SUPPLEMENT TO “THEORETICAL AND COMPUTATIONAL GUARANTEES OF MEAN FIELD VARIATIONAL INFERENCE FOR COMMUNITY DETECTION”

BY Anderson Y. Zhang and Harrison H. Zhou

University of Pennsylvania and Yale University

APPENDIX A: ADDITIONAL ALGORITHMS

In this section, we provide the detailed implementations of the batched Gibbs sampling and the iterative algorithm of MLE for community detection.

A.1. Batched Gibbs Sampling.

Algorithm 2: Batched Gibbs Sampling

Input: Adjacency matrix A, number of communities k, hyperparameters $\pi^{\text{pri}}, \alpha_p^{\text{pri}}, \beta_p^{\text{pri}}, \alpha_q^{\text{pri}}, \beta_q^{\text{pri}}$, some initializers $Z^{(0)}$, number of iterations S.

Output: Gibbs sampling $\hat{Z}, \hat{p}, \hat{q}$.

for $s = 1, 2, \ldots, S$ **do**

Update $\alpha_p^{(s)}, \beta_p^{(s)}, \alpha_q^{(s)}, \beta_q^{(s)}$ by

$$\alpha_p^{(s)} = \alpha_p^{\text{pri}} + \sum_{a=1}^{k}\sum_{i<j} A_{i,j} Z_{i,a}^{(s-1)} Z_{j,a}^{(s-1)}, \beta_p^{(s)} = \beta_p^{\text{pri}} + \sum_{a=1}^{k}\sum_{i<j}(1-A_{i,j}) Z_{i,a}^{(s-1)} Z_{j,a}^{(s-1)},$$
$$\alpha_q^{(s)} = \alpha_q^{\text{pri}} + \sum_{a\neq b}\sum_{i<j} A_{i,j} Z_{i,a}^{(s-1)} Z_{j,b}^{(s-1)}, \beta_q^{(s)} = \beta_q^{\text{pri}} + \sum_{a\neq b}\sum_{i<j}(1-A_{i,j}) Z_{i,a}^{(s-1)} Z_{j,b}^{(s-1)}.$$

Then generate $p^{(s)} \sim \text{Beta}(\alpha_p^{(s)}, \beta_p^{(s)})$ and $q^{(s)} \sim \text{Beta}(\alpha_q^{(s)}, \beta_q^{(s)})$ independently.

Define

$$t^{(s)} = \frac{1}{2}\log\frac{p^{(s)}(1-q^{(s)})}{(1-p^{(s)})q^{(s)}}, \quad \text{and } \lambda^{(s)} = \frac{1}{2t^{(s)}}\log\frac{1-q^{(s)}}{1-p^{(s)}}.$$

Then update $\pi^{(s)}$ with

$$\pi^{(s)} = h_{t^{(s)},\lambda^{(s)}}(Z^{(s-1)}),$$

where $h_{t,\lambda}(\cdot)$ is defined as in Equation (11). Independently generate each row of $Z^{(s)}$ from distributions

$$\mathbb{P}(Z_{i,\cdot}^{(s)} = e_a) = \pi_{i,a}^{(s)}, \forall a \in [k], \forall i \in [n].$$

end

We have $\hat{z} = z^{(S)}$, $\hat{p} = p^{(S)}$ and $\hat{q} = q^{(S)}$.

A.2. An Iterative Algorithm for Maximum Likelihood Estimation. We first define a mapping $h' : \Pi_0 \to \Pi_0$ as follows

$$[h'_\lambda(Z)]_{i,a} = \mathbb{I}\left[a = \arg\max_b \sum_{j\neq i} Z_{i,b}(A_{i,j} - \lambda)\right]. \tag{38}$$

Here if the maximizer is not unique, we simply pick the smallest index.

Algorithm 3: An Iterative Algorithm for MLE

Input: Adjacency matrix A, number of communities k, some initializers $z^{(0)}$, number of iterations S.

Output: Estimation $\hat{Z}, \hat{p}, \hat{q}$.

for $s = 1, 2, \dots, S$ **do**

1 Update $p^{(s)}, q^{(s)}$ by

$$p^{(s)} = \frac{\sum_{a=1}^k \sum_{i<j} A_{i,j} Z_{i,a}^{(s-1)} Z_{j,a}^{(s-1)}}{\sum_{a=1}^k \sum_{i<j} Z_{i,a}^{(s-1)} Z_{j,a}^{(s-1)}}$$

and

$$q^{(s)} = \frac{\sum_{a\neq b} \sum_{i<j} A_{i,j} Z_{i,a}^{(s-1)} Z_{j,b}^{(s-1)}}{\sum_{a\neq b} \sum_{i<j} Z_{i,a}^{(s-1)} Z_{j,b}^{(s-1)}}.$$

2 Define

$$t^{(s)} = \frac{1}{2} \log \frac{p^{(s)}(1-q^{(s)})}{(1-p^{(s)})q^{(s)}}, \quad \text{and } \lambda^{(s)} = \frac{1}{2t^{(s)}} \log \frac{1-q^{(s)}}{1-p^{(s)}}.$$

Then update $\pi^{(s)}$ with

$$Z^{(s)} = h'_{\lambda^{(s)}}(Z^{(s-1)}),$$

where $h'_\lambda(\cdot)$ is defined as in Equation (38).

end

3 We have $\hat{z} = z^{(S)}$, $\hat{p} = p^{(S)}$ and $\hat{q} = q^{(S)}$.

APPENDIX B: PROOFS OF THEOREMS

In this section, we first establish upper bounds on L_1^{sum} and L_2^{sum} which are needed in the proof of Theorem 6.1 in Section 6.3.2. Then we validate Theorem 3.1 where $\ell(\pi^{(0)}, \pi^*)$ is in a constant order of $\bar{n}_{\min}$, which complements the proof of Theorem 3.1 presented in Section 6.3. In addition, we give proofs of theorems stated in Section 4, including Theorem 4.1, Theorem 4.2 and Theorem 4.3.

B.1. Bounds on L_1^{sum} and L_2^{sum} for the Proof of Theorem 6.1. In this section, we establish upper bounds on L_1^{sum} and L_2^{sum} that are used in the proof of Theorem 6.1 in Section 6.3.2, i.e., Equations (36) and (37). Recall the definition of $\theta_{a,b}$ as in Equation (35). We have some properties on $\theta_{a,b}$ which will be useful for the upcoming analysis:

$$\|\theta_{a,b}\|_\infty \le 2 \tag{39}$$

$$\|\theta_{a,b}\|_1 \le \left\|\pi_{\cdot,a} - Z^*_{\cdot,a}\right\|_1 + \left\|\pi_{\cdot,b} - Z^*_{\cdot,b}\right\|_1 \le \|\pi - Z^*\|_1 \le \gamma \bar{n}_{\min}, \tag{40}$$

$$\text{and } \sum_{a=1}^{k}\sum_{b\neq a}\|\theta_{a,b}\|_1 \le 2k \sum_{a}\left\|\pi_{\cdot,a} - Z^*_{\cdot,a}\right\|_1 \le 2k\|\pi - Z^*\|_1 . \tag{41}$$

1. Bounds on L_1^{sum}. For any $i \in [n]$ such that $z_i = b$, we define

$$L_{1,i}(a,b,l) \triangleq \mathbb{I}\left[S^{(1)}_{i,a,b} \ge -\frac{(l+3/2)(n_a+n_b)I}{4mt} - S^{(3)}_{i,a,b}\right], \tag{42}$$

and $L'_{1,i}(a,b,l) \triangleq \exp(-l(n_a+n_b)I/(2m))L_{1,i}(a,b,l)$ so that

$$L_1^{\text{sum}} = \sum_{l=0}^{m-1}\sum_{a=1}^{k}\sum_{b\neq a}\sum_{i:z_i=b} L'_{1,i}(a,b,l).$$

We are going to obtain $\mathbb{E}L_1^{\text{sum}}$. By applying Markov inequality, we have

$$\mathbb{E}L_{1,i}(a,b,l) = \mathbb{P}\left[t^* S^{(1)}_{i,a,b} \ge -\frac{t^*(l+3/2)(n_a+n_b)I}{4mt} - t^* S^{(3)}_{i,a,b}\right]$$
$$\le \exp\left[\frac{t^*(l+3/2)(n_a+n_b)I}{4mt} + t^*\sum_{j\neq i}(\mathbb{E}A_{i,j}-\lambda)[\theta_{a,b}]_j\right]\mathbb{E}\exp\left[t^*\sum_{j\neq i}(Z^*_{j,a}-Z^*_{j,b})(A_{i,j}-\lambda)\right].$$

Let $X \sim \text{Ber}(q^*)$ and $Y \sim \text{Ber}(p^*)$. Recall that $z_i = b$. We have $|\{j \neq i : z_j = a\}| = n_a$ and $|\{j \neq i : z_j = b\}| = n_b - 1 \coloneqq n'_b$. Due to the underlying SBM structure, $\{A_{i,j}\}_{j:z_j=a}, \{A_{i,j}\}_{j:z_j=b}$ are independent and identical copies of X and Y, respectively. Thus,

$$\mathbb{E}\exp\left[t^*\sum_{j\neq i}(Z^*_{j,a}-Z^*_{j,b})(A_{i,j}-\lambda)\right] = \exp\left[-t^*\lambda(n_a-(n_b-1))\right]\prod_{j\neq i}\mathbb{E}\exp(t^*(Z^*_{j,a}-Z^*_{j,b})A_{i,j})$$
$$= \exp(-t^*\lambda(n_a-n'_b))\left[\mathbb{E}\exp(t^*X)\right]^{n_a}\left[\mathbb{E}\exp(-t^*Y)\right]^{n'_b}.$$

From Proposition C.1, we have $\mathbb{E}e^{t^*X}/\mathbb{E}e^{-t^*Y} = e^{t^*\lambda^*}$ and $\mathbb{E}e^{t^*X}\mathbb{E}e^{-t^*Y} = \exp(-I)$. This leads to

$$\begin{aligned}
&\mathbb{E}\exp\left[t^*\sum_{j\neq i}(Z^*_{j,a}-Z^*_{j,b})(A_{i,j}-\lambda)\right] \\
&= \exp(-t^*(\lambda-\lambda^*)(n_a-n'_b))\left[e^{-t^*\lambda^*}\frac{\mathbb{E}e^{t^*X}}{\mathbb{E}e^{-t^*Y}}\right]^{\frac{n_a-n'_b}{2}}\left[\mathbb{E}e^{t^*X}\mathbb{E}e^{-t^*Y}\right]^{\frac{n_a+n'_b}{2}} \\
&= \exp(-t^*(\lambda-\lambda^*)(n_a-n'_b))\exp\left[-\frac{(n_a+n'_b)I}{2}\right].
\end{aligned}$$

Therefore, the logarithm of $\mathbb{E}L'_{1,i}(a,b,l)$ is upper bounded by

$$\begin{aligned}
\log\left[\mathbb{E}L'_{1,i}(a,b,l)\right] &= -\frac{l(n_a+n_b)I}{2m}+\log\left[\mathbb{E}L_{1,i}(a,b,l)\right] \\
&\leq -\frac{l(n_a+n_b)I}{2m}+\frac{t^*(l+3/2)(n_a+n_b)I}{4mt}+t^*\sum_{j\neq i}(\mathbb{E}A_{i,j}-\lambda)[\theta_{a,b}]_j \\
&\quad - t^*(\lambda-\lambda^*)(n_a-n'_b)-\frac{(n_a+n'_b)I}{2}.
\end{aligned}$$

After combing like terms, we obtain

$$\begin{aligned}
\log\left[\mathbb{E}L'_{1,i}(a,b,l)\right] \leq& -\frac{(1+\frac{l}{m}-\frac{t^*(l+3/2)}{2mt})(n_a+n'_b)I}{2} \\
&- t^*(\lambda-\lambda^*)(n_a-n'_b)+t^*\sum_{j\neq i}(\mathbb{E}A_{i,j}-\lambda)[\theta_{a,b}]_j
\end{aligned} \tag{43}$$

We are going to show $\log[\mathbb{E}L'_{1,i}(a,b,l)] \leq -(1-\eta'')\bar{n}_{\min}I$ by some $\eta'' = o(1)$. We first present some properties of λ^*, t^* and I that will be helpful:

$$I \asymp (p^*-q^*)^2/p^*, \tag{44}$$

$$\lambda^* \in (q^*, p^*), \tag{45}$$

$$\text{and } t^* \asymp (p^*-q^*)/p^*. \tag{46}$$

Here Equations (44) and (45) are proved by Propositions C.2 and C.3 respectively. Equation (46) is due to $t^* \asymp \log(1+(p^*-q^*)/q^*) \asymp (p^*-q^*)/p^*$ under the assumption that $p^*, q^* = o(1)$, $p^* \asymp q^*$.

