$$

where

$$\begin{aligned}
V_1 &= E\{\exp\Big(\epsilon_1 (U_t)_1\Big) | (U_t)_2, \cdots, (U_t)_{n(n-1)/2}\}, \\
V_2 &= E\{\exp\Big(\epsilon_2 (U_t)_2\Big) | (U_t)_3, \cdots, (U_t)_{n(n-1)/2}\}, \\
&\vdots \\
V_{n(n-1)/2} &= E\{\exp\Big(\epsilon_{n(n-1)/2} (U_t)_{n(n-1)/2}\Big)\}.
\end{aligned}$$

According to the Hoeffding's lemma, we have

$$V_i \leq \exp\{\frac{\epsilon_i^2 C_1^2}{2}\}, \ i = 1, \cdots, n(n-1)/2.$$

Therefore,

$$E\{\exp\Big(\langle z, X_t^+ - E(X_t^+)\rangle\Big)\} \leq \prod_{i=1} \exp\{\frac{\epsilon_i^2 C_1^2}{2}\} = \exp\{\frac{C_1^2}{2}\langle \epsilon, \epsilon\rangle\}$$
$$= \exp\{\frac{C_1^2}{2}\langle \Sigma_1 z, z\rangle\}.$$

Therefore, X_t^+ is a subgaussian random vector. In addition, denote L as subgaussian norm of X_t^+ such that

$$L = inf\{\alpha \geq 0 : E(\exp(\langle z, X_t^+ - E(X_t^+)\rangle)) \leq \exp\{\alpha^2\langle \Sigma_1 z, z\rangle\}/2\}.$$

Then we have $L \leq \frac{C_1^2}{2}$.

Proof of Theorem 5.1

Given the independent model (4) defined in Section 2, we can simplify the likelihood ratio between a random membership z and the true membership z^* as

$$\log \frac{P_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z})}{P_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z^*})} = \frac{1}{M}\sum_{t=1}^{M}\sum_{i<j}\Big\{Y_{ij}^t \log\frac{\mu_{z_i z_j}}{\mu_{z_i^* z_j^*}} + (1-Y_{ij}^t)\log\frac{1-\mu_{z_i z_j}}{1-\mu_{z_i^* z_j^*}}\Big\}. \tag{1}$$

We define two transformation functions $f_1(x)$ and $f_2(x)$ as:

$$f_1(x) = \sqrt{\Big\{x\log\frac{\mu_{z_i z_j}}{\mu_{z_i^* z_j^*}} + (1-x)\log\frac{1-\mu_{z_i z_j}}{1-\mu_{z_i^* z_j^*}}\Big\}_+},$$
$$f_2(x) = \sqrt{\Big\{x\log\frac{\mu_{z_i z_j}}{\mu_{z_i^* z_j^*}} + (1-x)\log\frac{1-\mu_{z_i z_j}}{1-\mu_{z_i^* z_j^*}}\Big\}_-}.$$

where $\{\}_+$ and $\{\}_-$ are positive part and negative part of a random variable. The previous summation can be decomposed as positive part and negative part:

$$\log \frac{P_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z})}{P_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z^*})} = \frac{1}{M}\sum_{t=1}^{M}\sum_{i<j}\{f_1^2(Y_{ij}^t) - f_2^2(Y_{ij}^t)\}.$$

Define the vectorized edges in the t th sample network as:

$$X_t^+ = \{f_1(Y_{12}^t), f_1(Y_{13}^t), \cdots, f_1(Y_{N-1,N}^t)\}, X_t^- = \{f_2(Y_{12}^t), f_2(Y_{13}^t), \cdots, f_2(Y_{N-1,N}^t)\}. \tag{2}$$

Note that each element in X_t^+ or X_t^- is a bounded binary random variable. In addition, as $f_1(Y_{ij}^t)$ or $f_2(Y_{ij}^t)$ only rescale Y_{ij}^t then they preserve the within-community correlation among Y_{ij}^t. Then we consider the following quadratic forms

$$Q_1 = \sum_{t=1}^{M} \langle X_t^+, X_t^+ \rangle, Q_2 = \sum_{t=1}^{M} \langle X_t^-, X_t^- \rangle.$$

such that

$$\log \frac{P_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z})}{P_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z}^*)} = \frac{1}{M}(Q_1 - Q_2) \text{ and } E(\log \frac{P_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z})}{P_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z}^*)}) = \frac{1}{M}(EQ_1 - EQ_2).$$

Denote the centralized version quadratic forms Q_1 and Q_2 as $\mathcal{Q}_1$ and $\mathcal{Q}_2$ such that

$$\mathcal{Q}_1 = \sum_{t=1}^{M} \langle X_t^+ - E(X_t^+), X_t^+ - E(X_t^+) \rangle, \mathcal{Q}_2 = \sum_{t=1}^{M} \langle X_t^- - E(X_t^-), X_t^- - E(X_t^-) \rangle.$$

Denote the following quadratic difference as:

$$\Delta(Q_1, \mathcal{Q}_1) := (Q_1 - E(Q_1)) - (\mathcal{Q}_1 - E(\mathcal{Q}_1)) = 2\sum_{t=1}^{M} \langle E(X_t^+), X_t^+ - E(X_t^+) \rangle$$
$$\Delta(Q_2, \mathcal{Q}_2) := (Q_2 - E(Q_2)) - (\mathcal{Q}_2 - E(\mathcal{Q}_2)) = 2\sum_{t=1}^{M} \langle E(X_t^-), X_t^- - E(X_t^-) \rangle$$

For any $t > 0$, we have

$$\begin{aligned}
&P^*\Big\{\frac{P_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z})}{P_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z}^*)} > t\Big\} = P^*\Big\{(Q_1 - EQ_1) - (Q_2 - EQ_2) > M(\log t) - E(Q_1 - Q_2)\Big\} \\
&\le P^*\Big\{Q_1 - EQ_1 > \frac{M\log t - E(Q_1 - Q_2)}{2}\Big\} + P^*\Big\{Q_2 - EQ_2 < -\frac{M\log t - E(Q_1 - Q_2)}{2}\Big\} \\
&= P^*\Big\{\mathcal{Q}_1 - E\mathcal{Q}_1 > \frac{M\log t - E(Q_1 - Q_2)}{2} - \Delta(Q_1, \mathcal{Q}_1)\Big\} \\
&+ P^*\Big\{\mathcal{Q}_2 - E\mathcal{Q}_2 < -\frac{M\log t - E(Q_1 - Q_2)}{2} - \Delta(Q_2, \mathcal{Q}_2)\Big\}
\end{aligned}$$

where

$$
\begin{aligned}
&P^*\Big\{\mathcal{Q}_1-E\mathcal{Q}_1>\frac{M\log t-E(Q_1-Q_2)}{2}-\Delta(Q_1,\mathcal{Q}_1)\Big\} \\
\leq&\frac{1}{2}P^*\Big\{|\mathcal{Q}_1-E\mathcal{Q}_1|>\frac{M\log t-E(Q_1-Q_2)}{2}-\Delta(Q_1,\mathcal{Q}_1)\Big\} \qquad (3)\\
&P^*\Big\{\mathcal{Q}_2-E\mathcal{Q}_2>\frac{M\log t-E(Q_1-Q_2)}{2}-\Delta(Q_2,\mathcal{Q}_2)\Big\} \\
\leq&\frac{1}{2}P^*\Big\{|\mathcal{Q}_2-E\mathcal{Q}_2|>\frac{M\log t-E(Q_1-Q_2)}{2}-\Delta(Q_2,\mathcal{Q}_2)\Big\}. \qquad (4)
\end{aligned}
$$

Next, we estimate each of the term in (3). Given the $\{Y_{ij}^t\}_{t=1}^M$ are binary random variables and the setting that any two within-community edges $Y_{i_1j_1}$ and $Y_{i_2j_2}$ have a nonnegative correlation $corr(Y_{i_1j_1},Y_{i_2j_2})\geq 0$. Notice that

$$
corr\Big(f_1(Y_{i_1j_1}),f_1(Y_{i_2j_2})\Big)=\begin{cases} corr(Y_{i_1j_1},Y_{i_2j_2}) \text{ if } \mu_{z_iz_j}\geq\mu_{z_i^*z_j^*} \\ -corr(Y_{i_1j_1},Y_{i_2j_2}) \text{ if } \mu_{z_iz_j}<\mu_{z_i^*z_j^*}\end{cases}.
$$

We denote the covariance matrix of X_t^+ and X_t^- as Σ_1 and Σ_2. Notice that a term in (1) is zero only when its corresponding node membership is misclassified. Define the the number of nonzero term in (1) as N_r given $\|z-z^*\|_0=r$. Then we have $N_r=\frac{1}{2}rNM$. According to Lemma 1.1, X_t^+ is a subgaussian vector with a bounded subgaussian norm $L\leq C_1$ where C_1 is a positive constant and

$$
L=inf\{\alpha\geq 0: E(\exp(\langle z,X_t^+-E(X_t^+)\rangle))\leq\exp\{\alpha^2\langle\Sigma_1 z,z\rangle\}/2\}. \tag{5}
$$

Next we estimate $\|\Sigma_1\|_F,\|\Sigma_1\|_{op}$ and $\|\Sigma_2\|_F,\|\Sigma_2\|_{op}$ where $\|\cdot\|_F$ is the matrix Frobenius norm and $\|\cdot\|_{op}$ is the matrix spectral norm. Denote

$$
\Lambda=diag\Big(\sqrt{\mathrm{var}\{(X_t^+)_{12}\}},\sqrt{\mathrm{var}\{(X_t^+)_{13}\}},\cdots,\sqrt{\mathrm{var}\{(X_t^+)_{N-1,N}\}}\Big).
$$

Then $\|\Sigma_1\|_{op}=\|\Lambda R\Lambda\|_{op}\leq C_2\|R\|_{op}$ where R is the correlation matrix of X_t^+ and based on (C1),

$$
C_2\leq\max_{1\leq i<j\leq n}\mathrm{var}\{(X_t^+)_{ij}\}\leq\eta_N\max\{\log\frac{\zeta}{1-\zeta},\log\frac{1-\zeta}{\zeta}\}.
$$