The first term on the RHS of Equation (43) is upper bounded by $-(1-7/(8m))\bar{n}_{\min}I$ by the assumption $t^*/t = 1+o(1)$. Recall we assume $|t^*(\lambda-\lambda^*)| \leq \eta' t^*(p^*-q^*)$. By Equations (44) and (46) the second term is upper bounded by $\eta' \bar{n}_{\min} I$ up to a constant factor. For the last term on the RHS of Equation (43), since $|\lambda - \lambda^*| \leq \eta'(p^* - q^*)$ we have

$$\begin{aligned} t^* \left| \sum_{j \neq i} (\mathbb{E} A_{i,j} - \lambda)[\theta_{a,b}]_i \right| &\leq t^* \left| \sum_{j \neq i} (\mathbb{E} A_{i,j} - \lambda^*)[\theta_{a,b}]_i \right| + t^* \left| \sum_{j \neq i} (\lambda^* - \lambda)[\theta_{a,b}]_i \right| \\ &\leq (1 + \eta') t^* (p^* - q^*) \left\| \theta_{a,b} \right\|_1 \\ &\leq (1 + \eta') t^* (p^* - q^*) \gamma \bar{n}_{\min} \\ &\lesssim \gamma \bar{n}_{\min} I, \end{aligned}$$

where we use Equations (40) and (44) - (46).

As a consequence, there exists a sequence $\eta'' = o(1)$ that goes to zero slower than m^{-1}, γ, η', such that the summation of three terms on the RHS of Equation (43) is upper bounded by $-(1 - \eta'')\bar{n}_{\min} I$. Thus, from Equation (42) we have

$$\begin{aligned} \mathbb{E} L_1^{\text{sum}} &= \sum_{l=0}^{m-1} \sum_{a=1}^{k} \sum_{b \neq a} \sum_{i: z_i = b} \exp \left[\log \mathbb{E} L'_{1,i}(a, b, l) \right] \\ &\leq \sum_{l=0}^{m-1} \sum_{a=1}^{k} \sum_{b \neq a} \sum_{i: z_i = b} \exp(-(1 - \eta'')\bar{n}_{\min} I) \\ &\leq nmk \exp \left[-(1 - \eta'')\bar{n}_{\min} I \right]. \end{aligned}$$

Since η'' goes to 0 slower than m^{-1}, we have $\eta'' \geq m^{-1} \geq (\bar{n}_{\min} I)^{\frac{1}{4}}$ by Equation (34). Then by applying Markov inequality, we have

$$\mathbb{P} \left[L_1^{\text{sum}} \geq nmk \exp \left[-(1 - 2\eta'')\bar{n}_{\min} I \right] \right] \leq \exp \left[-\eta'' \bar{n}_{\min} I \right] \leq \exp \left[-2(\bar{n}_{\min} I)^{\frac{1}{2}} \right].$$

That is, with probability at least $1 - \exp[-2(\bar{n}_{\min} I)^{\frac{1}{2}}]$, Equation (36) holds.

2. Bounds on L_2^{sum}. Recall the definition of L_2^{sum} as

$$L_2^{\text{sum}} \triangleq \sum_{a=1}^{k} \sum_{b \neq a} \sum_{i: z_i = b} \mathbb{I} \left[A_{i,\cdot} - \mathbb{E} A_{i,\cdot}) \theta_{a,b} \geq \frac{\bar{n}_{\min} I}{4mt} \right].$$

Depending on network being dense or sparse, we consider two scenarios.

(1) Dense Scenario: $q^* \geq (\log n)/n$. In this scenario, we have a sharp

bound on $\|A - \mathbb{E}A\|_{\text{op}}$. First we observe that

$$\begin{aligned}\sum_{i:z_i=b} [(A_{i,\cdot} - \mathbb{E}A_{i,\cdot})\theta_{a,b}]^2 &= \theta_{a,b}^T \sum_{i:z_i=b} [(A_{i,\cdot} - \mathbb{E}A_{i,\cdot})^T (A_{i,\cdot} - \mathbb{E}A_{i,\cdot})]\theta_{a,b} \\ &\leq \theta_{a,b}^T \sum_{i} [(A_{i,\cdot} - \mathbb{E}A_{i,\cdot})^T (A_{i,\cdot} - \mathbb{E}A_{i,\cdot})]\theta_{a,b} \\ &= \theta_{a,b}^T [(A - \mathbb{E}A)^T (A - \mathbb{E}A)]\theta_{a,b}.\end{aligned}$$

By applying Markov inequality, we have

$$L_2^{\text{sum}} \leq \sum_{a=1}^{k} \sum_{b \neq a} \frac{\theta_{a,b}^T [(A - \mathbb{E}A)^T (A - \mathbb{E}A)]\theta_{a,b}}{(\bar{n}_{\min} I/(4mt))^2}.$$

Since $\|\theta_{a,b}\|_\infty \leq 2$, we have $\|\theta_{a,b}\|^2 \leq 2\|\theta_{a,b}\|_1$. Lemma C.3 shows $\|A - \mathbb{E}A\|_{\text{op}} \leq \sqrt{c_1 np}$ holds with probability at least $1 - n^{-r}$ for some constants $c_1, r > 0$. Together with Equation (41), we have

$$\begin{aligned}\sum_{a=1}^{k} \sum_{b \neq a} \theta_{a,b}^T [(A - \mathbb{E}A)^T (A - \mathbb{E}A)]\theta_{a,b} &\leq \sum_{a=1}^{k} \sum_{b \neq a} \|A - \mathbb{E}A\|_{\text{op}}^2 \|\theta_{a,b}\|^2 \\ &\leq \sum_{a=1}^{k} \sum_{b \neq a} 2c_1 np \|\theta_{a,b}\|_1 \\ &\leq 4c_1 knp \|\pi - Z^*\|_1.\end{aligned}$$

Thus, with probability at least $1 - n^{-r}$,

$$L_2^{\text{sum}} \leq \frac{4c_1 knp \|\pi - Z^*\|_1}{(\bar{n}_{\min} I/(4mt))^2}.$$

(2) Sparse Scenario: $q^* < (\log n)/n$. When the network is sparse, the previous upper bound on $\|A - \mathbb{E}A\|_{\text{op}}$ no longer holds. Instead, removing nodes with large degrees is required to yield provably sharp bound on $\|A - \mathbb{E}A\|_{\text{op}}$. Define $S = \{i \in [n], \sum_j A_{i,j} \geq 20np^*\}$. We define $\tilde{A}, \tilde{P}$ such that $\tilde{A}_{i,j} = A_{i,j}\mathbb{I}\{i, j \notin S\}$ and $\tilde{P}_{i,j} = (\mathbb{E}A_{i,j})\mathbb{I}\{i, j \notin S\}$. Then we have the

decomposition as

$$
\begin{aligned}
L_2(a,b) &\triangleq \sum_{i:z_i=b} \mathbb{I}\left[(A_{i,\cdot} - \mathbb{E}A_{i,\cdot})\theta_{a,b} \geq \frac{\bar{n}_{\min} I}{4mt}\right] \\
&\leq \sum_{i:z_i=b} \mathbb{I}\left[(\tilde{A}_{i,\cdot} - \tilde{P}_{i,\cdot})\theta_{a,b} \geq \frac{\bar{n}_{\min} I}{8mt}\right] \\
&\quad + \sum_{i:z_i=b} \mathbb{I}\left[\sum_{j\neq i}(A_{i,j} - \mathbb{E}A_{i,j})[\theta_{a,b}]_{i,j}\mathbb{I}\{i \in S \text{ or } j \in S\} \geq \frac{\bar{n}_{\min} I}{8mt}\right] \\
&:= L_{2,1}(a,b) + L_{2,2}(a,b).
\end{aligned}
$$

Define $L_{2,1}^{\text{sum}} \triangleq \sum_{a=1}^{k}\sum_{b\neq a} L_{2,1}(a,b)$. We have

$$
L_{2,1}^{\text{sum}} \leq \sum_{a=1}^{k}\sum_{b\neq a} \frac{\theta_{a,b}^T[(\tilde{A}-\tilde{P})^T(\tilde{A}-\tilde{P})]\theta_{a,b}}{(\bar{n}_{\min}I/(8mt))^2} \leq \sum_{a=1}^{k}\sum_{b\neq a} \frac{2\|\tilde{A}-\tilde{P}\|_{\text{op}}^2 \|\theta_{a,b}\|_1}{(\bar{n}_{\min}I/(8mt))^2}.
$$

Lemma C.4 shows $\|\tilde{A}-\tilde{P}\|_{\text{op}} \leq \sqrt{c_2 np}$ holds with probability at least $1-n^{-1}$ for some constant $c_2 > 0$. Then we have

$$
L_{2,1}^{\text{sum}} \leq \frac{4c_2knp\,\|\pi - Z^*\|_1}{(\bar{n}_{\min}I/(8mt))^2}.
$$

Lemma C.5 shows $\sum_{i,j}|A_{i,j} - \mathbb{E}A_{i,j}|\mathbb{I}\{i \in S\} \leq 20n^2p^*\exp(-5np^*)$ holds with probability at least $1-\exp(-5np^*)$. Then by applying Markov inequality, we have

$$
\begin{aligned}
L_{2,2}^{\text{sum}} &\triangleq \sum_{a=1}^{k}\left[\sum_{b\neq a} L_{2,2}(a,b)\right] \\
&\leq \sum_{a=1}^{k}\sum_{i,j=1}^{n} \frac{|A_{i,j} - \mathbb{E}A_{i,j}||[\theta_{a,b}]_{i,j}|\mathbb{I}\{i \in S \text{ or } j \in S\}}{\bar{n}_{\min}I/(8mt)} \\
&\leq \sum_{a=1}^{k} \frac{4\sum_{i,j}|A_{i,j} - \mathbb{E}A_{i,j}|\mathbb{I}\{i \in S\}}{\bar{n}_{\min}I/(8mt)} \\
&\leq \frac{80n^2kp^*\exp(-5np^*)}{\bar{n}_{\min}I/(8mt)}.
\end{aligned}
$$

As a consequence, we have

$$
L_2^{\text{sum}} \leq L_{2,1}^{\text{sum}} + L_{2,2}^{\text{sum}} \leq \frac{4c_2knp^*\,\|\pi - Z^*\|_1}{(\bar{n}_{\min}I/(8mt))^2} + \frac{80n^2kp^*\exp(-5np^*)}{\bar{n}_{\min}I/(8mt)},
$$

with probability at least $1-n^{-1}-\exp(-5np^*)$. By the bounds on L_1^{sum} and L_2^{sum}, and due to $t/t^* = 1+o(1)$, we obtain Equation (37).