Denote the largest eigenvalue of R as λ_R. From the Gershgorin circle theorem, we have

$$\lambda_R \le 1 + \max_{i=1,\cdots,N(N-1)/2} \sum_{j\neq i} |R_{ij}|.$$

Denote the number of node in the largest community is N_k. Note that the misclassification number of node $\|z - z^*\|_0 = r$ and edgewise correlation density λ both affect the sparsity of R, we have for each row in R:

$$\sum_{j\neq i} |R_{ij}| \le \rho N_k \min(r, \lambda N_k) \le \rho\kappa_2 N \min(r, \kappa_2 \lambda N),$$

where $\rho = \max_{i,j} R_{ij}$. Therefore, we have

$$\|\Sigma_1\|_{op} \le C\{1 + \rho\kappa_2\eta_N N \min(r, \kappa_2\lambda N)\},$$

for some constant C. Similarly we have a same upper bound for $\|\Sigma_2\|_{op}$. Notice that the dimension of R is $N_r \times N_r$ and $N_r \le rN$. In each row of R, the number of non-zero elements is less than $1 + N_k \min(r, \lambda N_k)$. Therefore, we have

$$\|\Sigma_1\|_F^2 \le C_2\rho^2 r\eta_N N\{1 + \kappa_2\eta_N N \min(r, \kappa_2\lambda N)\}.$$

Then we are able to estimate the upper bound for the first term in (3). According to the generalized Hanson-Wright inequality in (Chen and Yang (2018)), we have:

$$\frac{1}{2} P^*\Big\{|Q_1 - EQ_1| > s\Big\} \le \exp\Big\{ - C \min\Big(\frac{s^2}{L^4\|\Sigma_1\|_F^2\|A\|_F^2}, \frac{s}{L^2\|\Sigma_1\|_{op}\|A\|_{op}}\Big)\Big\}. \tag{6}$$

where $s = \frac{M\log t - E(Q_1 - Q_2)}{2} - \Delta(Q_1, \mathcal{Q}_1)$, $A = \boldsymbol{I}_{M\times M}$ and L is subgaussian norm of X_t^+ defined in (5). Then we have $L \le C_1$ and $\|A\|_F^2 = M, \|A\|_{op} = 1$. To estimate s, notice

$$\begin{aligned} E(Q_1 - Q_2) &= E[\sum_{t=1}^M \sum_{i<j} \Big\{Y_{ij}^t \log\frac{\mu_{z_i z_j}}{\mu_{z_i^* z_j^*}} + (1 - Y_{ij}^t)\log\frac{1-\mu_{z_i z_j}}{1-\mu_{z_i^* z_j^*}}\}] \\ &= -M\sum_{i<j}\Big\{\mu_{z_i^* z_j^*}\log\frac{\mu_{z_i^* z_j^*}}{\mu_{z_i z_j}} + (1-\mu_{z_i^* z_j^*})\log\frac{1-\mu_{z_i^* z_j^*}}{1-\mu_{z_i z_j}}\}, \end{aligned}$$

where there are total N_r non-zero terms in the summation. We introduce the function

$$k(x,y)=x\log(x/y)+(1-x)\log(1-x)/(1-y).$$

Notice that $k(x,y)>0$ for every $x,y\in(0,1)$. Then we define:

$$c^* := \min\{k(c_{ql},c_{q'l'})\}>0 \tag{7}$$

where the minimum are taken over $\{((q,l),(q',l'))\,|c_{q,l}\neq c_{q',l'}\}$. Given that $\eta_N=o_N(1)$, it can be shown that $k(\mu_{ql},\mu_{q'l'})\asymp \eta_N k(c_{ql},c_{q'l'})$. Combined with $N_r=\frac{1}{2}rNM$, we have $-E(Q_1-Q_2)>\frac{c^*}{2}r\eta_N NM$. To estimate $\Delta(Q_1,\mathcal{Q}_1)$, given all the elements in X_t^+ are bounded, we denote $\omega_1=\max_{1\le i<j\le n}E\{(X_t^+)_{ij}\},\omega_2=\max_{1\le i<j\le n}\mathrm{var}\{(X_t^+)_{ij}\}$

$$\begin{aligned}P\Big(|\Delta(Q_1,\mathcal{Q}_1)|>\frac{c^*}{2}rNM\Big)&\le P\Big(\omega_1|\sum_{t=1}^{M}\sum_{i=1}^{N_r}(X_{ti}^+-E(X_{ti}^+)|>\frac{c^*}{2}rNM\Big)\le\frac{\omega_1^2M\mathrm{var}(\sum_{i=1}^{N_r}X_{ti}^+)}{c^{*2}r^2N^2M^2/4}\\&\le\frac{\omega_1^2(\omega_2 rN(2+\rho\lambda rN))}{c^{*2}r^2N^2M}\le O(\frac{\eta_N}{M})\end{aligned}$$

Therefore, as M or N increases s is dominated by $-E(Q_1-Q_2)$ with probability approaching 1. Then for any fixed $t>0$, $s>\mathcal{O}_N(\frac{c^*}{2}rNM)$. Therefore, we have

$$\begin{aligned}&\min\Big(\frac{s^2}{L^4\|\Sigma_1\|_F^2\|A\|_F^2},\frac{s}{L^2\|\Sigma_1\|_{op}\|A\|_{op}}\Big)\\\ge&\min\Big(\frac{(\frac{c^*}{2}r\eta_N NM)^2}{C_1^2MC_2\rho^2rN\{1+\kappa_2\eta_N N\min(r,\kappa_2\lambda N)\}},\frac{\frac{c^*}{2}r\eta_N NM}{C_1C_2\{1+\rho\kappa_2\eta_N N\min(r,\kappa_2\lambda N)\}}\Big)\\\ge&C_3\frac{c^*r\eta_N NM}{1+\rho\kappa_2\eta_N N\min(r,\kappa_2\lambda N)}.\end{aligned}$$

where $C_3=\frac{c^*}{C_1^2C_2\rho^2}$. Hence for (6) we have:

$$\frac{1}{2}P^*\Big\{|Q_1-EQ_1|>s\Big\}\le\exp\Big\{-C\frac{c^*r\eta_N NM}{1+\rho\kappa_2\eta_N N\min(r,\kappa_2\lambda N)}\Big\},$$

where C is a positive constant. Follow Lemma 1.1, X_t^- is also subgaussian vector. Then we can obtain a same upper bound for

$$\frac{1}{2}P^*\Big\{|Q_2-EQ_2|>\frac{M\log t-E(Q_1-Q_2)}{2}\Big\}$$

in (3) through the above procedure. Therefore,

$$P^*\Big\{\frac{P_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z})}{P_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z^*})}>t\Big\}\le \exp\Big\{-C\frac{c^* r\eta_N NM}{1+\rho\kappa_2\eta_N N\min(r,\kappa_2\lambda N)}\Big\}.$$

Proof of Corollary 5.1

Given Theorem 5.1, we have

$$\begin{aligned}
&P_{Z*}\Big\{\sup_{\{z\neq z^*\}}\frac{L_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=z;\Theta)}{L_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=z^*;\Theta)}>t\Big\}\le P_{Z*}\Big\{\sum_{r=1}^N\sum_{\|z-z^*\|_1=r}\frac{L_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=z;\Theta)}{L_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=z^*;\Theta)}>t\Big\}\\
\le&\sum_{r=1}^N P_{Z*}\Big\{\sum_{\|z-z^*\|_1=r}\frac{L_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=z;\Theta)}{L_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=z^*;\Theta)}>t\Big\}\\
\le&\sum_{r=1}^N\binom{N}{r}(K-1)^r\exp\Big\{-C\frac{c^* r\eta_N NM}{1+\rho\kappa_2\eta_N N\min(r,\kappa_2\lambda N)}\Big\}\\
\le&\sum_{r=1}^N\binom{N}{r}\Big\{(K-1)\exp\Big\{-C\frac{c^*\eta_N NM}{1+\lambda\eta_N N^2}\Big\}\Big\}^r\le(1+\Big\{(K-1)\exp\Big\{-C\frac{c^*\eta_N NM}{1+\lambda\eta_N N^2}\Big\}\Big\})^N-1\\
\asymp&\mathcal{O}N\exp\Big\{-C\frac{c^*\eta_N NM}{1+\lambda\eta_N N^2}\Big\}
\end{aligned}$$

Proof of Theorem 5.2

We continue use the notations in the previous proof of Theorem 5.1. First decompose the proposed approximate likelihood in two parts:

$$\begin{aligned}
\log\frac{\tilde L(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z})}{\tilde L(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z^*})}=&\log\frac{P_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z})}{P_{ind}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z^*})}\\
&+\frac{1}{M}\sum_{t=1}^M\log\frac{1+\sum_{k=1}^K\max\Big\{\sum_{\substack{i<j;u<v\\(i,j)\neq(u,v)}}^N z_{ik}z_{jk}z_{uk}z_{vk}\rho_{ijuv}\hat Y_{ij}^{t,k}\hat Y_{uv}^{t,k},0\Big\}}{1+\sum_{k=1}^K\max\Big\{\sum_{\substack{i<j;u<v\\(i,j)\neq(u,v)}}^N z^*_{ik}z^*_{jk}z^*_{uk}z^*_{vk}\rho_{ijuv}\hat Y_{ij}^{t,k}\hat Y_{uv}^{t,k},0\Big\}}.
\end{aligned}$$

Notice that ρ_{ijuv} is the empirical estimator based in $\hat{Y}_{ij}^{t,k}$ and $\hat{Y}_{uv}^{t,k}$, then $\rho_{ijuv}\hat{Y}_{ij}^{t,k}\hat{Y}_{uv}^{t,k} > 0$ with high probability. Based on the mean value theorem, we have for some constant C_1 that

$$
\begin{aligned}
&\log \frac{1+\sum_{k=1}^{K} \max\Big\{ \sum_{\substack{i<j;u<v \\ (i,j)\neq(u,v)}}^{N} z_{ik}z_{jk}z_{uk}z_{vk}\rho_{ijuv}\hat{Y}_{ij}^{t,k}\hat{Y}_{uv}^{t,k}, 0\Big\}}{1+\sum_{k=1}^{K} \max\Big\{ \sum_{\substack{i<j;u<v \\ (i,j)\neq(u,v)}}^{N} z_{ik}^{*}z_{jk}^{*}z_{uk}^{*}z_{vk}^{*}\rho_{ijuv}\hat{Y}_{ij}^{t,k}\hat{Y}_{uv}^{t,k}, 0\Big\}} \\
=&C_1 \sum_{k=1}^{K} \Big\{ \max\Big(\sum_{\substack{i<j;u<v \\ (i,j)\neq(u,v)}}^{N} z_{ik}z_{jk}z_{uk}z_{vk}\rho_{ijuv}\hat{Y}_{ij}^{t,k}\hat{Y}_{uv}^{t,k}, 0\Big) - \max\Big(\sum_{\substack{i<j;u<v \\ (i,j)\neq(u,v)}}^{N} z_{ik}^{*}z_{jk}^{*}z_{uk}^{*}z_{vk}^{*}\rho_{ijuv}\hat{Y}_{ij}^{t,k}\hat{Y}_{uv}^{t,k}, 0\Big)\Big\} \\
\leq&C_1 \sum_{k=1}^{K} \Big\{ \sum_{\substack{i<j;u<v \\ (i,j)\neq(u,v)}}^{N} \big(z_{ik}z_{jk}z_{uk}z_{vk} - z_{ik}^{*}z_{jk}^{*}z_{uk}^{*}z_{vk}^{*}\big)\rho_{ijuv}\hat{Y}_{ij}^{t,k}\hat{Y}_{uv}^{t,k}\Big\}. \qquad (8)
\end{aligned}
$$

Notice in summation (8), the terms are non-zero only when $z_{ik}z_{jk}z_{uk}z_{vk} \neq z_{ik}^{*}z_{jk}^{*}z_{uk}^{*}z_{vk}^{*}$. We denote two node sets

$$
\begin{aligned}
\xi_1 &= \{(i,j,u,v) | z_{ik}z_{jk}z_{uk}z_{vk} = 1, z_{ik}^{*}z_{jk}^{*}z_{uk}^{*}z_{vk}^{*} = 0, k = 1, \cdots, K\}, \\
\xi_2 &= \{(i,j,u,v) | z_{ik}^{*}z_{jk}^{*}z_{uk}^{*}z_{vk}^{*} = 1, z_{ik}z_{jk}z_{uk}z_{vk} = 0, k = 1, \cdots, K\}.
\end{aligned}
$$

where $\#|\xi_1| = N_1$ and $\#|\xi_2| = N_2$. Given the number of misclassified nodes $\|z - z^{*}\|_0 = r$, we have $N_1 = \mathcal{O}(rN^3)$ and $N_2 = \mathcal{O}(rN^3)$. In the following, we construct the augmented edge vectors for the t th sample network by incorporating the vectorized pairwise edge interaction in (8) such that:

$$
\begin{aligned}
\tilde{X}_t^{+} &= \Big\{ X_t^{+}, \underbrace{\Big(\sqrt{\frac{C_1}{2}\{\rho_{ijuv}\hat{Y}_{ij}^{t,k}\hat{Y}_{uv}^{t,k}\}_{+}}\Big)_{1\times N_1}}_{\substack{(i,j,u,v)\in\xi_1 \\ z_{ik}z_{jk}z_{uk}z_{vk}=1 \\ k=1,\cdots,K}}, \underbrace{\Big(\sqrt{\frac{C_1}{2}\{-\rho_{ijuv}\hat{Y}_{ij}^{t,k}\hat{Y}_{uv}^{t,k}\}_{+}}\Big)_{1\times N_2}}_{\substack{(i,j,u,v)\in\xi_2 \\ z_{ik}^{*}z_{jk}^{*}z_{uk}^{*}z_{vk}^{*}=1 \\ k=1,\cdot,K}} \Big\}, \\
\tilde{X}_t^{-} &= \Big\{ X_t^{-}, \underbrace{\Big(\sqrt{\frac{C_1}{2}\{\rho_{ijuv}\hat{Y}_{ij}^{t,k}\hat{Y}_{uv}^{t,k}\}_{-}}\Big)_{1\times N_1}}_{\substack{(i,j,u,v)\in\xi_1 \\ z_{ik}z_{jk}z_{uk}z_{vk}=1 \\ k=1,\cdots,K}}, \underbrace{\Big(\sqrt{\frac{C_1}{2}\{-\rho_{ijuv}\hat{Y}_{ij}^{t,k}\hat{Y}_{uv}^{t,k}\}_{-}}\Big)_{1\times N_2}}_{\substack{(i,j,u,v)\in\xi_2 \\ z_{ik}^{*}z_{jk}^{*}z_{uk}^{*}z_{vk}^{*}=1 \\ k=1,\cdots,K}} \Big\}.
\end{aligned}
$$

where X_t^{+} and X_t^{-} are defined in (2). Denote the covariance matrix for $\tilde{X}_t^{+}$ and $\tilde{X}_t^{-}$ are $\tilde{\Sigma}_1$ and $\tilde{\Sigma}_2$ respectively. Since the second-order terms in X_t^{+} and X_t^{-} such as $\sqrt{\frac{C_1}{2}\{\rho_{ijuv}\hat{Y}_{ij}^{t,k}\hat{Y}_{uv}^{t,k}\}_{+}}$ only rescale the original edgewise interaction $\hat{Y}_{ij}^{t,k}\hat{Y}_{uv}^{t,k}$ then they preserve the third-order and fourth-order correlation within communities such that

$$|E\Big\{f_1(Y^t_{i_1j_1})\sqrt{\frac{C}{2}\{\rho_{ijuv}\hat{Y}^{t,k}_{i_2j_2}\hat{Y}^{t,k}_{i_3j_3}\}_+}\Big\}| = |E(\hat{Y}^{t,k}_{i_1j_1}\hat{Y}^{t,k}_{i_2j_2}\hat{Y}^{t,k}_{i_3j_3})|,$$

$$|E\Big\{f_2(Y^t_{i_1j_1})\sqrt{\frac{C}{2}\{\rho_{ijuv}\hat{Y}^{t,k}_{i_2j_2}\hat{Y}^{t,k}_{i_3j_3}\}_-}\Big\}| = |E(\hat{Y}^{t,k}_{i_1j_1}\hat{Y}^{t,k}_{i_2j_2}\hat{Y}^{t,k}_{i_3j_3})|,$$

$$|E\Big\{\sqrt{\frac{C}{2}\{\rho_{ijuv}\hat{Y}^{t,k}_{i_1j_1}\hat{Y}^{t,k}_{i_2j_2}\}_+}\sqrt{\frac{C}{2}\{\rho_{ijuv}\hat{Y}^{t,k}_{i_3j_3}\hat{Y}^{t,k}_{i_4j_4}\}_+}\Big\}| = |E(\hat{Y}^{t,k}_{i_1j_1}\hat{Y}^{t,k}_{i_2j_2}\hat{Y}^{t,k}_{i_3j_3}\hat{Y}^{t,k}_{i_4j_4})|,$$

$$|E\Big\{\sqrt{\frac{C}{2}\{\rho_{ijuv}\hat{Y}^{t,k}_{i_1j_1}\hat{Y}^{t,k}_{i_2j_2}\}_-}\sqrt{\frac{C}{2}\{\rho_{ijuv}\hat{Y}^{t,k}_{i_3j_3}\hat{Y}^{t,k}_{i_4j_4}\}_-}\Big\}| = |E(\hat{Y}^{t,k}_{i_1j_1}\hat{Y}^{t,k}_{i_2j_2}\hat{Y}^{t,k}_{i_3j_3}\hat{Y}^{t,k}_{i_4j_4})|.$$

Notice that each element in $\tilde{X}^+_t$ or $\tilde{X}^-_t$ is a bounded binary random variable. Follow the same procedure in Lemma 1.1, we can show that both $\tilde{X}^+_t$ and $\tilde{X}^-_t$ are subgaussian random vectors such that $L_1 \leq C_2, L_2 \leq C_2$ for some constant C_2 where L_1, L_2 are subgaussian norm of $\tilde{X}^+_t$ and $\tilde{X}^-_t$. Then consider the following quadratic forms:

$$\tilde{Q}_1 = \sum_{t=1}^{M}\langle \tilde{X}^+_t, \tilde{X}^+_t\rangle, \tilde{Q}_2 = \sum_{t=1}^{M}\langle \tilde{X}^-_t, \tilde{X}^-_t\rangle.$$

Therefore, we have

$$\log\frac{\tilde{L}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z})}{\tilde{L}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z}^*)} \leq \frac{1}{M}(\tilde{Q}_1 - \tilde{Q}_2).$$

Denote the centralized version quadratic forms $\tilde{Q}_1$ and $\tilde{Q}_2$ as $\tilde{\mathcal{Q}}_1$ and $\tilde{\mathcal{Q}}_2$ such that

$$\tilde{\mathcal{Q}}_1 = \sum_{t=1}^{M}\langle \tilde{X}^+_t - E(\tilde{X}^+_t), \tilde{X}^+_t - E(\tilde{X}^+_t)\rangle, \tilde{\mathcal{Q}}_2 = \sum_{t=1}^{M}\langle \tilde{X}^-_t - E(\tilde{X}^-_t), \tilde{X}^-_t - E(\tilde{X}^-_t)\rangle.$$

Denote the following quadratic difference as:

$$\Delta(\tilde{Q}_1, \tilde{\mathcal{Q}}_1) := (\tilde{Q}_1 - E(\tilde{Q}_1)) - (\tilde{\mathcal{Q}}_1 - E(\tilde{\mathcal{Q}}_1)) = 2\sum_{t=1}^{M}\langle E(\tilde{X}^+_t), \tilde{X}^+_t - E(\tilde{X}^+_t)\rangle$$
$$\Delta(\tilde{Q}_2, \tilde{\mathcal{Q}}_2) := (\tilde{Q}_2 - E(\tilde{Q}_2)) - (\tilde{\mathcal{Q}}_2 - E(\tilde{\mathcal{Q}}_2)) = 2\sum_{t=1}^{M}\langle E(\tilde{X}^-_t), \tilde{X}^-_t - E(\tilde{X}^-_t)\rangle$$

Similar to (3), for any fixed $t > 0$:

$$
\begin{aligned}
&P^*\Big\{\frac{\tilde{L}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z})}{\tilde{L}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z}^*)} > t\Big\} \le P^*\Big\{\frac{1}{M}(\tilde{Q}_1 - \tilde{Q}_2) > \log t\Big\} \\
\le & P^*\Big\{\tilde{Q}_1 - E\tilde{Q}_1 > \frac{M\log t - E(\tilde{Q}_1 - \tilde{Q}_2)}{2}\Big\} + P^*\Big\{\tilde{Q}_2 - E\tilde{Q}_2 < -\frac{M\log t - E(\tilde{Q}_1 - \tilde{Q}_2)}{2}\Big\} \\
= & P^*\Big\{\tilde{\mathcal{Q}}_1 - E\tilde{\mathcal{Q}}_1 > \frac{M\log t - E(\tilde{Q}_1 - \tilde{Q}_2)}{2} - \Delta(\tilde{Q}_1, \tilde{\mathcal{Q}}_1)\Big\} \\
+ & P^*\Big\{\tilde{\mathcal{Q}}_2 - E\tilde{\mathcal{Q}}_2 < -\frac{M\log t - E(\tilde{Q}_1 - \tilde{Q}_2)}{2} - \Delta(\tilde{Q}_2, \tilde{\mathcal{Q}}_2)\Big\} \\
\le & \frac{1}{2}P^*\Big\{|\tilde{\mathcal{Q}}_1 - E\tilde{\mathcal{Q}}_1| > \frac{M\log t - E(\tilde{Q}_1 - \tilde{Q}_2)}{2} - \Delta(\tilde{Q}_1, \tilde{\mathcal{Q}}_1)\Big\} \\
+ & \frac{1}{2}P^*\Big\{|\tilde{\mathcal{Q}}_2 - E\tilde{\mathcal{Q}}_2| > \frac{M\log t - E(\tilde{Q}_1 - \tilde{Q}_2)}{2} - \Delta(\tilde{Q}_2, \tilde{\mathcal{Q}}_2)\Big\}.
\end{aligned} \tag{9}
$$