B.2. Proof of Theorem 3.1 for the case where $\ell(\pi^{(0)}, \pi^*)$ is in a constant order of $\bar{n}_{\min}$. For any π such that $\ell(\pi,\pi^*) \leq c_{\text{init}}\bar{n}_{\min}$, we are going to show when c_{init} is sufficiently small

$$\ell(h_{t,\lambda}(\pi), Z^*) \leq n\exp(-\bar{n}_{\min}I/25) + \frac{\ell(\pi, Z^*)}{2\sqrt{nI/[wk[n/\bar{n}_{\min}]^2]}}, \tag{47}$$

with probability at least $1-\exp(-\bar{n}_{\min}I/10)-n^{-r}$ for some constant $r>0$. If it holds, for any $\pi^{(0)}$ such that $\ell(\pi^{(0)}, Z^*) = c\bar{n}_{\min}$ for some constant $c \leq c_{\text{init}}$, the term $n\exp(-\bar{n}_{\min}I/25)$ is dominated by $\ell(\pi^{(0)}, Z^*)/\sqrt{nI/[wk[n/\bar{n}_{\min}]^2]}$ which implies

$$\ell(\pi^{(1)}, Z^*) \leq n\exp(-(1-\eta)/\bar{n}_{\min}I) + \frac{\ell(\pi^{(0)}, Z^*)}{\sqrt{nI/[wk[n/\bar{n}_{\min}]^2]}}.$$

It also implies $\ell(\pi^{(1)}, Z^*) = o(\bar{n}_{\min})$, which means after the first iteration, the results in Section 6.3 can be directly applied and the proof is complete.

The proof of Equation (47) mainly follows that of Theorem 6.1. We have

$$\left\|[h_{t,\lambda}(\pi)]_{i,\cdot} - Z^*_{i,\cdot}\right\|_1 \leq 2w\sum_{a\neq z_i} 1\wedge\exp\left[2t\sum_{j\neq i}(\pi_{j,a}-\pi_{j,z_i})(A_{i,j}-\lambda)\right].$$

Note that the inequality $1\wedge\exp(-x) \leq f(x_0) + \mathbb{I}\{x \geq x_0\}$ holds for any $x_0 \geq 0$. By taking $x_0 = (n_a+n_{z_i})I/4$, we have

$$\left\|[h_{t,\lambda}(\pi)]_{i,\cdot} - Z^*_{i,\cdot}\right\|_1 \leq 2w\sum_{a\neq z_i}\left[\exp\left[-\frac{(n_a+n_{z_i})I}{4}\right] + \mathbb{I}\left[\sum_{j\neq i}(\pi_{j,a}-\pi_{j,z_i})(A_{i,j}-\lambda) \geq -\frac{(n_a+n_{z_i})I}{8t}\right]\right],$$

and consequently,

$$\begin{aligned}\|h_{t,\lambda}(\pi) - Z^*\|_1 &\leq 2wnk\exp(-\bar{n}_{\min}I/2) \\ &\quad + 2w\sum_{a=1}^{k}\sum_{b\neq a}\sum_{i:z_i=b}\mathbb{I}\Bigg[\sum_{j\neq i}(\pi_{j,a}-\pi_{j,b})(A_{i,j}-\lambda) \geq -\frac{(n_a+n_{z_i})I}{8t}\Bigg]\Bigg].\end{aligned}$$

Define $\theta_{a,b}$ the same way as in Section 6.3, and by the same argument, we have

$$\|h_{t,\lambda}(\pi) - Z^*\|_1 \leq 2wnk\exp(-\bar{n}_{\min}I/2) + 2w\sum_{a=1}^{k}\sum_{b\neq a}\sum_{i:z_i=b}\mathbb{I}\left[(A_{i,\cdot} - \mathbb{E}A_{i,\cdot})\theta_{a,b} \geq \frac{\bar{n}_{\min}I}{8t}\right]$$

$$+ 2w\sum_{a=1}^{k}\sum_{b\neq a}\sum_{i:z_i=b}\mathbb{I}\Bigg[\sum_{j\neq i}(Z^*_{j,a} - Z^*_{j,b})(A_{i,j} - \lambda) \geq -\frac{(n_a+n_b)I}{4t} - \sum_{j\neq i}(\mathbb{E}A_{i,j} - \lambda)[\theta_{a,b}]_j\Bigg].$$

From Lemma C.1, when c_{init} is sufficiently small, with probability at least $1 - e^3 5^{-n}$ we have

$$\max\left\{\frac{|t-t^*|}{(p^*-q^*)/p^*}, \frac{|\lambda - \lambda^*|}{(p^*-q^*)}\right\} \leq 24c_0c_{\text{init}}. \tag{48}$$

Proposition C.3 shows that $\lambda^* \in (q^* + c(p^*-q^*), q^* + (1-c)(p^*-q^*))$ for some positive constant $0 < c < 1/2$. Therefore, when c_{init} is sufficiently small, we have $\lambda \in (q^*, p^*)$. Thus,

$$\left|\sum_{j\neq i}(\mathbb{E}A_{i,j} - \lambda)[\theta_{a,b}]_j\right| \leq (p^*-q^*)\|\theta_{a,b}\|_1 \leq (p^*-q^*)\|\pi - Z^*\|_1 \leq c_{\text{init}}(p^*-q^*)\bar{n}_{\min},$$

where we use Equation (40). By Equations (44) - (46), it is smaller than $(n_a + n_{z_i})/(8t)$ when c_{init} is sufficiently small. As a consequence, we have

$$\|h_{t,\lambda}(\pi) - Z^*\|_1 \leq 2wnk\exp(-\bar{n}_{\min}I/2) + 2w\sum_{a=1}^{k}\sum_{b\neq a}\sum_{i:z_i=b}\mathbb{I}\left[(A_{i,\cdot} - \mathbb{E}A_{i,\cdot})\theta_{a,b} \geq \frac{\bar{n}_{\min}I}{8t}\right]$$

$$+ 2w\sum_{a=1}^{k}\sum_{b\neq a}\sum_{i:z_i=b}\mathbb{I}\Bigg[\sum_{j\neq i}(Z^*_{j,a} - Z^*_{j,b})(A_{i,j} - \lambda) \geq -\frac{(n_a+n_b)I}{8t}\Bigg].$$

Define $L_1^{\text{sum}} = \sum_{a=1}^k\sum_{b\neq a}\sum_{i:z_i=b}\mathbb{I}\left[\sum_{j\neq i}(Z^*_{j,a} - Z^*_{j,b})(A_{i,j} - \lambda) \geq -(n_a+n_b)I/(8t)\right]$ and $L_2^{\text{sum}} = \sum_{a=1}^k\sum_{b\neq a}\sum_{i:z_i=b}\mathbb{I}\left[(A_{i,\cdot} - \mathbb{E}A_{i,\cdot})\theta_{a,b} \geq \bar{n}_{\min}I/(8t)\right]$. Our analysis on them is quite similar to that in Section 6.3. By Markov inequality,

$$\begin{aligned}
\mathbb{E}L_1^{\text{sum}} &= \sum_{a=1}^{k}\sum_{b\neq a}\sum_{i:z_i=b}\mathbb{P}\left[t^*\sum_{j\neq i}(Z^*_{j,a} - Z^*_{j,b})(A_{i,j} - \lambda) \geq -t^*(n_a+n_b)I/(8t)\right] \\
&\leq \sum_{a=1}^{k}\sum_{b\neq a}\sum_{i:z_i=b}\exp\left[\frac{t^*(n_a+n_b)I}{8t} - t^*(\lambda - \lambda^*)(n_a - n_b)\right]\mathbb{E}\exp\left[t^*\sum_{j\neq i}(Z^*_{j,a} - Z^*_{j,b})(A_{i,j} - \lambda^*)\right] \\
&\leq \sum_{a=1}^{k}\sum_{b\neq a}\sum_{i:z_i=b}\exp\left[\frac{t^*(n_a+n_b)I}{8t} - t^*(\lambda - \lambda^*)(n_a - n_b) - \frac{(n_a+n_b)I}{2}\right].
\end{aligned}$$

By Equations (44) - (46) and (48), when c_{init} is small enough, $t^*/t \leq 2$ and $t^*|\lambda - \lambda^*| \leq I/6$. Thus

$$\mathbb{E}L_1^{\text{sum}} \leq nk\exp(-\bar{n}_{\min}I/12).$$

Hence, with probability at least $1 - \exp(-\bar{n}_{\min}I/24)$,

$$L_1^{\text{sum}} \leq nk\exp(-\bar{n}_{\min}I/24).$$

For L_2^{sum} we use the same argument as in Section 6.3 and obtain

$$L_2^{\text{sum}} \leq \frac{4c_2knp^*\left\|\pi - Z^*\right\|_1}{(\bar{n}_{\min}I/(8t))^2} + \frac{80n^2kp^*\exp(-5np^*)}{\bar{n}_{\min}I/(8t)},$$

with probability at least $1-n^{-r}-\exp(-5np^*)$ for some constants $r, c_1, c_2 > 0$. Recall that

$$\left\|h_{t,\lambda}(\pi) - Z^*\right\|_1 \leq 2wnk\exp(-\bar{n}_{\min}I/2) + 2wL_1^{\text{sum}} + 2wL_2^{\text{sum}}.$$

Using the same argument as in Section 6.3, we conclude with

$$\left\|h_{t,\lambda}(\pi) - Z^*\right\|_1 \leq n\exp(-\bar{n}_{\min}I/25) + \frac{1}{2\sqrt{nI/[wk[n/\bar{n}_{\min}]^2]}}\left\|\pi - Z^*\right\|_1,$$

with probability at least $1 - \exp(-\bar{n}_{\min}I/10) - n^{-r}$.

B.3. Proof of Theorem 4.1. Define $t^* = \frac{1}{2}\log\frac{p^*(1-q^*)}{q^*(1-p^*)}$ and $\lambda^* = \frac{1}{2t^*}\log\frac{1-q^*}{1-p^*}$. By the same simplification we derive in Theorem 2.1, we have

$$\hat{\pi}^{\text{MF}} = \arg\max_{\pi\in\Pi_1} f'(\pi; A),$$

where

$$f'(\pi; A) = \langle A + \lambda^* I_n - \lambda^* 1_n 1_n^T, \pi\pi^T\rangle - \frac{1}{t^*}\sum_{i=1}^n \text{KL}(\text{Categorical}(\pi_{i,\cdot})\|\text{Categorical}(\pi_{i,\cdot}^{\text{pri}})).$$

Recall the definition of $h_{t,\lambda}(\cdot)$ as in Equation (11). A key observation is that $\hat{\pi}^{\text{MF}} = h_{t^*,\lambda^*}(\hat{\pi}^{\text{MF}})$, otherwise if there exists some $i \in [n]$ such that $[h_{t^*,\lambda^*}(\hat{\pi}^{\text{MF}})]_{i,\cdot}$ not equal to $\hat{\pi}_{i,\cdot}^{\text{MF}}$. This indicates the implementation of CAVI update on the i-th row of π will make change, leading to the decrease of $f'(\cdot; A)$. This contradicts with the fact that $\hat{\pi}^{\text{MF}}$ is the global minimizer.

The fixed-point property of $\hat{\pi}^{\text{MF}}$ is the key to our analysis. It involves three steps.

- *Step One.* For any π such that $\ell(\pi, Z^*) = o(\bar{n}_{\min})$, by the same analysis as in the proof of Theorem 3.1, we are able to show that there exist constant $r > 0$ and sequence $\eta = o(1)$ such that

$$\|h_{t^*,\lambda^*}(\pi) - Z^*\|_1 \leq n \exp(-(1-\eta)\bar{n}_{\min} I) + \frac{\|\pi - Z^*\|_1}{\sqrt{nI/[wk[n/\bar{n}_{\min}]^2]}},$$

with probability at least $1 - \exp[-(\bar{n}_{\min} I)^{\frac{1}{2}}] - n^{-r}$.
- *Step Two.* Lemma C.6 presents some loose upper bound for $\ell(\hat{\pi}^{\mathrm{MF}}, Z^*)$. That is, under the assumption $\rho n I/[wk^2[n/\bar{n}_{\min}]^2] \to \infty$, with probability at least $1 - e^3 5^{-n}$, we have

$$\ell(\hat{\pi}^{\mathrm{MF}}, Z^*) \leq o(\bar{n}_{\min}).$$

- *Step Three.* Using the property that $h_{t^*,\lambda^*}(\hat{\pi}^{\mathrm{MF}}) = \hat{\pi}^{\mathrm{MF}}$, we have

$$\left\|\hat{\pi}^{\mathrm{MF}} - Z^*\right\|_1 \leq n \exp(-(1-\eta)\bar{n}_{\min} I) + \frac{\left\|\hat{\pi}^{\mathrm{MF}} - Z^*\right\|_1}{\sqrt{nI/[wk[n/\bar{n}_{\min}]^2]}}$$

holds with probability at least $1 - \exp[-(\bar{n}_{\min} I)^{\frac{1}{2}}] - n^{-r}$. Then we obtain the desired result by simple algebra.