Next we estimate $\|\tilde{\Sigma}_1\|_F, \|\tilde{\Sigma}_1\|_{op}$ and $\|\tilde{\Sigma}_2\|_F, \|\tilde{\Sigma}_2\|_{op}$. Denote

$$
\tilde{\Lambda} = diag(\Lambda, \underbrace{sd\Big(\sqrt{\frac{1}{2}\{\rho_{ijuv}\hat{Y}_{ij}^{t,k}\hat{Y}_{uv}^{t,k}\}_+}\Big)_{1\times N_1}}_{\substack{(i,j,u,v)\in\xi_1 \\ z_{ik}z_{jk}z_{uk}z_{vk}=1 \\ k=1,\cdots,K}}, \underbrace{sd\Big(\sqrt{\frac{1}{2}\{-\rho_{ijuv}\hat{Y}_{ij}^{t,k}\hat{Y}_{uv}^{t,k}\}_+}\Big)_{1\times N_2}}_{\substack{(i,j,u,v)\in\xi_2 \\ z^*_{ik}z^*_{jk}z^*_{uk}z^*_{vk}=1 \\ k=1,\cdots,K}}),
$$

then $\|\tilde{\Sigma}_1\|_{op} = \|\tilde{\Lambda}\tilde{R}\tilde{\Lambda}\|_{op} \le C_3\|\tilde{R}\|_{op}$ where $\tilde{R}$ is the correlation matrix of $\tilde{X}_t^+$ and C_3 is the largest variance of elements in $\tilde{X}_t^+$. Denote the largest eigenvalue of $\tilde{R}$ as $\lambda_{\tilde{R}}$. From the Gershgorin circle theorem, we have

$$
\lambda_{\tilde{R}} \le 1 + \max_i \sum_{j\neq i}\left|\tilde{R}_{ij}\right|.
$$

Given that the misclassification number of node $\|z - z^*\|_0 = r$, edgewise correlation density λ and condition C3, for each row in $\tilde{R}$, there exists some constant C_4>0 such that:

$$
\sum_{j\neq i}|R_{ij}| \le C_4\rho N_k \min(r, \lambda N_k) = C_4\rho\kappa_2 N\min(r, \kappa_2\lambda N), \tag{10}
$$

where $\rho = \max_{i,j}\tilde{R}_{ij}$. Therefore, we have

$$
\|\tilde{\Sigma}_1\|_{op} \le C_3\{1 + C_4\rho\kappa_2 N\min(r, \kappa_2\lambda N)\}.
$$

Similarly, $\|\tilde{\Sigma}_2\|_{op}$ follows a same upper bound. Notice that the dimension of $\tilde{R}$ is $(N_r + N_1 + N_2) \times (N_r + N_1 + N_2)$. Under the condition C3, in each row of $\tilde{R}$, the number of non-zero elements is less than $1 + C_4 N_k \min(r, \lambda N_k)$. Therefore, we have for a constant $C' > 0$:

$$
\begin{aligned}
\|\tilde{\Sigma}_1\|_F^2 \leq & C_3 \rho^2 (N_r + N_1 + N_2)\{1 + C_4 \kappa_2 N \min(r, \kappa_2 \lambda N)\} \\
\leq & C' \rho^2 (rN + rN^3)\{1 + C_4 \kappa_2 N \min(r, \kappa_2 \lambda N)\}.
\end{aligned}
$$

According to the generalized Hanson-Wright inequality in (Chen and Yang (2018)):

$$
\frac{1}{2} P^* \Big\{ |\tilde{Q}_1 - E\tilde{Q}_1| > s \Big\} \leq \exp\Big\{ -C \min\Big(\frac{s^2}{L_1^4 \|\tilde{\Sigma}_1\|_F^2 \|A\|_F^2}, \frac{s}{L_1^2 \|\tilde{\Sigma}_1\|_{op} \|A\|_{op}} \Big) \Big\}, \tag{11}
$$

where $s = \frac{M \log t - E(\tilde{\mathcal{Q}}_1 - \tilde{\mathcal{Q}}_2)}{2} - \Delta(\tilde{Q}_1, \tilde{\mathcal{Q}}_1)$, $A = \boldsymbol{I}_{M\times M}$ and L_1 is subgaussian norm of $\tilde{X}_t^+$. Notice $\|A\|_F^2 = M$, $\|A\|_{op} = 1$. Given (8), we have

$$
E(\tilde{Q}_1 - \tilde{Q}_2) = E(Q_1 - Q_2) + C_1 \sum_{k=1}^{K} \Big\{ \sum_{\substack{i<j; u<v \\ (i,j)\neq(u,v)}}^{N} (z_{ik} z_{jk} z_{uk} z_{vk} - z_{ik}^* z_{jk}^* z_{uk}^* z_{vk}^*) \rho_{ijuv} E(\hat{Y}_{ij}^{t,k} \hat{Y}_{uv}^{t,k}) \Big\}.
$$

Denote ρ_{min} as the lower bound of all non-zero correlation among edges such that $E(\hat{Y}_{ij}^{t,k} \hat{Y}_{uv}^{t,k}) = \rho_{ijuv} \geq \rho_{min}$. Given the edges from different communities are independent and within-community correlation density λ, we have for some positive constant C_5,

$$
\#|\{(i,j,u,v) : E(\hat{Y}_{ij}^{t,k} \hat{Y}_{uv}^{t,k}) \neq 0, (i,j,u,v) \in \xi_2\}| = \lambda N_1 = \lambda C_5 r N^3,
$$

$$
\#|\{(i,j,u,v) : E(\hat{Y}_{ij}^{t,k} \hat{Y}_{uv}^{t,k}) \neq 0, (i,j,u,v) \in \xi_1\}| \leq \lambda \binom{r}{4}.
$$

Assume that $r \leq cN$ for some constant $0 < c < 1$, we have for some constant $c_0 > 0$:

$$
-E(\tilde{Q}_1 - \tilde{Q}_2) \geq \frac{c^*}{2} rNM + \lambda M \frac{C_1 \rho_{min}^2}{2} (C_5 r N^3 - \binom{r}{4}) \geq c_0 r (c^* \eta_N N + \lambda N^3) M.
$$

To estimate $\Delta(\tilde{Q}_1, \tilde{\mathcal{Q}}_1)$, given all the elements in $\tilde{X}_t^+$ are bounded, we denote

$$
\omega_3 = \max_i E\{(\tilde{X}_t^+)_i\}, \omega_4 = \max_i \mathrm{var}\{(\tilde{X}_t^+)_i\}.
$$

According to the definition of $\tilde{X}_t^+$ and $N_1 = \mathcal{O}(rN^3), N_2 = \mathcal{O}(rN^3)$, there exists a positive constant C^+ such that $\#|\tilde{X}_t^+| = \frac{rN}{2} + C^+ rN^3$, therefore

$$\begin{aligned}
&P\Big(|\Delta(\tilde{Q}_1, \tilde{\mathcal{Q}}_1)| > c_0 r(c^*\eta_N N + \lambda N^3)M\Big) \\
&\leq P\Big(\omega_3 |\sum_{t=1}^{M} \sum_{i=1}^{\#|\tilde{X}_t^+|} (\tilde{X}_{ti}^+ - E(\tilde{X}_{ti}^+)| > c_0 r(c^*\eta_N N + \lambda N^3)M\Big) \\
&\leq \frac{\omega_3^2 M \mathrm{var}(\sum_{i=1}^{\#|\tilde{X}_t^+|} \tilde{X}_{ti}^+)}{c_0^2 r^2 (c^*\eta_N N + \lambda N^3)^2 M^2}
\end{aligned} \tag{12}$$

From the assumption (C3), there exists a positive constant ω_5 such that

$$\begin{aligned}
&\mathrm{var}(\sum_{i=1}^{\#|\tilde{X}_t^+|} \tilde{X}_{ti}^+) = \sum_{i=1}^{\#|\tilde{X}_t^+|} \mathrm{var}(\tilde{X}_{ti}^+) + \sum_{i,j} \mathrm{cov}(\tilde{X}_{ti}^+, \tilde{X}_{tj}^+) \\
\leq &\omega_4 (\frac{rN}{2} + C^+ rN^3) + w_4 \rho (\frac{\lambda r^2 N^2}{4} + rN^3 \cdot \omega_5 \lambda N^2 + rN \cdot \omega_5 \lambda N^2)
\end{aligned} \tag{13}$$

Through combining (12) and (13), give $\lambda > 0$ we have

$$P\Big(|\Delta(\tilde{Q}_1, \tilde{\mathcal{Q}}_1)| > c_0 r(c^*\eta_N N + \lambda N^3)M\Big) \leq \frac{\omega_3^2 \omega_4}{c_0^2 M} (\frac{1}{2r\lambda^2 N^5} + \frac{C^+}{r\lambda^2 N^3} + \frac{\rho}{4\lambda N^4} + \frac{\rho\omega_5}{r\lambda N} + \frac{\rho\omega_5}{r\lambda N^3})$$

Therefore, given $N > \mathcal{O}_N(\frac{1}{\lambda})$ and M, N increasing, s is dominated by $-E(\tilde{Q}_1 - \tilde{Q}_2)$ with probability approaching 1. Given any fixed $t > 0$, $s > \mathcal{O}_N(r(c^*\eta_N N + \lambda N^3)M)$. For the first term in (11),

$$\frac{s^2}{L_1^4 \|\tilde{\Sigma}_1\|_F^2 \|A\|_F^2} \geq \frac{r^2 (c^*\eta_N N + \lambda N^3)^2 M^2}{L_1^4 C' \rho^2 (rN + rN^3)\{1 + C_4 \kappa_2 N \min(r, \kappa_2 \lambda N)\} M}.$$

For the second term in (11),

$$\frac{s}{L_1^2 \|\tilde{\Sigma}_1\|_{op} \|A\|_{op}} \geq \frac{r(c^*\eta_N N + \lambda N^3)M}{L_1^2 C' \{1 + C_4 \rho \kappa_2 N \min(r, \kappa_2 \lambda N)\}}.$$

Given $\lambda > 0$, we have for some constant $C_6 > 0$

$$\min\Big(\frac{s^2}{L_1^4 \|\tilde{\Sigma}_1\|_F^2 \|A\|_F^2}, \frac{s}{L_1^2 \|\tilde{\Sigma}_1\|_{op} \|A\|_{op}}\Big) \geq C_6 \frac{r\lambda NM(c^*\eta_N + \lambda N^2)}{1 + C_4 \rho \kappa_2 N \min(r, \kappa_2 \lambda N)}. \tag{14}$$

Follow the same procedure we can show a upper bound for $P^*\Big\{|\tilde{Q}_2 - E\tilde{Q}_2| > s\Big\}$ with a same order to (14). Combined with (9) and (11), we have for $\lambda > 0$ and some constant $C > 0$:

$$P_{Z*}\Big\{\frac{\tilde{L}(\boldsymbol{Y}|\boldsymbol{Z}=z;\Theta)}{\tilde{L}(\boldsymbol{Y}|\boldsymbol{Z}=z^*;\Theta)} > t\Big\} \leq \exp\Big\{-C\frac{r\lambda NM(c^*\eta_N+\lambda N^2)}{1+C_4\rho\kappa_2 N\min(r,\kappa_2\lambda N)}\Big\},$$

Proof of Corollary 5.2

The proof follows a similar discussion for Corollary 5.1.