B.4. Proof of Theorem 4.2. By law of total expectation, we have

(49)
$$\begin{aligned}\mathbb{E}_{Z^{(s+1)}}\left[\left\|Z^{(s+1)} - Z^*\right\|_1 \middle| A, Z^{(0)}\right] &= \mathbb{E}_{\pi^{(s+1)}}\left[\mathbb{E}_{Z^{(s+1)}}\left[\left\|Z^{(s+1)} - Z^*\right\|_1 \middle| \pi^{(s+1)}, A, Z^{(0)}\right] \middle| A, Z^{(0)}\right] \\ &= \mathbb{E}_{\pi^{(s+1)}}\left[\left\|\pi^{(s+1)} - Z^*\right\|_1 \middle| A, Z^{(0)}\right],\end{aligned}$$

where the first equation is due to that the conditional expectation of $Z^{(s+1)}$ is $\pi^{(s+1)}$. We are going to build the connection between $\pi^{(s)}$ and $\pi^{(s+1)}$. In Algorithm 2, there are intermediate steps between $\pi^{(s)}$ and $\pi^{(s+1)}$ as follows:

$$\pi^{(s)} \rightsquigarrow Z^{(s)} \rightsquigarrow (p^{(s+1)}, q^{(s+1)}) \to (t^{(s+1)}, \lambda^{(s+1)}) \to \pi^{(s+1)},$$

where we use the plain right arrow ($\to$) to indicate deterministic generation and the curved right arrow ($\rightsquigarrow$) to indicate random generation. Despite a slight abuse of notation, we define $\pi^{(0)} = Z^{(0)}$.

Analogous to the proof of Theorem 3.1 in Section 6.3, we assume $\ell(Z^{(0)}, Z^*) = o(\bar{n}_{\min})$. The proof for the case $\ell(Z^{(0)}, Z^*)$ in the same order of $\bar{n}_{\min}$ is similar and thus is omitted.

Let $\gamma = o(1)$ be any sequence goes to 0 when n grows. We define a series of events as follows:

- global event $\mathcal{F}$: We define $\mathcal{F}$ exactly the same way as we define in the proof of Theorem 3.1 in Section 6.3 with respect to sequences γ and η', and we have $\mathbb{P}(\mathcal{F}) \geq 1 - \exp[-(\bar{n}_{\min} I)^{\frac{1}{2}})] - n^{-r}$ for some constant $r > 0$. We have $\eta' = o(1)$ whose value will be determined later.
- global event $\mathcal{G}$: Consider any $Z \in \Pi_1$ such that $\|Z - Z^*\|_1 \leq \gamma \bar{n}_{\min}$. Define

$$\alpha_p = \alpha_p^{\text{pri}} + \sum_{a=1}^{k} \sum_{i<j} A_{i,j} Z_{i,a} Z_{j,a}, \beta_p = \beta_p^{\text{pri}} + \sum_{a=1}^{k} \sum_{i<j} (1 - A_{i,j}) Z_{i,a} Z_{j,a},$$
$$\alpha_q = \alpha_q^{\text{pri}} + \sum_{a \neq b} \sum_{i<j} A_{i,j} Z_{i,a} Z_{j,b}, \beta_q = \beta_q^{\text{pri}} + \sum_{a \neq b} \sum_{i<j} (1 - A_{i,j}) Z_{i,a} Z_{j,b}.$$

Define $\mathcal{G}$ be the event that

$$\max \left\{ \left| \frac{\alpha_p}{\alpha_p + \beta_p} - p^* \right|, \left| \frac{\alpha_q}{\alpha_q + \beta_q} - q^* \right| \right\} \leq \eta''(p^* - q^*)$$

holds uniformly over all the eligible Z for some sequence $\eta'' = o(1)$. Then by the same analysis as in Lemma C.1, we have $\mathbb{P}(\mathcal{G}) \geq 1 - e^3 5^{-n}$.
- local events $\{\mathcal{H}_1^{(s)}\}_{s=1}^S$: We define $\mathcal{H}_1^{(s)} = \{\left\|\pi^{(s)} - Z^*\right\|_1 \geq \gamma \bar{n}_{\min}/2\}$.
- local events $\{\mathcal{H}_2^{(s)}\}_{s=1}^S$: We define $\mathcal{H}_2^{(s)} = \{\left\|Z^{(s)} - Z^*\right\|_1 \geq \gamma \bar{n}_{\min}\}$. For the conditional probability, we have

$$\begin{aligned}
&\mathbb{P}(\mathcal{H}_2^{(s)} = 1 | \mathcal{H}_1^{(s)} = 0) \\
&\leq \mathbb{P}\left[\left| \sum_{i=1}^{n} \left[\left\|Z_{i,\cdot}^{(s)} - Z_{i,\cdot}^*\right\|_1 - \left\|\pi_{i,\cdot}^{(s)} - Z_{i,\cdot}^*\right\|_1 \right] \right| \geq \gamma \bar{n}_{\min} - \left\|\pi^{(s)} - Z^*\right\|_1 \middle| \mathcal{H}_1^{(s)} = 0 \right] \\
&\leq \mathbb{P}\left[\left| \sum_{i=1}^{n} \left[\left\|Z_{i,\cdot}^{(s)} - Z_{i,\cdot}^*\right\|_1 - \left\|\pi_{i,\cdot}^{(s)} - Z_{i,\cdot}^*\right\|_1 \right] \right| \geq \gamma \bar{n}_{\min}/2 \middle| \mathcal{H}_1^{(s)} = 0 \right]
\end{aligned}$$

Since $\left\|\pi^{(s)} - Z^*\right\|_1 \leq \gamma \bar{n}_{\min}/2$ given $\mathcal{H}_1^{(s)} = 0$ by Bernstein inequality, we have

$$\begin{aligned}
\mathbb{P}(\mathcal{H}_2^{(s)} = 1 | \mathcal{H}_1^{(s)} = 0) &\leq \exp\left[- \frac{(\gamma \bar{n}_{\min})^2/8}{\left\|\pi^{(s)} - Z^*\right\|_1 + \gamma \bar{n}_{\min}/6} \right] \\
&\leq \exp\left[-3(\gamma \bar{n}_{\min})^2/16 \right].
\end{aligned}$$

- local events $\{\mathcal{H}_3^{(s)}\}_{s=1}^S$: We define $\mathcal{H}_3^{(s)} = \{|t^{(s)} - t^*| \geq \eta'(p^* - q^*)/p^*, \text{ or } |\lambda^{(s)} - \lambda^*| \geq \eta'(p^* - q^*)\}$. If the global event $\mathcal{G}$ holds and the local event $\mathcal{H}_2^{(s)}$

does not hold, we have

$$\max\left\{\left|\frac{\alpha_p^{(s+1)}}{\alpha_p^{(s+1)}+\beta_p^{(s+1)}}-p^*\right|,\left|\frac{\alpha_q^{(s+1)}}{\alpha_q^{(s+1)}+\beta_q^{(s+1)}}-q^*\right|\right\}\leq \eta''(p^*-q^*).$$

Note that $\alpha_p^{(s+1)}+\beta_p^{(s+1)}=\alpha_p^{\text{pri}}+\beta_p^{\text{pri}}+\sum_{a=1}^k\sum_{i<j}Z_{i,a}^{(s)}Z_{j,a}^{(s)}\geq n^2/k$. Using the tail bound of Beta distribution (Lemma C.7) we are able to show

$$\begin{aligned}
&\mathbb{P}\left[\left|p^{(s+1)}-\frac{\alpha_p^{(s+1)}}{\alpha_p^{(s+1)}+\beta_p^{(s+1)}}\right|\geq \eta''(p^*-q^*)\middle|\mathcal{H}_2^{(s)}=0,\mathcal{G}=1\right]\\
&\leq \exp\left[-\eta''^2n^2\frac{(p^*-q^*)^2}{2p^*}\right]\\
&\leq \exp\left[-\eta''^2n^2I/2\right],
\end{aligned}$$

where the last inequality is due to Proposition C.2. This leads to

$$\mathbb{P}\left[\left|p^{(s+1)}-p^*\right|\geq 2\eta''(p^*-q^*)\middle|\mathcal{H}_2^{(s)}=0,\mathcal{G}=1\right]\leq \exp\left[-\eta''^2n^2I/2\right].$$

And similar result holds for $q^{(s+1)}$. Then by the same analysis as in the proof of Lemma C.1, $\max\{|p^{(s+1)}-p^*|,|q^{(s+1)}-q^*|\}\leq 2\eta''(p^*-q^*)$ leads to

$$\max\left\{\frac{|t^{(s+1)}-t^*|}{(p^*-q^*)/p^*},\frac{|\lambda^{(s+1)-\lambda^*}|}{p^*-q^*}\right\}\leq 16c_0\eta''.$$

By taking $\eta'=16c_0\eta''$, we obtain

$$\mathbb{P}(\mathcal{H}_3^{(s+1)}=1|\mathcal{H}_2^{(s)}=0,\mathcal{G}=1)\leq 2\exp\left[-\eta''^2n^2I/2\right].$$

Note that events $\mathcal{F}$ and $\mathcal{G}$ are about the adjacency matrix A. The events $\mathcal{H}_1^{(s)},\mathcal{H}_2^{(s)}$ and $\mathcal{H}_3^{(s+1)}$ are for $\pi^{(x)}$, $Z^{(s)}$ and $(p^{(s+1)},q^{(s+1)})$ respectively. With all the above events defined, we can continue our analysis for Equation (49). Under the event $\mathcal{F}\cap\mathcal{G}\cap(\mathcal{H}_1^{(s)}\cup\mathcal{H}_2^{(s)}\cup\mathcal{H}_3^{(s+1)})^C$ we have

$$\left\|\pi^{(s+1)}-Z^*\right\|_1\leq n\exp(-(1-\eta)\bar{n}_{\min}I)+c_n\left\|\pi^{(s)}-Z^*\right\|_1, \tag{50}$$

where $c_n=[nI/[wk[n/\bar{n}_{\min}]^2]]^{-1/2}$. As a consequence, under the event $\mathcal{F}\cap\mathcal{G}\cap(\prod_{v=0}^s\mathcal{H}_1^{(v)}\cup\mathcal{H}_2^{(v)}\cup\mathcal{H}_3^{(v+1)})^C$, we have

$$\left\|\pi^{(s+1)}-Z^*\right\|_1\leq n\exp(-(1-2\eta)\bar{n}_{\min}I)+c_n^s\left\|\pi^{(0)}-Z^*\right\|_1.$$

Therefore, we have

(51)

$$\mathbb{E}_{\pi^{(s+1)}}\left[\left\|\pi^{(s+1)}-Z^*\right\|_1\middle|\mathcal{H}_1^{(0)}=0,\mathcal{F}=1,\mathcal{G}=1\right]\leq n\exp(-(1-2\eta)\bar{n}_{\min}I)$$
$$+c_n^s\left\|\pi^{(0)}-Z^*\right\|_1+n\mathbb{P}\left[\prod_{v=1}^{s}\mathcal{H}_1^{(v)}\cup\mathcal{H}_2^{(v)}\cup\mathcal{H}_3^{(v+1)}\middle|\mathcal{H}_1^{(0)}=0,\mathcal{F}=1,\mathcal{G}=1\right].$$

Due to the small value of c_n, if $\left\|\pi^{(s)}-Z^*\right\|_1\leq\gamma\bar{n}_{\min}$, Equation (50) immediately implies $\left\|\pi^{(s+1)}-Z^*\right\|_1\leq\gamma\bar{n}_{\min}$. This implies that under the event $\mathcal{F}\cup\mathcal{G}$ we have

$$\mathcal{H}_1^{(s+1)}\subset\mathcal{H}_1^{(s)}\cup\mathcal{H}_2^{(s)}\cup\mathcal{H}_3^{(s+1)},\forall s\geq 0,$$

and consequently,

$$\prod_{v=0}^{s}\mathcal{H}_1^{(v)}\cup\mathcal{H}_2^{(v)}\cup\mathcal{H}_3^{(v+1)}\subset\mathcal{H}_1^{(0)}\prod_{v=0}^{s}\mathcal{H}_2^{(v)}\cup\mathcal{H}_3^{(v+1)},\forall s\geq 1.$$

Thus,

$$\begin{aligned}(52)\qquad &\mathbb{P}\left[\prod_{v=0}^{s}\mathcal{H}_1^{(v)}\cup\mathcal{H}_2^{(v)}\cup\mathcal{H}_3^{(v+1)}\middle|\mathcal{H}_1^{(0)}=0,\mathcal{F}=1,\mathcal{G}=1\right]\\ &\leq\mathbb{P}\left[\prod_{v=0}^{s}\mathcal{H}_2^{(v)}\cup\mathcal{H}_3^{(v+1)}\middle|\mathcal{H}_1^{(0)}=0,\mathcal{F}=1,\mathcal{G}=1\right]\\ &\leq\sum_{v=0}^{s}\mathbb{P}(\mathcal{H}_2^{(v)}=1|\mathcal{H}_1^{(v)}=0)+\sum_{v=0}^{n}\mathbb{P}(\mathcal{H}_3^{(v+1)}=1|\mathcal{H}_2^{(v)}=0,\mathcal{G}=1)\\ &\leq(s+1)\left[\exp\left[-3(\gamma\bar{n}_{\min})^2/16\right]+2\exp\left[-\eta''^2n^2I/2\right]\right].\end{aligned}$$

Note that $\mathbb{P}(\mathcal{H}_1^{(0)}=0,\mathcal{F}=1,\mathcal{G}=1)\geq 1-\exp[-(\bar{n}_{\min}I)^{\frac{1}{2}})]-n^{-r}-e^35^{-n}-\epsilon$. Recall we define $\pi^{(0)}=Z^{(0)}$. By Equations (49), (51) and (52), we have

$$\mathbb{E}_{Z^{(s+1)}}\left[\left\|Z^{(s+1)}-Z^*\right\|_1\middle|A,Z^{(0)}\right]\leq n\exp(-(1-2\eta)\bar{n}_{\min}I)+c_n^s\left\|Z^{(0)}-Z^*\right\|_1+(s+1)nb_n,$$

with probability at least $1-\exp[-(\bar{n}_{\min}I)^{\frac{1}{2}})]-n^{-r}-e^35^{-n}-\epsilon$, where $b_n=\exp\left[-3(\gamma\bar{n}_{\min})^2/16\right]+2\exp\left[-\eta''^2n^2I/2\right]$.

B.5. Proof of Theorem 4.3. Note the similarity between Algorithm 3 and Algorithm 1. We can prove Theorem 4.3 with almost the identical argument used in the proof of Theorem 3.1, thus omitted.

APPENDIX C: STATEMENTS AND PROOFS OF AUXILIARY LEMMAS AND PROPOSITIONS

We include all the auxiliary propositions and lemmas in this section.

C.1. Statements and Proofs of Lemmas and Propositions for Theorem 3.1.

LEMMA C.1. *Let c_{init} be some sufficiently small constant. Consider any $\pi \in \Pi_1$ such that $\|\pi - Z^*\|_1 \leq c_{init} n/k$. Let $\alpha_p, \beta_p, \alpha_q, \beta_q, t, \lambda$ be the outputs after one step CAVI iteration from π described in Algorithm 1. That is, they are defined as Equations (28) - (31). Define*

$$\hat{p} = \frac{\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}A_{i,j}}{\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}}, \text{ and } \hat{q} = \frac{\sum_{i<j}\sum_{a\neq b} \pi_{i,a}\pi_{j,b}A_{i,j}}{\sum_{i<j}\sum_{a\neq b}\pi_{i,a}\pi_{j,b}}.$$

Under the same assumption as in Theorem 3.1, there exists some sequence $\epsilon = o(1)$ such that with probability at least $1-e^3 5^{-n}$, the following inequality holds

$$\max\left\{\frac{|\hat{p}-p^*|}{p^*-q^*}, \frac{|\hat{q}-q^*|}{p^*-q^*}, \frac{|t-t^*|}{(p^*-q^*)/p^*}, \frac{|\lambda-\lambda^*|}{p^*-q^*}\right\} \leq \epsilon + 24c_0\frac{\|\pi - Z^*\|_1}{n/k},$$

uniformly over all the eligible π. In addition if we further assume c_{init} goes to 0, the LHS of the above inequality will be simply upper bounded by ϵ.

PROOF. We are going to obtain tight bounds on $|\hat{p}-p^*|$ and $|\hat{q}-q^*|$ first. Note that we have the "variance-bias" decomposition as in

$$|\hat{p}-p^*| \leq \frac{|\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}(A_{i,j}-\mathbb{E}A_{i,j})|}{\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}} + \left|\frac{\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}\mathbb{E}A_{i,j}}{\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}} - p^*\right|.$$

We have concentration inequality holds for the numerator in the first term by Lemma C.2. That is, with probability at least $1-e^3 5^{-n}$, we have

$$\left|\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}(A_{i,j}-\mathbb{E}A_{i,j})\right| = \left|\frac{1}{2}\langle A-\mathbb{E}A, \pi\pi^T\rangle\right| \leq 3n\sqrt{np^*}$$

holds uniformly over all $\pi \in \Pi_1$. For the denominator, we have

$$\frac{n^2}{2} \geq \sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a} = \frac{1}{2}\sum_{a=1}^k \|\pi_{\cdot,a}\|_1^2 \geq \frac{n^2}{2k},$$

since $\sum_{a=1}^k \|\pi_{\cdot,a}\|_1 = n$. Thus, we are able to obtain an upper bound on the first term as

$$\frac{|\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}(A_{i,j}-\mathbb{E}A_{i,j})|}{\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}} \leq 6\sqrt{\frac{k^2p^*}{n}}.$$

For the second term, since $\mathbb{E}A_{i,j} = p^*\sum_{a=1}^k Z^*_{i,a}Z^*_{j,a} + q^*(1-\sum_{a=1}^k Z^*_{i,a}Z^*_{j,a})$, we have

$$\begin{aligned}\left|\frac{\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}\mathbb{E}A_{i,j}}{\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}} - p^*\right| &= (p^*-q^*)\frac{\left|\sum_{i<j}\left[\sum_{a=1}^k \pi_{i,a}\pi_{j,a}\right]\left[\sum_{a=1}^k 1 - Z^*_{i,a}Z^*_{j,a}\right]\right|}{\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}} \\ &= (p^*-q^*)\frac{\left|\langle \pi\pi^T, 11^T - Z^*Z^{*T}\rangle\right|}{\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}} \\ &= (p^*-q^*)\frac{\left|\langle \pi\pi^T - Z^*Z^{*T}, 11^T - Z^*Z^{*T}\rangle\right|}{\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}},\end{aligned}$$

where in the last inequality we use the orthogonality between Z^*Z^{*T} and $11^T - Z^*Z^{*T}$. For its numerator, we have

$$\begin{aligned}\left|\langle \pi\pi^T - Z^*Z^{*T}, 11^T - Z^*Z^{*T}\rangle\right| &\leq \left\|\pi\pi^T - Z^*Z^{*T}\right\|_1 \\ &\leq \|\pi - Z^*\|_1 \left(\|\pi\|_1 + \|Z^*\|_1\right) \\ &\leq \|\pi - Z^*\|_1 \left(2\|Z^*\|_1 + \|\pi - Z^*\|_1\right) \\ &\leq 3n\|\pi - Z^*\|_1.\end{aligned}$$

This leads to

$$\left|\frac{\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}\mathbb{E}A_{i,j}}{\sum_{i<j}\sum_{a=1}^k \pi_{i,a}\pi_{j,a}} - p^*\right| \leq \frac{3n\|\pi - Z^*\|_1(p^*-q^*)}{n^2/k} \leq 3kn^{-1}(p^*-q^*)\|\pi - Z^*\|_1.$$

Thus,

$$|\hat{p} - p^*| \leq 6\sqrt{\frac{k^2p^*}{n}} + 3kn^{-1}(p^*-q^*)\|\pi - Z^*\|_1 \leq \left[\sqrt{\frac{k^2p^*}{n(p^*-q^*)^2}} + \frac{3\|\pi - Z^*\|_1}{n/k}\right](p^*-q^*).$$

Similar result holds for $|\hat{q} - q^*|$. Denote $\eta_0 = \sqrt{\frac{k^2p^*}{n(p^*-q^*)^2}} + \frac{3\|\pi - Z^*\|_1}{n/k}$, thus

$$\max\{|\hat{p}-p^*|, |\hat{q}-q^*|\} \leq \eta_0(p^*-q^*).$$

By the assumption of nI in Equation (19) and Proposition C.2, we have $n(p^*-q^*)^2/(k^2p^*) \asymp nI/k^2 \to \infty$. Therefore, the first term in η_0 goes to 0. The second term in η_0 is at most $3c_{\text{init}}$ which implies $\eta_0 \leq 4c_{\text{init}}$.

By the fact that the digamma function satisfies $\psi(x) \in (\log(x-1/2), \log x), \forall x \geq 1/2$, we have

$$\begin{aligned}
\psi(\alpha_p) - \psi(\beta_p) &\geq \log \frac{\alpha_p - 1/2}{\beta_p} \\
&= \log \left[\frac{\left[\alpha_p^{\text{pri}} - 1/2 + \sum_{i<j} \sum_{a=1}^k \pi_{i,a}\pi_{j,a}A_{i,j}\right] / \left[\sum_{i<j} \sum_{a=1}^k \pi_{i,a}\pi_{j,a}\right]}{1 + \left[\beta_p^{\text{pri}} - \sum_{i<j} \sum_{a=1}^k \pi_{i,a}\pi_{j,a}A_{i,j}\right] / \left[\sum_{i<j} \sum_{a=1}^k \pi_{i,a}\pi_{j,a}\right]} \right] \\
&= \log \left[\frac{\hat{p} + (\alpha_p^{\text{pri}} - 1/2) / \left[\sum_{i<j} \sum_{a=1}^k \pi_{i,a}\pi_{j,a}\right]}{1 - \hat{p} + \beta_p^{\text{pri}} / \left[\sum_{i<j} \sum_{a=1}^k \pi_{i,a}\pi_{j,a}\right]} \right].
\end{aligned}$$

Recall that we have shown $\sum_{i<j} \sum_{a=1}^k \pi_{i,a}\pi_{j,a}$ lies in the interval of $(n^2/(2k), n^2/2)$. By Equation (19), there exists a sequence $\eta' = o(1)$ such that $\alpha_p, \beta_p \leq \eta'(p^* - q^*)n^2/k$. Then we have

$$\psi(\alpha_p) - \psi(\beta_p) \geq \log \frac{p^* - |p^* - \hat{p}| - \eta'(p^* - q^*)}{1 - p^* + |p^* - \hat{p}| + \eta'(p^* - q^*)}.$$

Similar analysis leads to

$$\psi(\alpha_q) - \psi(\beta_q) \leq \log \frac{q^* + |q^* - \hat{q}| + \eta'(p^* - q^*)}{1 - q^* - |q^* - \hat{q}| - \eta'(p^* - q^*)}.$$

Together we have

$$\begin{aligned}
t - t^* &\geq \log \left[\frac{p^* - |p^* - \hat{p}| - \eta'(p^* - q^*)}{1 - p^* + |p^* - \hat{p}| + \eta'(p^* - q^*)} \frac{1 - q^* - |q^* - \hat{q}| - \eta'(p^* - q^*)}{q^* + |q^* - \hat{q}| + \eta'(p^* - q^*)} \right] - t^* \\
&\geq \log \left[\left[1 - \frac{|p^* - \hat{p}| + \eta'(p^* - q^*)}{q^*}\right]^4 \frac{p^*(1 - q^*)}{q^*(1 - p^*)} \right] - t^* \\
&= 4 \log \left[1 - (\eta_0 + \eta') \frac{p^* - q^*}{q^*}\right].
\end{aligned}$$