Proof of Theorem 5.3

Follow the notations introduced in Theorem 5.1 and Theorem 5.2, we further define that $\boldsymbol{w} = \max \frac{P^{(s)}(Z_i=q)}{P^{(s)}(Z_i=l)}$, $i = 1, \cdots, N$, $q, l = 1, \cdots, K$. Let $\boldsymbol{E}$ stands for the operator of expectation step in Algorithm 1 in Section 4.

We first consider the misclassification of updated estimated membership for node s, e.g., $\boldsymbol{E}(z_s)$ from the current estimation α_s. We denote that $\boldsymbol{\alpha}_{-s}$ as the probability estimations of nodes' memberships at current step except node s and assume the true membership of node s is b, i.e., $z_s^* = b$. If we use the marginal likelihood, then:

$$\begin{aligned}
&\|\boldsymbol{E}(z_s) - z_s^*\|_1 = \\
&|\frac{P(z_s=1)\tilde{L}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s};z_s=1)}{\sum_{q=1}^K P(z_s=q)\tilde{L}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s};z_s=q)} - 0| + \cdots + |\frac{P(z_s=b)\tilde{L}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s};z_s=b)}{\sum_{q=1}^K P(z_s=K)\tilde{L}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s};z_s=K)} - 1| \\
&\leq 2\frac{\sum_{q\neq b} P(z_s=q)\tilde{L}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s};z_s=q)}{\sum_{q=1}^K P(z_s=q)\tilde{L}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s};Z_s=q)} \leq 2\boldsymbol{w}\sum_{q\neq b}^K \frac{\tilde{L}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s};z_s=q)}{\tilde{L}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s};z_s=b)} \\
&= 2\boldsymbol{w}\sum_{q\neq b}^K \min[1, \exp\{\log\tilde{L}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s};Z_s=q) - \log\tilde{L}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s};z_s=b)\}].
\end{aligned} \tag{15}$$

Then given node s belongs to different communities while the estimated membership for other nodes $\boldsymbol{\alpha}_{-s}$ are fixed. We decompose the proposed approximate likelihood into marginal part and correlation part in the following: $\log\tilde{L}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s};z_s) = \log L_{mar}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s};z_s) + \log L_{cor}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s};z_s)$. The marginal likelihood $\log L_{mar}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s}; z_s)$,

$$\begin{aligned}
&\log L_{mar}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s}; z_s = a)\\
&= \frac{1}{M}\sum_{t=1}^{M}\left[\underbrace{log\prod_{q,l}^{K}\prod_{i\neq j\neq s}^{N}\left\{\mu_{ql}^{Y_{ij}^t}(1-\mu_{ql})^{(1-Y_{ij}^t)}\right\}^{\alpha_{iq}\alpha_{jl}}}_{\text{not depend on } z_s} + \prod_{q=1}^{K}\prod_{i\neq s}^{N}\left\{\mu_{qa}^{Y_{is}^t}(1-\mu_{qa})^{(1-Y_{is}^t)}\right\}^{\alpha_{iq}}\right].
\end{aligned}$$

Therefore, the discrepancy among marginal likelihood is

$$\begin{aligned}
&\log L_{mar}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s}; z_s = a) - \log L_{mar}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s}; z_s = b)\\
=&\frac{1}{M}\sum_{t=1}^{M}\sum_{q=1}^{K}\sum_{i\neq s}^{N}\left[\alpha_{iq}\{Y_{is}^t\log\frac{\hat{\mu}_{qa}}{\hat{\mu}_{qb}} + (1-Y_{is}^t)\log\frac{1-\hat{\mu}_{qa}}{1-\hat{\mu}_{qb}}\}\right]\\
=&\frac{1}{M}\sum_{t=1}^{M}\sum_{q=1}^{K}\sum_{i\neq s}^{N}\left[\alpha_{iq}\{Y_{is}^t\log\frac{\mu_{qa}}{\mu_{qb}} + (1-Y_{is}^t)\log\frac{1-\mu_{qa}}{1-\mu_{qb}}\}\right]\\
+&\frac{1}{M}\sum_{t=1}^{M}\sum_{q=1}^{K}\sum_{i\neq s}^{N}\left[\alpha_{iq}\{Y_{is}^t\log\frac{\mu_{qa}\hat{\mu}_{qb}}{\hat{\mu}_{qa}\mu_{qb}} + (1-Y_{is}^t)\log\frac{(1-\mu_{qa})(1-\hat{\mu}_{qb})}{(1-\hat{\mu}_{qa})(1-\hat{\mu}_{qb})}\}\right]
\end{aligned}$$

We can decompose the marginal discrepancy into four parts:

$$\begin{aligned}
&\log L_{mar}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s}; z_s = a) - \log L_{mar}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s}; z_s = b)\\
=&\underbrace{\frac{1}{M}\sum_{t=1}^{M}\sum_{q=1}^{K}\sum_{i\neq s}^{N}(\alpha_{iq}-z_{iq}^*)\{Y_{is}^t - E(Y_{is}^t)\}(\log\frac{\mu_{qa}}{\mu_{qb}} - \log\frac{1-\mu_{qa}}{1-\mu_{qb}})}_{\boldsymbol{A_1}}\\
+&\underbrace{\frac{1}{M}\sum_{t=1}^{M}\sum_{q=1}^{K}\sum_{i\neq s}^{N}\left[(\alpha_{iq}-z_{iq}^*)\{EY_{is}^t\log\frac{\mu_{qa}}{\mu_{qb}} + (1-EY_{is}^t)\log\frac{1-\mu_{qa}}{1-\mu_{qb}}\}\right]}_{\boldsymbol{A_2}}\\
+&\underbrace{\frac{1}{M}\sum_{t=1}^{M}\sum_{q=1}^{K}\sum_{i\neq s}^{N}\left[z_{iq}^*\{Y_{is}^t\log\frac{\mu_{qa}}{\mu_{qb}} + (1-Y_{is}^t)\log\frac{1-\mu_{qa}}{1-\mu_{qb}}\}\right]}_{\boldsymbol{A_3}}\\
+&\underbrace{\frac{1}{M}\sum_{t=1}^{M}\sum_{q=1}^{K}\sum_{i\neq s}^{N}\left[\alpha_{iq}\{Y_{is}^t\log\frac{\mu_{qa}\hat{\mu}_{qb}}{\hat{\mu}_{qa}\mu_{qb}} + (1-Y_{is}^t)\log\frac{(1-\mu_{qa})(1-\hat{\mu}_{qb})}{(1-\hat{\mu}_{qa})(1-\hat{\mu}_{qb})}\}\right]}_{\boldsymbol{A_4}}.
\end{aligned}$$

For the correlation part, we consider the pairwise interaction terms in the $\log L_{cor}(\boldsymbol{Y}|\boldsymbol{\alpha})$. Notice that for $t = 1, \cdots, M$

$$\sum_{\substack{i<j;k<g\\(i,j)\neq(k,g)}}^{N} \alpha_{ia}\alpha_{ja}\alpha_{ka}\alpha_{ga}\rho_{ijkg}\hat{Y}_{ij}^{t,a}\hat{Y}_{kg}^{t,a} = (\sum_{i\neq s}^{N} \alpha_{sa}\alpha_{ia}\hat{Y}_{si}^{t,a})(\sum_{i<j}^{N} \alpha_{ia}\alpha_{ja}\hat{Y}_{ij}^{t,a}) - \sum_{i\neq s}^{N}(\alpha_{ia}\hat{Y}_{si}^{t,a})^2 + A_a^t,$$

where A_q^t does not depend on z_s. Since the first term $(\sum_{i\neq s}^{N} \alpha_{sa}\alpha_{ia}\hat{Y}_{si}^{t,a})(\sum_{i<j}^{N} \alpha_{ia}\alpha_{ja}\hat{Y}_{ij}^{t,a}) = o(N^3)$ and the second term $\sum_{i\neq s}^{N}(\alpha_{ia}\hat{Y}_{si}^{t,a})^2 = o(N)$, without loss of generality, we can keep the first dominating term when N is large. For the correlation part $\log L_{cor}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s}; z_s)$, if $\alpha_{sq} = 0,\ q \neq a$ and $\alpha_{sa} = 1$:

$$\begin{aligned}\log L_{cor}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s}; Z_s = a) =& \frac{1}{M}\sum_{t=1}^{M}\Big\{1 + \sum_{q=1}^{K}\frac{\rho_q}{2}\max(\sum_{\substack{i<j;k<g\\(i,j)\neq(k,g)}}^{N} \alpha_{iq}\alpha_{jq}\alpha_{kq}\alpha_{gq}\hat{Y}_{ij}^{t,q}\hat{Y}_{kg}^{t,q}, 0)\Big\}\\ =& 1 + \underbrace{\frac{1}{M}\sum_{t=1}^{M}\sum_{q=1}^{K}\frac{\rho_q}{2}A_q^t}_{\boldsymbol{A}} + \underbrace{\frac{\rho_a}{2}(\sum_{i\neq s}^{N}\alpha_{sa}\alpha_{ia}\hat{Y}_{si}^{t,a})(\sum_{i<j}^{N}\alpha_{ia}\alpha_{ja}\hat{Y}_{ij}^{t,a})}_{\boldsymbol{B}_a}.\end{aligned}$$