Recall that we assume $c_0 p^* < q^* < p^*$. Thus $(\eta_0 + \eta')(p^* - q^*)/p^* \leq 5c_{\text{init}}c_0$. When c_{init} is sufficiently small, we have $(\eta_0 + \eta')(p^* - q^*)/p^* \leq 1/2$. Then using the fact $-x \geq \log(1-x) \geq -2x, \forall x \in (0, 1/2)$. We have

$$t - t^* \geq -8(\eta_0 + \eta')(p^* - q^*)/q^*.$$

Analogously we can obtain the same upper bound on $\hat{t} - t^*$, and then

$$|t - t^*| \leq 8c_0(\eta_0 + \eta') \frac{p^* - q^*}{p^*}.$$

Identical analysis can be applied towards bounds on $|\hat{\lambda} - \lambda^*|$. Note that

$$\log \frac{\beta_p}{\alpha_p + \beta_p} = \log \left[\frac{1 - \hat{p} + \beta_p^{\text{pri}} / \left[\sum_{i<j} \sum_{a=1}^k \pi_{i,a}\pi_{j,a} \right]}{1 + (\alpha_p^{\text{pri}} + \beta_p^{\text{pri}}) / \left[\sum_{i<j} \sum_{a=1}^k \pi_{i,a}\pi_{j,a} \right]} \right],$$

similarly for α_q, β_q. Omitting the immediate steps, we end up with

$$|\lambda - \lambda^*| = | [\psi(\beta_q) - \psi(\alpha_q + \beta_q)] - [\psi(\beta_p) - \psi(\alpha_p + \beta_p)] - \lambda^* | \leq 8(\eta_0 + \eta')(p^* - q^*).$$

The proof is complete after we unify and rephrase all the aforementioned results. □

LEMMA C.2. *Let $A \in [0,1]^{n\times n}$ such that $A = A^T$ and $A_{i,i} = 0, \forall i \in [n]$. Assume $\{A_{i,j}\}_{i<j}$ are independent random variable, and there exists $p \leq 1$ such that $9n^{-1} \leq \frac{2}{n(n-1)} \sum_{i<j} Var(A_{i,j}) \leq p$, and then we have*

$$\sup_{\pi \in \Pi_1} \left| \langle A - \mathbb{E}A, \pi\pi^T \rangle \right| \leq 6n\sqrt{np},$$

with probability at least $1 - e^3 5^{-n}$.

PROOF. This result is a direct consequence of Grothendieck inequality [2] (see also Theorem 3.1 of [3] for a rephrased statement) on the matrix $A - \mathbb{E}A$. The Lemma 4.1 of [3] proves that with probability at least $1 - e^3 5^{-n}$,

$$\sup_{s,t \in \{-1,1\}^n} \left| \sum_{i,j} (A_{i,j} - \mathbb{E}A_{i,j}) s_i t_j \right| \leq 3n\sqrt{np}.$$

Then by applying Grothendieck inequality we obtain

$$\sup_{\|X_i\|_2 \leq 1, \forall i \in [n]} \left| \sum_{i,j} (A_{i,j} - \mathbb{E}A_{i,j}) X_i^T X_j \right| \leq 3cn\sqrt{np},$$

where c is a positive constant smaller than 2. This concludes with

$$\sup_{\pi \in \Pi_1} \left| \langle A - \mathbb{E}A, \pi\pi^T \rangle \right| \leq 6n\sqrt{np},$$

□

PROPOSITION C.1. *Assume $0 < q < p < 1$. Let $X \sim Ber(q)$ and $Y \sim Ber(p)$. Recall the definition $\lambda = \log \frac{1-q}{1-p} / \log \frac{p(1-q)}{q(1-p)}$, $t = \frac{1}{2} \log \frac{p(1-q)}{q(1-p)}$ and $I = -2\log[\sqrt{pq} + \sqrt{(1-p)(1-q)}]$. Then the following two equations hold*

$$e^{t\lambda} = \left(\frac{\mathbb{E}e^{tX}}{\mathbb{E}e^{-tY}} \right)^{\frac{1}{2}}, \quad \text{and} \quad \mathbb{E}e^{tX}\mathbb{E}e^{-tY} = \exp(-I). \tag{53}$$

PROOF. The proof is straightforward and all by calculation. Note that $\mathbb{E}\exp(tX) = pe^t + 1 - p$ and $\mathbb{E}\exp(tY) = qe^t + 1 - q$. We can easily obtain

$$\mathbb{E}e^{tX}\mathbb{E}e^{-tY} = (pe^t + 1 - p)(qe^{-t} + 1 - q) = (\sqrt{pq} + \sqrt{(1-p)(1-q)})^2 = \exp(-I).$$

We can justify the first part of Equation (53) in a similar way. □

LEMMA C.3. *[Theorem 5.2 of [5]] Let $A \in \{0,1\}^{n\times n}$ be a symmetric binary matrix with $A_{i,i} = 0, \forall i \in [n]$, and $\{A_{i,j}\}_{i<j}$ are independent Bernoulli random variable. If $p \triangleq \max_{i,j} \mathbb{E}A_{i,j} \geq \log n/n$. Then there exist constants $c, r > 0$ such that*

$$\|A - \mathbb{E}A\|_{\mathrm{op}} \leq c\sqrt{np},$$

with probability at least $1 - n^{-r}$.

The following lemma on the operator norm of sparse networks is from [1]. In the original statement of Lemma 12 in [1], "with probability $1 - o(1)$" is stated. However, its proof in [1] gives explicit form of the probability that the statement holds, which is at least $1 - n^{-1}$.

LEMMA C.4. *[Lemma 12 of [1]] Suppose M is random symmetric matrix with zero on the diagonal whose entries above the diagonal are independent with the following distribution*

$$M_{i,j} = \begin{cases} 1 - p_{i,j}, & w.p.\ p_{i,j}; \\ -p_{i,j}, & w.p.\ 1 - p_{i,j}. \end{cases}$$

Let $p \triangleq \max_{i,j} p_{i,j}$ and $\tilde{M}$ be the matrix obtained from M by zeroing out all the rows and columns having more than $20np$ positive entries. Then there exists some constant $c > 0$ such that

$$\|\tilde{M}\|_{\mathrm{op}} \leq c\sqrt{np},$$

holds with probability at least $1 - n^{-1}$.

LEMMA C.5. *Let $A \in \{0,1\}^{n\times n}$ be a symmetric binary matrix with $A_{i,i} = 0, \forall i \in [n]$, and $\{A_{i,j}\}_{i<j}$ are independent Bernoulli random variable. Let $p \geq \max_{i,j} \mathbb{E}A_{i,j}$. Define $S = \{i \in [n], \sum_j A_{i,j} \geq 20np\}$ and $Z_i = \sum_j |A_{i,j} - \mathbb{E}A_{i,j}|\mathbb{I}\{i \in S\}$. Then with probability at least $1 - \exp(-5np)$, we have*

$$\sum_i Z_i \leq 20n^2 p \exp(-5np).$$

PROOF. Note that $\mathbb{E}\sum_j |A_{i,j} - \mathbb{E}A_{i,j}| \le 2np(1-p) \le 2np$. For any $s \ge 20np$, we have

$$\begin{aligned}
\mathbb{P}(Z_i > s) &\le \mathbb{P}\left[\sum_j |A_{i,j} - \mathbb{E}A_{i,j}| - \mathbb{E}\sum_j |A_{i,j} - \mathbb{E}A_{i,j}| > s - 2np\right] \\
&\le \exp\left[-\frac{\frac{1}{2}(s-2np)^2}{np + \frac{1}{3}(s-2np)}\right] \\
&\le \exp(-s/2),
\end{aligned}$$

by implementing Bernstein inequality. Applying Bernstein inequality again we have

$$\begin{aligned}
\mathbb{P}(Z_i > 0) &= \mathbb{P}\left[\sum_j A_{i,j} \ge 20np\right] \\
&\le \mathbb{P}\left[\sum_j A_{i,j} - \mathbb{E}\sum_j A_{i,j} \ge 18np\right] \\
&\le \exp\left[-\frac{(18np)^2/2}{np + 18np/3}\right] \\
&\le \exp(-21np/2).
\end{aligned}$$

Thus, we are able to bound $\mathbb{E}Z_i$ with

$$\begin{aligned}
\mathbb{E}Z_i &\le \int_0^{20np} \mathbb{P}(Z_i > 0)\,\mathrm{d}s + \int_{20np}^{\infty} \mathbb{P}(Z_i > s)\,\mathrm{d}s \\
&\le 20np\exp(-21np/2) + \int_{20np}^{\infty} \exp(-s/2) \\
&\le 20np\exp(-10np).
\end{aligned}$$

By Markov inequality, we have

$$\begin{aligned}
\mathbb{P}\left[\sum_{i,j} |A_{i,j} - \mathbb{E}A_{i,j}|\mathbb{I}\{i \in S\} \ge 20n^2p\exp(-5np)\right] &= \mathbb{P}\left[\sum_i Z_i \ge 20n^2p\exp(-5np)\right] \\
&\le \frac{n\mathbb{E}Z_1}{20n^2p\exp(-5np)} \\
&\le \exp(-5np).
\end{aligned}$$

□

PROPOSITION C.2. *Under the assumption that* $0 < q < p = o(1)$. *For* $I = -2\log\left[\sqrt{pq} + \sqrt{(1-p)(1-q)}\right]$ *we have*

$$I = (1 + o(1))(\sqrt{p} - \sqrt{q})^2.$$

Consequently, $(p-q)^2/(4p) \le I \le (p-q)^2/p$.

PROOF. It is a partial result of Lemma B.1 in [6]. □

PROPOSITION C.3. *Define* $\lambda = \log\frac{1-q}{1-p}/\log\frac{p(1-q)}{q(1-p)}$. *For any* $p, q > 0$ *such that* $p, q = o(1)$ *and* $p \asymp q$, *there exists a constant* $0 < c < 1/2$ *such that*

$$\frac{\lambda - q}{p - q} \in (c, 1 - c).$$

PROOF. First we are going to establish the lower bound. Let $x = p - q$, and then we can rewrite λ as

$$\lambda = \frac{1}{1 + \frac{\log(1+x/q)}{\log(1+x/(1-q-x))}}.$$

Case I: $x \ge q/10$. Define $s = (p-q)/q$. Since $p \asymp q$ we have $s \ge 1/10$ and also upper bounded by some constant. We have

$$\begin{aligned}
\frac{\lambda - q}{p - q} &= \frac{1}{s}\left[\frac{1}{q}\frac{1}{1 + \frac{\log(1+s)}{\log(1+sq/(1-(s+1)q))}} - 1\right] \\
&= \frac{1}{s}\left[\frac{(1-q)\log(1+sq/(1-(s+1)q)) - q\log(1+s)}{q\log(1+sq/(1-(s+1)q)) + q\log(1+s)}\right] \\
&\ge \frac{1}{s}\frac{(1-q)\frac{sq}{1-(s+1)q} - q\log(1+s)}{2q\log(1+s)} \\
&\ge \frac{1}{8}\frac{1-q}{\log(1+s)},
\end{aligned}$$

which is lower bounded by some constant $c > 0$.

Case II: $x < q/10$. By Taylor theorem, there exist constants $0 \le \epsilon_1, \epsilon_2 \le 1/10$ such that

$$\log\left[1 + \frac{x}{q}\right] = \frac{x}{q} - \frac{1-\epsilon_1}{2}\left[\frac{x}{q}\right]^2,$$

$$\text{and } \log\left[1 + \frac{x}{1-q-x}\right] = \frac{x}{1-q-x} - \frac{1-\epsilon_2}{2}\left[\frac{x}{1-q-x}\right]^2.$$

Thus, we have

$$\frac{\log(1+\frac{x}{q})}{\log(1+\frac{x}{1-q-x})} = \frac{q(1-q)^2 - \left[2q(1-q) + \frac{1-\epsilon_1}{2}(1-q)^2\right]x + c_1x^2 + c_2x^3}{q^2(1-q) - \frac{3-\epsilon_2}{2}q^2x},$$

where $c_1 = (1-\epsilon_1)(1-q) + q$ and $c_2 = -(1-\epsilon_1)/2$. Thus,

$$\begin{aligned}\frac{\lambda - q}{p-q} &= \frac{1}{x}\left[\frac{q^2(1-q) - \frac{3-\epsilon_2}{2}q^2x}{q(1-q) - \left[2q(1-q) + \frac{1-\epsilon_1}{2}(1-q)^2 + \frac{3-\epsilon_2}{2}q^2\right]x + c_1x^2 + c_2x^3} - q\right] \\ &= \frac{\left[\frac{1}{2}q(1-q) + \frac{\epsilon_2}{2}q^2(1-q) - \frac{\epsilon_1}{2}(1-q)^2q\right] + c_1qx + c_2qx^2}{q(1-q) - \left[2q(1-q) + \frac{1-\epsilon_1}{2}(1-q)^2 + \frac{3-\epsilon_2}{2}q^2\right]x + c_1x^2 + c_2x^3}\end{aligned}$$

Note that $|c_1|, |c_2| \leq 1$. We have

$$\frac{\lambda - q}{p-q} \geq \frac{\frac{1}{4}q(1-q)}{2q(1-q)} \geq 1/8.$$

By using exactly the same discussion, we can show $(p-\lambda)/(p-q) > c$. Thus, we proved the desired bound stated in the proposition. □

C.2. Statements and Proofs of Lemmas and Propositions for Theorem 4.1.

LEMMA C.6. *Let $Z^* \in \Pi_0$. Assume $p^*, q^* = o(1)$ and $p^* \asymp q^*$. Define t^*, λ^* and $\hat{\pi}^{MF}$ the same way as in Theorem 4.1. If $nI/[k\log kw] \to \infty$, we have with probability at least $1 - e^3 5^{-n}$,*

$$\left\|Z^*Z^{*T} - \hat{\pi}^{MF}(\hat{\pi}^{MF})^T\right\|_1 \lesssim n^2/\sqrt{nI}.$$

If we further assume $Z^* \in \Pi_0^{(\rho,\rho')}$ with arbitrary ρ, ρ', and then we have with probability at least $1 - e^3 5^{-n}$,

$$\ell(\hat{\pi}^{MF}, Z^*) \lesssim \rho^{-1} n\sqrt{k^2/(nI)}.$$

PROOF. Form Lemma C.2, with probability at least $1 - e^3 5^{-n}$, we have uniformly for all $\pi \in \Pi_1$

$$|\langle A - \mathbb{E}A, \pi\pi^T\rangle| \leq 6n\sqrt{np^*}. \tag{54}$$

In the remaining part of the proof, we always assume the above event holds. Denote $f'(\pi) = \langle A + \lambda^* I_n - \lambda^* 1_n 1_n^T, \pi\pi^T\rangle - (t^*)^{-1}\sum_{i=1}^n \mathrm{KL}(\pi_{i,\cdot}\|\pi_{i,\cdot}^{\mathrm{pri}})$ for any

$\pi \in \Pi_1$. Here we adopt the notation $\mathrm{KL}(\pi_{i,\cdot}\|\pi^{\mathrm{pri}}_{i,\cdot})$ short for $\mathrm{KL}(\mathrm{Categorical}(\pi_{i,\cdot})\|\mathrm{Categorical}(\pi^{\mathrm{pri}}_{i,\cdot}))$, and we do it in the same way in the rest part of the proof. Thus,

$$\begin{aligned}
\langle \mathbb{E}A + \lambda^* I_n - \lambda^* 1_n 1_n^T, \hat{\pi}^{\mathrm{MF}}(\hat{\pi}^{\mathrm{MF}})^T\rangle &\geq \langle A + \lambda^* I_n - \lambda^* 1_n 1_n^T, \hat{\pi}^{\mathrm{MF}}(\hat{\pi}^{\mathrm{MF}})^T\rangle - 6n\sqrt{np^*} \\
&= f'(\hat{\pi}^{\mathrm{MF}}) - 6n\sqrt{np^*} + (t^*)^{-1}\sum_{i=1}^n \mathrm{KL}(\hat{\pi}^{\mathrm{MF}}_{i,\cdot}\|\pi^{\mathrm{pri}}_{i,\cdot}) \\
&\geq f'(Z^*) - 6n\sqrt{np^*} + (t^*)^{-1}\sum_{i=1}^n \mathrm{KL}(\hat{\pi}^{\mathrm{MF}}_{i,\cdot}\|\pi^{\mathrm{pri}}_{i,\cdot}) \\
&\geq \langle \mathbb{E}A + \lambda^* I_n - \lambda^* 1_n 1_n^T, Z^* Z^{*T}\rangle - 12n\sqrt{np^*} \\
&\quad + (t^*)^{-1}\sum_{i=1}^n \mathrm{KL}(\hat{\pi}^{\mathrm{MF}}_{i,\cdot}\|\pi^{\mathrm{pri}}_{i,\cdot}) - (t^*)^{-1}\sum_{i=1}^n \mathrm{KL}(Z^*_{i,\cdot}\|\pi^{\mathrm{pri}}_{i,\cdot}),
\end{aligned}$$

where we use Equation (54) twice in the first and last inequality. Note that for any $\pi \in \Pi_1$, we have

$$|\mathrm{KL}(\pi_{i,\cdot}\|\pi^{\mathrm{pri}}_{i,\cdot})| \leq |\sum_j \pi_{i,j}\log \pi_{i,j}| + |\sum_j \pi_{i,j}\log \pi^{\mathrm{pri}}_{i,j}| \leq \log k + \log w,$$

where the second inequality is due to $0 \geq \sum_j \pi_{i,j}\log\pi_{i,j} = \mathrm{KL}(\pi_{i,\cdot}\|k^{-1}1_k) - \log k \geq -\log k$, where $k^{-1}1_k$ can be explicitly written as a length-k vector $(1/k, 1/k, \ldots, 1/k)$. Then we have

$$\left|\sum_{i=1}^n \mathrm{KL}(\hat{\pi}^{\mathrm{MF}}_{i,\cdot}\|\pi^{\mathrm{pri}}_{i,\cdot}) - \sum_{i=1}^n \mathrm{KL}(Z^*_{i,\cdot}\|\pi^{\mathrm{pri}}_{i,\cdot})\right| \leq 2n\log kw.$$

Thus,

$$\langle \mathbb{E}A + \lambda^* I_n - \lambda^* 1_n 1_n^T, Z^* Z^{*T} - \hat{\pi}^{\mathrm{MF}}(\hat{\pi}^{\mathrm{MF}})^T\rangle \leq 12n\sqrt{np^*} + 2(t^*)^{-1} n\log kw.$$

By Proposition C.4, we have

$$\langle \mathbb{E}A + \lambda^* I_n - \lambda^* 1_n 1_n^T, Z^* Z^{*T} - \hat{\pi}^{\mathrm{MF}}(\hat{\pi}^{\mathrm{MF}})^T\rangle \geq 2(p^* - q^*)\left[\left(1 - \frac{\lambda^* - q^*}{p^* - q^*}\right)\alpha + \frac{\lambda^* - q^*}{p^* - q^*}\gamma\right],$$

where $\alpha = \langle Z^* Z^{*T} - \hat{\pi}^{\mathrm{MF}}(\hat{\pi}^{\mathrm{MF}})^T, Z^* Z^{*T} - I_n\rangle/2$ and $\gamma = \langle \hat{\pi}^{\mathrm{MF}}(\hat{\pi}^{\mathrm{MF}})^T - Z^* Z^{*T}, 1_n 1_n^T - Z^* Z^{*T}\rangle/2$. By Proposition C.3, there exists a constant $c > 0$ such that

$$\langle \mathbb{E}A + \lambda^* I_n - \lambda^* 1_n 1_n^T, Z^* Z^{*T} - \hat{\pi}^{\mathrm{MF}}(\hat{\pi}^{\mathrm{MF}})^T\rangle \geq 2c(p^* - q^*)(\alpha + \gamma). \tag{55}$$

Note that the following inequality holds

$$\begin{aligned}2(\alpha+\gamma) &= \left\|Z^*Z^{*T}-\hat{\pi}^{\mathrm{MF}}(\hat{\pi}^{\mathrm{MF}})^T\right\|_1 - \langle Z^*Z^{*T}-\hat{\pi}^{\mathrm{MF}}(\hat{\pi}^{\mathrm{MF}})^T, I_n\rangle/2 \\ &\geq \left\|Z^*Z^{*T}-\hat{\pi}^{\mathrm{MF}}(\hat{\pi}^{\mathrm{MF}})^T\right\|_1 - n/2.\end{aligned}$$

These together lead to

$$\left\|Z^*Z^{*T}-\hat{\pi}^{\mathrm{MF}}(\hat{\pi}^{\mathrm{MF}})^T\right\|_1 \leq \frac{1}{c(p^*-q^*)}\left[12n\sqrt{np^*}+2(t^*)^{-1}n\log kw + c(p^*-q^*)n/2\right].$$

Note that $t^* \asymp (p^*-q^*)/p^*$ when $p^* \asymp q^*$. Together by Proposition C.2, as long as $nI/[k\log kw] \to \infty$, the last two terms in the RHS of the above formula is dominated by the first term. Thus,

$$\left\|Z^*Z^{*T}-\hat{\pi}^{\mathrm{MF}}(\hat{\pi}^{\mathrm{MF}})^T\right\|_1 \lesssim \frac{n^2}{\sqrt{nI}}.$$

If we further assume $Z^* \in \Pi_0^{(\rho,\rho')}$, Proposition C.5 and Equation (55) lead to

$$\langle \mathbb{E}A + \lambda^* I_n - \lambda^* 1_n 1_n^T, Z^*Z^{*T} - \hat{\pi}^{\mathrm{MF}}(\hat{\pi}^{\mathrm{MF}})^T\rangle \geq \frac{\rho c n(p^*-q^*)}{8k}\ell(\hat{\pi}^{\mathrm{MF}}, Z^*).$$

So we have

$$\begin{aligned}\ell(\hat{\pi}^{\mathrm{MF}}, Z^*) &\leq \frac{8k}{\rho c n(p^*-q^*)}(12n\sqrt{np^*}+2(t^*)^{-1}n\log kw) \\ &\leq \frac{192k}{\rho c}\sqrt{\frac{np^*}{(p^*-q^*)^2}}.\end{aligned}$$

□

Before we state the remaining lemmas and propositions used in the Proof of Lemma C.6, we first introduce two definitions. For any $\pi, \pi' \in [0,1]^{n\times k}$, define $\alpha(\pi;\pi') = \langle \pi'\pi'^T - \pi\pi^T, \pi'\pi'^T - I_n\rangle/2$ and $\gamma(\pi;\pi') = \langle \pi\pi^T - \pi'\pi'^T, 1_n 1_n^T - \pi'\pi'^T\rangle/2$.