Through the Taylor expansion, the discrepancy of correlation information when node s belongs to different communities a and b:

$$\begin{aligned}\log L_{cor}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s}; Z_s = a) - \log L_{cor}(\boldsymbol{Y}|\boldsymbol{\alpha}_{-s}; Z_s = b) &= \log(1 + \boldsymbol{A} + \boldsymbol{B_a}) - \log(1 + \boldsymbol{A} + \boldsymbol{B_b})\\ &= \log(1 + \frac{\boldsymbol{B_a} - \boldsymbol{B_b}}{1 + \boldsymbol{A} + \boldsymbol{B_b}}) \leq \boldsymbol{C}_A(\boldsymbol{B_a} - \boldsymbol{B_b}),\end{aligned}$$

where $\boldsymbol{C}_A$ is a constant relating to the gradient of function $\log(1 + 1/x)$ at $\boldsymbol{A}$. Then we set $\rho = \min \rho_q, q = 1, \cdots, K$

$$\begin{aligned}\boldsymbol{B_a} - \boldsymbol{B_b} =& (\sum_{i\neq s}^{N}\alpha_{ia}\hat{Y}_{si}^{t,a})(\sum_{i<j}^{N}\alpha_{ia}\alpha_{ja}\hat{Y}_{ij}^{t,a}) - (\sum_{i\neq s}^{N}\alpha_{ib}\hat{Y}_{si}^{t,b})(\sum_{i<j}^{N}\alpha_{ib}\alpha_{jb}\hat{Y}_{ij}^{t,b})\\ &\leq \frac{\rho}{4}\Big(\langle\alpha_a \otimes vec(\alpha_a^T\alpha_a), \hat{Y}_{\cdot s}^{t,a} \otimes vec(\hat{\boldsymbol{Y}}^{t,a})\rangle - \langle\alpha_b \otimes vec(\alpha_b^T\alpha_b), \hat{Y}_{\cdot s}^{t,b} \otimes vec(\hat{\boldsymbol{Y}}^{t,b})\rangle\Big).\end{aligned}$$

For the simplicity of notation, we define and decompose the correlation discrepancy as followings:

$$
\begin{aligned}
\boldsymbol{B} :=& \sum_{t=1}^{M} \frac{\rho C_A}{4M} \Big(\langle \alpha_a \otimes vec(\alpha_a^T \alpha_a), \hat{Y}_{\cdot s}^{t,a} \otimes vec(\hat{\boldsymbol{Y}}^{t,a}) \rangle - \langle \alpha_b \otimes vec(\alpha_b^T \alpha_b), \hat{Y}_{\cdot i}^{t,b} \otimes vec(\hat{\boldsymbol{Y}}^{t,b}) \rangle \Big) \\
=& \underbrace{\frac{\rho C_A}{4M} \sum_{t=1}^{M} \Big(\langle \alpha_a \otimes vec(\alpha_a^T \alpha_a) - z_a^* \otimes vec(z_a^{*T} z_a^*), \hat{Y}_{\cdot s}^{t,a} \otimes vec(\hat{\boldsymbol{Y}}^{t,a}) \rangle -}_{} \\
& \underbrace{\langle \alpha_b \otimes vec(\alpha_b^T \alpha_b) - z_b^* \otimes vec(z_b^{*T} z_b^*), \hat{Y}_{\cdot s}^{t,b} \otimes vec(\hat{\boldsymbol{Y}}^{t,b}) \rangle \Big)}_{\text{misclassification error:}\boldsymbol{B}_1} \\
& + \underbrace{\frac{\rho C_A}{4M} \sum_{t=1}^{M} \Big(\langle z_a^* \otimes vec(z_a^{*T} z_a^*), \hat{Y}_{\cdot s}^{t,a} \otimes vec(\hat{\boldsymbol{Y}}^{t,a}) \rangle - \langle z_b^* \otimes vec(z_b^{*T} z_b^*), \hat{Y}_{\cdot s}^{t,b} \otimes vec(\hat{\boldsymbol{Y}}^{t,b}) \rangle \Big)}_{\text{estimation bias:}\boldsymbol{B}_2}.
\end{aligned}
$$

Notice that $\min\{1, \exp(x)\} \leq \exp(x_0) + \sum_{l=0}^{m-1} \frac{1-\exp(x_0)}{m} \mathbb{1}\{x \geq (1 - l/m)x_0\}$ and set $x_0 = -\alpha' MN$, where $\alpha' = \frac{\lambda(c^* \eta_N + \lambda N^2)}{1+\lambda N^2}$. Given (15), we have:

$$
E\|\boldsymbol{\alpha}^{s+1} - \boldsymbol{z}^*\|_1 \leq 2\boldsymbol{w}NK \exp(-\alpha' MN) + 2\boldsymbol{w} \sum_{l=0}^{m-1} \sum_{a=1}^{K} \sum_{b \neq a} \sum_{i: z_i^* = b} \frac{1 - exp(\alpha' MN)}{m} E(\boldsymbol{L}_2) \tag{16}
$$

where $E(\boldsymbol{L}) = \mathbb{P}\Big(\boldsymbol{A} + \boldsymbol{B} \geq \frac{m-l}{m} x_0\Big)$. For some specific $t > 0$,

$$
\begin{aligned}
& \mathbb{P}\Big(\boldsymbol{A} + \boldsymbol{B} \geq \frac{m-l}{m} x_0\Big) = \mathbb{P}\Big(\boldsymbol{A_1} + \boldsymbol{A_2} + \boldsymbol{A_3} + \boldsymbol{A_4} + \boldsymbol{B_1} + \boldsymbol{B_2} \geq \frac{m-l}{m} x_0\Big) \\
\leq & \mathbb{P}\Big(\boldsymbol{A_1} + \boldsymbol{B_1} \geq t\Big) + \mathbb{P}\Big(\boldsymbol{A_3} + \boldsymbol{B_2} \geq \frac{m-l}{m} x_0 - t - \boldsymbol{A_2} - \boldsymbol{A_4}\Big).
\end{aligned} \tag{17}
$$

We then transfer $\boldsymbol{A_3} + \boldsymbol{B_2}$ into a quadratic form. For each community $q, q = 1, \cdots, K$ define the transformations:

$$
\begin{aligned}
f_q^+(x) &= \sqrt{\Big[z_{iq}^* \{Y_{is}^t \log \frac{\mu_{qa}}{\mu_{qb}} + (1 - Y_{is}^t) \log \frac{1 - \mu_{qa}}{1 - \mu_{qb}}\}\Big]_+}, \\
f_q^-(x) &= \sqrt{\Big[z_{iq}^* \{Y_{is}^t \log \frac{\mu_{qa}}{\mu_{qb}} + (1 - Y_{is}^t) \log \frac{1 - \mu_{qa}}{1 - \mu_{qb}}\}\Big]_-}, \\
X_t^+ &= \{f_1^+(Y_{1s}^t), \cdots, f_1^+(Y_{ns}^t), f_2^+(Y_{1s}^t), \cdots, f_2^+(Y_{Ns}^t), \cdots, f_K^+(Y_{1s}^t), \cdots, f_K^+(Y_{Ns}^t)\}, \\
X_t^- &= \{f_1^-(Y_{1s}^t), \cdots, f_1^-(Y_{Ns}^t), f_2^-(Y_{1s}^t), \cdots, f_2^-(Y_{Ns}^t), \cdots, f_K^-(Y_{1s}^t), \cdots, f_K^-(Y_{Ns}^t)\}.
\end{aligned}
$$

Notice that the total number of non-zero terms in X_t^+ or X_t^- is N. We define the node sets

$$\tilde{\xi}_a = \{(i_1, i_2, i_3)|z^*_{i_1a}z^*_{i_2a}z^*_{i_3a} = 1\} \quad \tilde{\xi}_b = \{(i_1, i_2, i_3)|z^*_{i_1b}z^*_{i_2b}z^*_{i_3b} = 1\}.$$

Note $\#|\tilde{\xi}_a| = o(N_a^3)$ and $\#|\tilde{\xi}_b| = o(N_b^3)$ where N_a and N_b are number of node in community a and b. We further define augmented edges vectors:

$$\bar{X}_t^+ = \Bigg(X_t^+, \underbrace{\Big(\frac{C_A}{4}\sqrt{\{\rho_{i_1si_2i_3}\hat{Y}^{t,a}_{i_1s}\hat{Y}^{t,a}_{i_2i_3}\}_+}\Big)_{1\times\#|\tilde{\xi}_a|}}_{(i_1,i_2,i_3)\in\tilde{\xi}_a}, \underbrace{\Big(\frac{C_A}{4}\sqrt{\{-\rho_{i_1si_2i_3}\hat{Y}^{t,b}_{i_1s}\hat{Y}^{t,b}_{i_2i_3}\}_+}\Big)_{1\times\#|\tilde{\xi}_b|}}_{(i_1,i_2,i_3)\in\tilde{\xi}_b}\Bigg),$$
$$\bar{X}_t^- = \Bigg(X_t^-, \underbrace{\Big(\frac{C_A}{4}\sqrt{\{\rho_{i_1si_2i_3}\hat{Y}^{t,a}_{i_1s}\hat{Y}^{t,a}_{i_2i_3}\}_-}\Big)_{1\times\#|\tilde{\xi}_a|}}_{(i_1,i_2,i_3)\in\tilde{\xi}_a}, \underbrace{\Big(\frac{C_A}{4}\sqrt{\{-\rho_{i_1si_2i_3}\hat{Y}^{t,b}_{i_1s}\hat{Y}^{t,b}_{i_2i_3}\}_-}\Big)_{1\times\#|\tilde{\xi}_b|}}_{(i_1,i_2,i_3)\in\tilde{\xi}_b}\Bigg).$$

Denote the covariance of $\bar{X}_t^+$ and $\bar{X}_t^-$ as $\bar{\Sigma}_1$ and $\bar{\Sigma}_2$. Note that each element in $\bar{X}_t^+$ or $\bar{X}_t^-$ is a bounded binary random variable. Similarly, $\bar{X}_t^+$ and $\bar{X}_t^-$ are subgaussian vectors. Therefore,

$$\boldsymbol{A_3} + \boldsymbol{B_2} = \frac{1}{M}\sum_{t=1}^{M}\Big(\langle\bar{X}_t^+, \bar{X}_t^+\rangle - \langle\bar{X}_t^-, \bar{X}_t^-\rangle\Big) = \frac{1}{M}(\bar{Q}_1 - \bar{Q}_2),$$
$$E(\boldsymbol{A_3} + \boldsymbol{B_2}) = \frac{1}{M}(E\bar{Q}_1 - E\bar{Q}_2).$$

Denote $s = \frac{m-l}{m}x_0 - t - \boldsymbol{A_2} - \boldsymbol{A_4} - E(\boldsymbol{A_3} + \boldsymbol{B_2})$, we estimate $E(\boldsymbol{A_3} + \boldsymbol{B_2})$, $\boldsymbol{A_2}$ and $\boldsymbol{A_4}$ in the following. Given $z^*_s = b$ and the result in (7), we have for some constant $c > 0$ and $q = 1, \cdots, K$:

$$E\Big[\{Y^t_{is}\log\frac{\mu_{qa}}{\mu_{qb}} + (1 - Y^t_{is})\log\frac{1-\mu_{qa}}{1-\mu_{qb}}\}\Big] = \mu_{qb}\log\frac{\mu_{qa}}{\mu_{qb}} + (1-\mu_{qb})\log\frac{1-\mu_{qa}}{1-\mu_{qb}} < -c < 0.$$

Then

$$E\boldsymbol{A_3} = \frac{1}{M}\sum_{t=1}^{M}\sum_{q=1}^{K}\sum_{i\neq s}^{N}\Big[z^*_{iq}\{\mu_{qb}\log\frac{\mu_{qa}}{\mu_{qb}} + (1-\mu_{qb})\log\frac{1-\mu_{qa}}{1-\mu_{qb}}\}\Big] < -c^*\eta_N(N-1).$$

Given edges from different communities are independent and correlation density λ, there exists a constant $C > 0$ such that

$$
\begin{aligned}
E\boldsymbol{B_2} =& \frac{\rho C_A}{4}\Big[\langle \alpha_a \otimes vec(\alpha_a^T\alpha_a), E\{\hat{Y}_{\cdot s}^{t,a} \otimes vec(\hat{\boldsymbol{Y}}^{t,a})\}\rangle - \langle \alpha_b \otimes vec(\alpha_b^T\alpha_b), E\{\hat{Y}_{\cdot i}^{t,b} \otimes vec(\hat{\boldsymbol{Y}}^{t,b})\}\rangle\Big] \\
=& -\frac{\rho C_A}{4}\langle \alpha_b \otimes vec(\alpha_b^T\alpha_b), E\{\hat{Y}_{\cdot i}^{t,b} \otimes vec(\hat{\boldsymbol{Y}}^{t,b})\}\rangle \leq -C\lambda N_b^3.
\end{aligned}
$$

Therefore, $-E(\boldsymbol{A_3}+\boldsymbol{B_2}) \geq c'(c^*\eta_N N + \lambda N^3)$ for some positive constant c'. Based on condition C1 that $\mu_{ql}, q, l = 1, \cdots, K$ are bounded, it can be shown that $|EY_{is}^t \log\frac{\mu_{qa}}{\mu_{qb}} + (1-EY_{is}^t)\log\frac{1-\mu_{qa}}{1-\mu_{qb}}|$ is bounded then $|\boldsymbol{A_2}| = \mathcal{O}_N(N)$. From condition (C5), we have

$$
\log\frac{\gamma_1}{\gamma_2} \leq \log\frac{\mu_{qa}\hat{\mu}_{qb}}{\hat{\mu}_{qa}\mu_{qb}} \leq \log\frac{\gamma_2}{\gamma_1} \quad \text{and} \quad \log\frac{1-\gamma_2}{1-\gamma_1} \leq \log\frac{(1-\mu_{qa})(1-\hat{\mu}_{qb})}{(1-\hat{\mu}_{qa})(1-\hat{\mu}_{qb})} \leq \log\frac{1-\gamma_1}{1-\gamma_2}
$$

Define $\gamma = \max\{-\log\frac{\gamma_1}{\gamma_2}, \frac{\gamma_2}{\gamma_1}, -\frac{1-\gamma_2}{1-\gamma_1}, \frac{1-\gamma_1}{1-\gamma_2}\}$. Then we have

$$
\begin{aligned}
|\boldsymbol{A_4}| =& |\frac{1}{M}\sum_{t=1}^{M}\sum_{q=1}^{K}\sum_{i\neq s}^{N}\Big[\alpha_{iq}\{Y_{is}^t \log\frac{\mu_{qa}\hat{\mu}_{qb}}{\hat{\mu}_{qa}\mu_{qb}} + (1-Y_{is}^t)\log\frac{(1-\mu_{qa})(1-\hat{\mu}_{qb})}{(1-\hat{\mu}_{qa})(1-\hat{\mu}_{qb})}\}\Big]| \\
\leq& \gamma|\sum_{q=1}^{K}\sum_{i\neq s}^{N}\alpha_{iq}| \leq \gamma N
\end{aligned}
$$

Therefore we have $|\boldsymbol{A_2}+\boldsymbol{A_4}| = \mathcal{O}_N(N)$. We choose $t = -\frac{E(\boldsymbol{A_3}+\boldsymbol{B_2})}{2}$ and $x_0 = -\alpha' MN$ where $\alpha' > 0$. As the function of node size N, M and λ are constrained in the range $M \leq o(N^{2-\frac{\eta}{2}})$ and $\lambda N^{\frac{\eta}{2}} > 1$, where η is defined in condition C4. Then $\frac{m-l}{m}x_0 = o_N(E(\boldsymbol{A_3}+\boldsymbol{B_2}))$. Therefore, $E(\boldsymbol{A_3}+\boldsymbol{B_2})$ is dominant term in s such that $s \geq -C'\lambda N^3$ where $C' > 0$ is a constant. Follow a similar discussion in (10) and condition C3, we have the upper bound for $\|\bar{\Sigma}_1\|_{op}$:

$$
\|\bar{\Sigma}_1\|_{op} \leq c_0(1 + c_1\lambda N^2).
$$

In addition, from $\#|X_t^+| = N$, $\#|\bar{\xi}_a| = o(N_a^3)$, $\#|\bar{\xi}_b| = o(N_b^3)$ and condition (C3), we have the upper bound for $\|\bar{\Sigma}_1\|_F^2$:

$$
\|\bar{\Sigma}_1\|_F^2 \leq C_1 N(1 + c_1\lambda N^2) + C_2 N^3(1 + c_2\lambda N^2),
$$

where C_1, C_2, c_1, c_2 are constants. Then we estimate the upper bound for the second term in (17) following the similar decentralized quadratic decomposition in Theorem 2.5.1 and Theorem 2.5.3:

$$\mathbb{P}\Big(\boldsymbol{A_3}+\boldsymbol{B_2}\geq\frac{m-l}{m}x_0-t-\boldsymbol{A_2}-\boldsymbol{A_4}\Big)=\mathbb{P}\Big\{(\bar{Q}_1-E\bar{Q}_1)-(\bar{Q}_2-E\bar{Q}_2)>Ms\Big\}$$
$$\leq\frac{1}{2}\mathbb{P}\Big\{|\bar{Q}_1-E\bar{Q}_1|>\frac{Ms}{2}\Big\}+\frac{1}{2}\mathbb{P}\Big\{|\bar{Q}_2-E\bar{Q}_2|>\frac{Ms}{2}\Big\}.$$

According to the generalized Hanson-Wright inequality in (Chen and Yang (2018)):

$$\frac{1}{2}\mathbb{P}\Big\{|\bar{Q}_1-E\bar{Q}_1|>s\Big\}\leq\exp\Big\{-C\min\Big(\frac{s^2M^2}{\bar{L}_1^4\|\bar{\Sigma}_1\|_F^2\|A\|_F^2},\frac{sM}{\bar{L}_1^2\|\bar{\Sigma}_1\|_{op}\|A\|_{op}}\Big)\Big\},\tag{18}$$

where $A=\boldsymbol{I}_{M\times M}$ and $\bar{L}_1$ is subgaussian norm of $\bar{X}_t^+$. Notice that

$$\frac{s^2M^2}{\bar{L}_1^4\|\bar{\Sigma}_1\|_F^2\|A\|_F^2}\geq\frac{(C'\lambda N^3)^2M^2}{\bar{L}_1^4\{C_1N(1+c_1\lambda N^2)+C_2N^3(1+c_2\lambda N^2)\}M}.$$

$$\frac{sM}{\bar{L}_1^2\|\bar{\Sigma}_1\|_{op}\|A\|_{op}}\geq\frac{C'\lambda N^3M}{\bar{L}_1^2c_0(1+c_3\lambda N^2)}.$$

Given $\lambda N^{\frac{\eta}{2}}>1$, we have for some constant $C^*>0$

$$C\min\Big(\frac{s^2M^2}{\bar{L}_1^4\|\bar{\Sigma}_1\|_F^2\|A\|_F^2},\frac{sM}{\bar{L}_1^2\|\bar{\Sigma}_1\|_{op}\|A\|_{op}}\Big)\geq C^*\lambda MN.$$

The upper bound for $\mathbb{P}\Big\{|\bar{Q}_2-E\bar{Q}_2|>\frac{Ms}{2}\Big\}$ can be similarly obtained. Therefore,

$$\mathbb{P}\Big(\boldsymbol{A_3}+\boldsymbol{B_2}\geq\frac{m-l}{m}x_0-t-\boldsymbol{A_2}\Big)\leq\exp(-C'\lambda MN).$$

Next, we estimate the term $\mathbb{P}\Big(\boldsymbol{A_1}+\boldsymbol{B_1}\geq t\Big)$. Notice

$$E(\boldsymbol{A_1})=E\Big[\frac{1}{M}\sum_{t=1}^{M}\sum_{q=1}^{K}\sum_{i\neq s}^{N}(\alpha_{iq}-z_{iq}^*)\{Y_{is}^t-E(Y_{is}^t)\}(\log\frac{\mu_{qa}}{\mu_{qb}}-\log\frac{1-\mu_{qa}}{1-\mu_{qb}})\Big]=0,$$
$$E(\boldsymbol{B_1})=\frac{\rho C_A}{4M}\sum_{t=1}^{M}\Big[\langle\alpha_a\otimes vec(\alpha_a^T\alpha_a)-z_a^*\otimes vec(z_a^{*T}z_a^*),E\{\hat{Y}_{\cdot s}^{t,a}\otimes vec(\hat{\boldsymbol{Y}}^{t,a})\}\rangle-$$
$$\langle\alpha_b\otimes vec(\alpha_b^T\alpha_b)-z_b^*\otimes vec(z_b^{*T}z_b^*),E\{\hat{Y}_{\cdot s}^{t,b}\otimes vec(\hat{\boldsymbol{Y}}^{t,b})\}\rangle\Big].$$

Given condition (C4) such that $\|\boldsymbol{\alpha} - z^*\|_1 = cN^{1-\eta}, 0 < \eta < 1$,

$$\boldsymbol{B}_1 = \frac{\rho C_A}{4M} \sum_{t=1}^{M} \Big\{ \langle \alpha_a \otimes vec(\alpha_a^T \alpha_a) - z_a^* \otimes vec(z_a^{*T} z_a^*), \hat{Y}_{\cdot s}^a \otimes vec(\hat{\mathbf{Y}}^a) \rangle - \\ \langle \alpha_b \otimes vec(\alpha_b^T \alpha_b) - z_b^* \otimes vec(z_b^{*T} z_b^*), \hat{Y}_{\cdot s}^b \otimes vec(\hat{\mathbf{Y}}^b) \rangle \Big\}.$$