PROPOSITION C.4. *Define $P = Z^*BZ^{*T} - pI_n$, with $B = q1_k1_k^T + (p-q)I_k$. We have the equation*

$$\langle P + \lambda I_n - \lambda 1_n 1_n^T, Z^*Z^{*T} - \pi\pi^T\rangle = 2(p-q)\left[\left(1-\frac{\lambda-q}{p-q}\right)\alpha(\pi;Z^*) + \frac{\lambda-q}{p-q}\gamma(\pi;Z^*)\right].$$

PROOF. Note that $Z^*BZ^{*T} - pI_n = (p-q)Z^*Z^{*T} + q1_n1_n^T$. We have

$$\begin{aligned}
\langle P + \lambda I_n - \lambda 1_n 1_n^T, Z^*Z^{*T} - \pi\pi^T \rangle &= (p-q)\langle Z^*Z^{*T} - \frac{\lambda - q}{p-q}1_n1_n^T + \frac{\lambda - p}{p-q}I_n, Z^*Z^{*T} - \pi\pi^T\rangle \\
&= (p-q)\langle Z^*Z^{*T} - I_n, Z^*Z^{*T} - \pi\pi^T\rangle \\
&\quad + (\lambda - q)\langle I_n - 1_n1_n^T, Z^*Z^{*T} - \pi\pi^T\rangle \\
&= (p-\lambda)\langle Z^*Z^{*T} - I_n, Z^*Z^{*T} - \pi\pi^T\rangle \\
&\quad + (\lambda - q)\langle Z^*Z^{*T} - 1_n1_n^T, Z^*Z^{*T} - \pi\pi^T\rangle \\
&= 2(p-q)\alpha(\pi; Z^*) + 2(\lambda - q)\gamma(\pi; Z^*).
\end{aligned}$$

Consequently, we obtain the desired bound. □

PROPOSITION C.5. *If* $Z^* \in \Pi_0^{(\rho,\rho')}$, $\pi \in \Pi_1$, *we have*

$$\alpha(\pi; Z^*) + \gamma(\pi; Z^*) \geq \frac{\rho n}{16k}\ell(\pi, Z^*).$$

PROOF. We use α, γ instead of $\alpha(\pi; Z^*), \gamma(\pi; Z^*)$ for simplicity. Without loss of generality we assume $\|\pi - Z^*\|_1 = \ell(\pi, Z^*)$. Define $\mathcal{C}_u = \{i : Z^*_{i,u} = 1\}$ and $L_{u,v} = \sum_{i \in \mathcal{C}_u} \pi_{i,v}$. We have the equality $\sum_v L_{u,v} = |\mathcal{C}_u|$ and also

$$\alpha = \frac{1}{2}\sum_u \left[|\mathcal{C}_u|^2 - \sum_{i,j \in \mathcal{C}_u}\sum_w \pi_{i,w}\pi_{j,w}\right] = \frac{1}{2}\sum_u\left[|\mathcal{C}_u|^2 - \sum_w L_{u,w}^2\right] = \frac{1}{2}\sum_u\sum_{w \neq w'} L_{u,w}L_{u,w'}$$

and $\gamma = \frac{1}{2}\sum_{u \neq v}\sum_{i \in \mathcal{C}_u, j \in \mathcal{C}_v}\sum_w \pi_{i,w}\pi_{j,w} = \frac{1}{2}\sum_{u\neq v}\sum_w L_{u,w}L_{v,w}$.

We define $[k]$ into two disjoint subsets S_1 and S_2 where

$$\begin{aligned}
S_1 &= \left\{u \in [k] : \forall v \neq u, L_{u,v} \leq \frac{3}{4}|\mathcal{C}_u|\right\}, \\
\text{and } S_2 &= \left\{i \in [k] : \exists v \neq u, L_{u,v} > \frac{3}{4}|\mathcal{C}_u|\right\}.
\end{aligned}$$

Define $L_u = \sum_{v \neq u} L_{u,v}$. For any $u \in S_1$, if $L_{u,u} \geq |\mathcal{C}_u|/4$, we have $|\mathcal{C}_u|^2 - \sum_w L_{u,w}^2 \geq L_{u,u}L_u \geq |\mathcal{C}_u|L_u/4$. If $L_{u,u} < \frac{1}{4}|\mathcal{C}_u|$ we have $|\mathcal{C}_u|^2 - \sum_w L_{u,w}^2 \geq \frac{3}{8}|\mathcal{C}_u|^2 \geq |\mathcal{C}_u|L_u/4$ as well. This leads to

$$\alpha \geq \frac{1}{2}\sum_{u \in S_1}\left[|\mathcal{C}_u|^2 - \sum_w L_{u,w}^2\right] \geq \frac{1}{8}\sum_{u \in S_1}|\mathcal{C}_u|L_u.$$

For any $u \in S_2$ there exists a $v \neq u$ such that $L_{u,v} > \frac{3}{4}|\mathcal{C}_u|$. We must have $L_{u,u}+L_{v,v} \geq L_{u,v}+L_{v,u}$ otherwise $\|\pi - Z^*\|_1 = \ell(\pi, Z^*)$ does not hold since we can switch the u-th and v-th columns of π to make $\|\pi - Z^*\|_1$ smaller. Consequently, we have $L_{v,v} \geq L_u/2$. So we have $\sum_{u' \neq u} \sum_w L_{u,w} L_{u',w} \geq L_{u,v} L_{v,v} \geq 3|\mathcal{C}_u| L_u/8$. Then we have

$$\gamma \geq \frac{1}{2} \sum_{u \in S_2} \sum_{u' \neq u} \sum_w L_{u,w} L_{u',w} \geq \frac{3}{8} \sum_{u \in S_2} |\mathcal{C}_u| L_u.$$

Thus,

$$\alpha + \gamma \geq \frac{1}{16} \sum_u |\mathcal{C}_u| L_u \geq \frac{\rho n}{16k} \sum_u L_u \geq \frac{\rho n}{16k} \|\pi - Z^*\|_1 = \frac{\rho n}{16k} \ell(\pi, Z^*).$$

□

C.3. Statements and Proofs of Lemmas and Propositions for Theorem 4.2.

Lemma C.7. *Let $X \sim Beta(\alpha, \beta)$ where $\alpha = n^2 p$ and $\beta = n^2(1-p)$ with $p = o(1)$. Let $\eta = o(1)$. Then we have*

$$\mathbb{P}(|X - p| \geq \eta p) \leq \exp(-\eta^2 n^2 p/2).$$

Proof. Note X has the same distribution as $Y/(Y+Z)$ where Y and Z are independent χ^2 random variables with $Y \sim \chi^2(2\alpha)$ and $Z \sim \chi^2(2\beta)$. Then by using tail bound of χ^2 distribution (i.e., Proposition C.6)

$$\begin{aligned}\mathbb{P}(|X - p| \geq \eta p) &\leq \mathbb{P}(|Y - 2n^2 p| \geq 2\eta n^2 p) + \mathbb{P}(|Y + Z - 2n^2| \geq \eta n^2) \\ &\leq 2\exp(-\eta^2 n^2 p/4) + 2\exp(-\eta^2 n^2/16) \\ &\leq \exp(-\eta^2 n^2 p/2).\end{aligned}$$

□

Proposition C.6. *Let $X \sim \chi^2(k)$ we have*

$$\mathbb{P}\Big(|X - k| \geq kt\Big) \leq 2\exp(-kt^2/8), \forall t \in (0,1).$$

Proof. See Lemma 1 of [4]. □

APPENDIX D: GENERAL DERIVATIONS OF CAVI FOR VARIATIONAL INFERENCE

In this section, we provide the derivation from Equation (3) to Equation (4). First we have

(56)

$$\begin{aligned} \mathrm{KL}(\mathbf{q}(x)\|\mathbf{p}(x|y)) &= \mathbb{E}_{\mathbf{q}(x)}\Big[\log\frac{\mathbf{q}(x)}{\mathbf{p}(x|y)}\Big] \\ &= \mathbb{E}_{\mathbf{q}(x)}[\log\mathbf{q}(x)] - \mathbb{E}_{\mathbf{q}(x)}[\log\mathbf{p}(x|y)] \\ &= \mathbb{E}_{\mathbf{q}(x)}[\log\mathbf{q}(x)] - \mathbb{E}_{\mathbf{q}(x)}[\log\mathbf{p}(x,y)] + \log\mathbf{p}(y) \\ &= -(\mathbb{E}_{\mathbf{q}(x)}[\log\mathbf{p}(x,y)] - \mathbb{E}_{\mathbf{q}(x)}[\log\mathbf{q}(x)]) + \log\mathbf{p}(y) \\ &= -\left[\mathbb{E}_{\mathbf{q}(x)}[\log\mathbf{p}(y|x)] - \mathrm{KL}(\mathbf{q}(x)\|\mathbf{p}(x))\right] + \log\mathbf{p}(y). \end{aligned}$$

Thus, to minimize $\mathrm{KL}(\mathbf{q}(x)\|\mathbf{p}(x|y))$ w.r.t. $\mathbf{q}(x)$ is equivalent to maximize $\mathbb{E}_{\mathbf{q}(x)}[\log\mathbf{p}(y|x)] - \mathrm{KL}(\mathbf{q}(x)\|\mathbf{p}(x))$.

Recall we have independence under both $\mathbf{p}$ and $\mathbf{q}$ for $\{x_i\}_{i=1}^n$. For simplicity, denote x_{-i} to be $\{x_j\}_{j\neq i}$ and $\mathbf{q}_{-i}$ to be $\prod_{j\neq i}\mathbf{q}_j$. We have the decomposition

$$\begin{aligned} b_i(\mathbf{q}_i) &\triangleq \mathbb{E}_{\mathbf{q}(x)}[\log\mathbf{p}(x,y)] - \mathbb{E}_{\mathbf{q}(x)}[\log\mathbf{q}(x)] \\ &= \mathbb{E}_{\mathbf{q}_i}\left[\mathbb{E}_{\mathbf{q}_{-i}}\left[\log\mathbf{p}(x_i,x_{-i},y)\right]\right] - \mathbb{E}_{\mathbf{q}_i}\left[\mathbb{E}_{\mathbf{q}_{-i}}[\log\mathbf{q}(x_i,x_{-i})]\right] \\ &= \mathbb{E}_{\mathbf{q}_i}\left[\mathbb{E}_{\mathbf{q}_{-i}}\left[\log\mathbf{p}(x_i|x_{-i},y)\right]\right] - \mathbb{E}_{\mathbf{q}_i}\left[\log\mathbf{q}_i(x_i)\right] + \mathrm{const} \\ &= -\mathbb{E}_{\mathbf{q}_i}\log\frac{\log\mathbf{q}_i(x_i)}{c^{-1}\exp\left[\mathbb{E}_{\mathbf{q}_{-i}}\left[\log\mathbf{p}(x_i|x_{-i},y)\right]\right]} + \mathrm{const}, \end{aligned}$$

where the constant includes all terms not depending on x_i and $c = \sum_{x_i}\exp\left[\mathbb{E}_{\mathbf{q}_{-i}}\left[\log\mathbf{p}(x_i|x_{-i},y)\right]\right]$ which is also independent of x_i. It is obvious that to solve Equation (3) is equivalent to

$$\begin{aligned} \hat{\mathbf{q}}_i &= \arg\max_{\mathbf{q}_i} b_i(\mathbf{q}_i) \\ &= \arg\min_{\mathbf{q}_i}\mathrm{KL}\left[\mathbf{q}_i\|c^{-1}\exp\left[\mathbb{E}_{\mathbf{q}_{-i}}\left[\log\mathbf{p}(x_i|x_{-i},y)\right]\right]\right]. \end{aligned}$$

Immediately we have $\hat{\mathbf{q}}_i(x_i) = c^{-1}\exp\left[\mathbb{E}_{\mathbf{q}_{-i}}\left[\log\mathbf{p}(x_i|x_{-i},y)\right]\right]$. Or we may write it as

$$\hat{\mathbf{q}}_i(x_i) \propto \exp\left[\mathbb{E}_{\mathbf{q}_{-i}}\left[\log\mathbf{p}(x_i|x_{-i},y)\right]\right].$$

REFERENCES

[1] Peter Chin, Anup Rao, and Van Vu. Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *COLT*, pages 391–423, 2015.

[2] Alexandre Grothendieck. Résumé de la théorie métrique des produits tensoriels topologiques. *Resenhas do Instituto de Matemática e Estatística da Universidade de São Paulo*, 2(4):401–481, 1996.

[3] Olivier Guédon and Roman Vershynin. Community detection in sparse networks via Grothendiecks inequality. *Probability Theory and Related Fields*, 165(3-4):1025–1049, 2016.

[4] Beatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, pages 1302–1338, 2000.

[5] Jing Lei and Alessandro Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015.

[6] Anderson Y Zhang and Harrison H Zhou. Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, 44(5):2252–2280, 2016.