Notice that for any community $a = 1, \cdots, K$,

$$\begin{aligned} \|(vec(\alpha_a^T \alpha_a) - vec(z_a^{*T} z_a^*))\|_2 &\leq \|\alpha_a \otimes (\alpha_a - z_a^*)\|_2 + \|(\alpha_a - z_a^*) \otimes z_a^*\|_2 \\ &\leq \|\alpha_a\|_2 \|(\alpha_a - z_a^*)\|_2 + \|(\alpha_a - Z_a^*)\|_2 \|z_a^*\|_2, \\ \|E(\hat{Y}_{\cdot s}^{t,a})\|_2 \leq \sqrt{\frac{N}{\hat{\mu}_{aa}(1-\hat{\mu}_{aa})}}, &\quad \|E(\hat{\mathbf{Y}}^{t,a})\|_2 \leq \sqrt{\frac{N^2}{\hat{\mu}_{aa}(1-\hat{\mu}_{aa})}}. \end{aligned}$$

Therefore, we have

$$\begin{aligned} &\langle \alpha_a \otimes vec(\alpha_a^T \alpha_a) - z_a^* \otimes vec(z_a^{*T} z_a^*), E\{\hat{Y}_{\cdot s}^{t,a} \otimes vec(\hat{\mathbf{Y}}^{t,a})\} \rangle \\ \leq &\|\alpha_a \otimes vec(\alpha_a^T \alpha_a) - z_a^* \otimes vec(z_a^{*T} z_a^*)\|_2 \|E\{\hat{Y}_{\cdot s}^{t,a} \otimes vec(\hat{\mathbf{Y}}^{t,a})\}\|_2 \\ \leq &\Big(\|\alpha_a \otimes vec(\alpha_a^T \alpha_a) - vec(z_a^{*T} z_a^*))\|_2 + \|(\alpha_a - z_a^*) \otimes vec(z_a^{*T} z_a^*)\|_2 \Big) \|E\{\hat{Y}_{\cdot s}^{t,a} \otimes vec(\hat{\mathbf{Y}}^{t,a})\}\|_2 \\ \leq &\|\alpha_a - z_a^*\|_2 \cdot \Big(\|\alpha_a\|_2^2 + \|z_a^*\|_2^2 + \|\alpha_a\|_2 \|z_a^*\|_2 \Big) \cdot \|E(\hat{Y}_{\cdot s}^{t,a})\|_2 \cdot \|E(\hat{\mathbf{Y}}^{t,a})\|_2 \leq \frac{3N * N^{3/2}}{\hat{\mu}_{aa}(1-\hat{\mu}_{aa})} \|\alpha_a - z_a^*\|_2. \end{aligned}$$

Since $\|\alpha_a - z_a^*\|_2 = \sqrt{\|\alpha_a - z_a^*\|_2^2} \leq \sqrt{\|\boldsymbol{\alpha} - z^*\|_1}$ for any $a = 1, \cdots, K$, then for some constant $C > 0$,

$$|E(\boldsymbol{B}_1)| \leq CN^{3-\frac{\eta}{2}}.$$

We define edge vectors $\tilde{Y}_t, t = 1, \cdots, M$ and membership vector $\boldsymbol{\theta}_{a,b}$ as:

$$\begin{aligned} \tilde{Y}_t &= \Big\{ \underbrace{Y_{\cdot s}^t - E(Y_{\cdot s}^t), \cdots, Y_{\cdot s}^t - E(Y_{\cdot s}^t)}_{NK}, \hat{Y}_{\cdot s}^{t,a} \otimes vec(\hat{\mathbf{Y}}^{t,a}), \hat{Y}_{\cdot s}^{t,b} \otimes vec(\hat{\mathbf{Y}}^{t,b}) \Big\}, \\ \boldsymbol{\theta}_{a,b} &= \Big[\underbrace{(\alpha_{iq} - z_{iq}^*)(\log \frac{\mu_{qa}}{\mu_{qb}} - \log \frac{1-\mu_{qa}}{1-\mu_{qb}}}_{i=1\cdots,N}), \cdots, \underbrace{(\alpha_{iK} - z_{iK}^*)(\log \frac{\mu_{Ka}}{\mu_{Kb}} - \log \frac{1-\mu_{Ka}}{1-\mu_{Kb}}),}_{i=1\cdots,N} \\ &\frac{\rho C_A}{4} \{\alpha_a \otimes vec(\alpha_a^T \alpha_a) - z_a^* \otimes vec(z_a^{*T} z_a^*)\}, \frac{\rho C_A}{4} \{\alpha_b \otimes vec(\alpha_b^T \alpha_b) - z_b^* \otimes vec(z_b^{*T} z_b^*)\} \Big]. \end{aligned}$$

Notice for $a, b = 1, \cdots, K$, we have

$$\begin{aligned}\|\boldsymbol{\theta}_{a,b}\|_2^2 &\leq \mu_2\|\boldsymbol{\alpha} - z^*\|_2^2 + \|\alpha_a \otimes vec(\alpha_a^T\alpha_a) - z_a^* \otimes vec(z_a^{*T} z_a^*)\|_2^2 + \|\alpha_b \otimes vec(\alpha_b^T\alpha_b) - z_b^* \otimes vec(z_b^{*T} z_b^*)\|_2^2 \\ &\leq \mu_2\|\boldsymbol{\alpha} - z^*\|_1 + C_1 N^2(\|\alpha_a - z_a^*\|_1 + \|\alpha_b - z_b^*\|_1),\end{aligned}$$

where $\mu_2 := \max\{(\log\frac{\mu_{qa}}{\mu_{qb}} - \log\frac{1-\mu_{qa}}{1-\mu_{qb}})\}$, $q = 1, \cdots, K$ and $C_1 > 0$ is a constant. Then we can transform $\mathrm{var}(\boldsymbol{A_1} + \boldsymbol{B_1})$ into

$$\mathrm{var}(\boldsymbol{A_1} + \boldsymbol{B_1}) = \frac{1}{M}\sum_{t=1}^{M}\mathrm{var}(\boldsymbol{\theta}_{a,b}\tilde{Y}_t) = \frac{1}{M}\sum_{t=1}^{M}\boldsymbol{\theta}_{a,b}^T Cov(\tilde{Y}_t, \tilde{Y}_t)\boldsymbol{\theta}_{a,b} \leq \frac{1}{M}\|Cov(\tilde{Y}_t, \tilde{Y}_t)\|_{op}\|\boldsymbol{\theta}_{a,b}\|_2^2.$$

From the condition (C3) and same discussion in (10), we have for some constant $C > 0$ and $c > 0$:

$$\|Cov(\tilde{Y}_t, \tilde{Y}_t)\|_{op} \leq C(1 + c\lambda N^2).$$

Given $\frac{1}{\lambda} = o(N^{\frac{\eta}{2}})$, we have $E(\boldsymbol{A_1} + \boldsymbol{B_1}) = o_N(E(\boldsymbol{A_3} + \boldsymbol{B_2}))$ then the $E(\boldsymbol{A_3} + \boldsymbol{B_2})$ is dominating in the term $\{t - E(\boldsymbol{A_1} + \boldsymbol{B_1})\}^2$. Based on the Markov inequality, for some constant $C_2 > 0$

$$\begin{aligned}\mathbb{P}\Big(\boldsymbol{A_1} + \boldsymbol{B_1} \geq t\Big) \leq& \frac{\mathrm{var}(\boldsymbol{A_1} + \boldsymbol{B_1})}{\{t - E(\boldsymbol{A_1} + \boldsymbol{B_1})\}^2} \leq \frac{\|Cov(\tilde{Y}_t, \tilde{Y}_t)\|_{op}\|\boldsymbol{\theta}_{a,b}\|_2^2}{M\{c'(N + \lambda N^3)\}^2} \\ \leq& \frac{C(1 + c\lambda N^2)\{\mu_2\|\boldsymbol{\alpha} - z^*\|_1 + C_1N^2(\|\alpha_a - z_a^*\|_1 + \|\alpha_b - z_b^*\|_1)\}}{(c'(N + \lambda N^3))^2 M} \\ \leq& \frac{2Cc\{\mu_2\|\boldsymbol{\alpha} - z^*\|_1 + C_1N^2(\|\alpha_a - z_a^*\|_1 + \|\alpha_b - z_b^*\|_1)\}}{c'^2(1 + \sqrt{\lambda}N^2)^2 M} \\ \leq& C_2\frac{N^{\eta/4}(\|\alpha_a - z_a^*\|_1 + \|\alpha_b - z_b^*\|_1)}{(1 + \lambda N^{2+\frac{\eta}{4}})M}.\end{aligned}$$

Combined upper bound of $\mathbb{P}\Big(\boldsymbol{A_1} + \boldsymbol{B_1} \geq t\Big)$ and $\mathbb{P}\Big(\boldsymbol{A_3} + \boldsymbol{B_2} \geq s\Big)$ with (16), there exists positive constant $c_1 > 0, c_2 > 0, c_3 > 0$ such that:

$$\begin{aligned}&E\|\boldsymbol{\alpha}^{s+1} - \boldsymbol{z^*}\|_1 \leq 2\boldsymbol{w}NK\exp(-\alpha' MN) + 2\boldsymbol{w}\sum_{l=0}^{m-1}\sum_{a=1}^{K}\sum_{b\neq a}\sum_{i:z_i^*=b}\frac{1 - exp(\alpha' MN)}{m}E(\boldsymbol{L}_2) \\ \leq& 2\boldsymbol{w}KN\exp(-\alpha' MN) + 2\boldsymbol{w}mKN\exp(-C'\lambda MN) + 2\boldsymbol{w}mKNC_2\frac{N^{\eta/4}(\|\alpha_a - z_a^*\|_1 + \|\alpha_b\ ! - z_b^*\|_1)}{(1 + \lambda N^{2+\frac{\eta}{4}})M} \\ \leq& c_1NK\exp(-c_2\alpha' MN) + \frac{c_3N^{1+\frac{\eta}{4}}\|\boldsymbol{\alpha}^s - \boldsymbol{z^*}\|_1}{(1 + \lambda N^{2+\frac{\eta}{4}})M}.\end{aligned}$$

References

Chen, X. and Yang, Y. (2018), "Hanson-Wright inequality in Hilbert spaces with application to K-means clustering for non-Euclidean data," *arXiv preprint arXiv:1810.11180*.