Supplementary material for "A Theoretical Case Study of Structured Variational Inference for Community Detection"

This supplementary document contains detailed proofs and derivation of theoretical results presented in the main paper "A Theoretical Case Study of Structured Variational Inference for Community Detection", and additional experimental results. In particular, Section A contains the detailed derivation of updates of the Variational Inference with Pairwise Structure (VIPS) algorithm. Section B contains detailed proofs of the theoretical results presented in the main paper. Section C contains details on how to generalize VIPS to $K = 2$ with unbalanced community sizes and $K = 3$ with equal sized communities, and experimental results on robustness to parameter mis-specification. Section D contains additional experimental results and figures.

A Detailed Derivation of the Updates of VIPS

In the main paper Eq. (5), the Evidence Lower BOund (ELBO) for pairwise structured variational inference is

$$\begin{aligned}\mathcal{L}(Q;\pi,B) =&\frac{1}{2}\mathbb{E}_Q\sum_{i\neq j,a,b} Z_{ia}Z_{jb}(A_{ij}^{zz}\alpha_{ab}+f(\alpha_{ab}))+\frac{1}{2}\mathbb{E}_Q\sum_{i\neq j,a,b} Y_{ia}Y_{jb}(A_{ij}^{yy}\alpha_{ab}+f(\alpha_{ab}))\\ &+\mathbb{E}_Q\sum_{i\neq j,a,b} Z_{ia}Y_{jb}(A_{ij}^{zy}\alpha_{ab}+f(\alpha_{ab}))+\mathbb{E}_Q\sum_{i,a,b} Z_{ia}Y_{ib}(A_{ii}^{zy}\alpha_{ab}+f(\alpha_{ab}))\\ &-\sum_{i=1}^{m}\mathrm{KL}(Q(z_i,y_i)||P(z_i)P(y_i))\end{aligned}$$

where $\alpha_{ab} = \log(B_{ab}/(1-B_{ab}))$ and $f(\alpha) = -\log(1+e^{\alpha})$. Denote the first four terms in ELBO as T_1, T_2, T_3, T_4, where T_1, T_2 correspond to the likelihood of the blocks A^{zz} and A^{yy} in the adjacency matrix, T_3 corresponds to the likelihood of $(z_i, y_j), i \neq j$ and T_4 corresponds to (z_i, y_i). Plugging in the marginal density of the independent nodes in T_1, T_2, T_3 and joint density of the dependent nodes in T_4, we have

$$\begin{aligned}T_1 =&\frac{1}{2}\sum_{i\neq j}\Big\{[(1-\phi_i)(1-\phi_j)+\phi_i\phi_j](A_{ij}^{zz}\log\frac{p}{1-p}+\log(1-p))+\\ &[(1-\phi_i)\phi_j+\phi_i(1-\phi_j)](A_{ij}^{zz}\log\frac{q}{1-q}+\log(1-q))\Big\}\end{aligned}\tag{A.1}$$

$$\begin{aligned}T_2 =&\frac{1}{2}\sum_{i\neq j}\Big\{[(1-\xi_i)(1-\xi_j)+\xi_i\xi_j](A_{ij}^{yy}\log\frac{p}{1-p}+\log(1-p))+\\ &[(1-\xi_i)\xi_j+\xi_i(1-\xi_j)](A_{ij}^{yy}\log\frac{q}{1-q}+\log(1-q))\Big\}\end{aligned}\tag{A.2}$$

$$\begin{aligned}T_3 =&\sum_{i\neq j}\Big\{[(1-\phi_i)(1-\xi_j)+\phi_i\xi_j](A_{ij}^{zy}\log\frac{p}{1-p}+\log(1-p))+\\ &[(1-\phi_i)\xi_j+\phi_i(1-\xi_j)](A_{ij}^{zy}\log\frac{q}{1-q}+\log(1-q))\Big\}\end{aligned}\tag{A.3}$$

$$\begin{aligned}T_4 =&\sum_{i}\Big\{(1-\psi_i^{01}-\psi_i^{10})(A_{ii}^{zy}\log\frac{p}{1-p}+\log(1-p))+\\ &(\psi_i^{01}+\psi_i^{10})(A_{ii}^{zy}\log\frac{q}{1-q}+\log(1-q))\Big\}\end{aligned}\tag{A.4}$$

The KL regularization term (6) is

$$\mathrm{KL}(Q(z_i,y_i)||P(z_i)P(y_i)) =\psi_i^{00}\log\frac{\psi_i^{00}}{(1-\pi)^2}+\psi_i^{01}\log\frac{\psi_i^{01}}{\pi(1-\pi)}+$$

$$\psi_i^{10}\log\frac{\psi_i^{10}}{\pi(1-\pi)}+\psi_i^{11}\log\frac{\psi_i^{11}}{\pi^2}$$
$$=\sum_{0\le c,d\le 1}\psi_i^{cd}\log\frac{\psi_i^{cd}}{\pi^c\pi^d(1-\pi)^{1-c}(1-\pi)^{1-d}}$$

To take the derivative of $\mathcal{L}(Q;\pi,B)$ with respect to $\psi_i^{cd}, cd\neq 0$, we first have the derivative of the KL term

$$\frac{\partial}{\partial\psi_i^{cd}}\mathrm{KL}(Q(z_i,y_i)||P(z_i)P(y_i))=\log\frac{\psi_i^{cd}}{\pi^{c+d}(1-\pi)^{2-c-d}}-\log\frac{\psi_i^{00}}{(1-\pi)^2} \tag{A.5}$$
$$=\log\frac{\psi_i^{cd}}{1-\psi_i^{01}-\psi_i^{10}-\psi_i^{11}}\qquad(\pi=\frac{1}{2}) \tag{A.6}$$

Denote the right hand side of Eq. (A.6) as $\theta_i^{cd}\coloneqq\log\frac{\psi_i^{cd}}{1-\psi_i^{01}-\psi_i^{10}-\psi_i^{11}}$. For the reconstruction terms, denoting $T(a,p)\coloneqq a\log(\frac{p}{1-p})+\log(1-p)$ for simplicity, the derivative can be computed as

$$\begin{aligned}\frac{\partial}{\partial\psi_i^{10}}(\sum T_k)=&\sum_{j,j\neq i}\Big[(2\phi_j-1)T(A_{ij}^{zz},p)-(2\phi_j-1)T(A_{ij}^{zz},q)\Big]+\\&\sum_{j,j\neq i}\Big[(2\xi_j-1)T(A_{ij}^{zy},p)-(2\xi_j-1)T(A_{ij}^{zy},q)\Big]+\\&\Big[-T(A_{ii}^{zy},p)+T(A_{ii}^{zy},q)\Big]\end{aligned} \tag{A.7}$$

$$\begin{aligned}\frac{\partial}{\partial\psi_i^{01}}(\sum T_k)=&\sum_{j,j\neq i}\Big[(2\xi_j-1)T(A_{ij}^{yy},p)-(2\xi_j-1)T(A_{ij}^{yy},q)\Big]+\\&\sum_{j,j\neq i}\Big[(2\phi_j-1)T(A_{ji}^{zy},p)-(2\phi_j-1)T(A_{ji}^{zy},q)\Big]+\\&\Big[-T(A_{ii}^{zy},p)+T(A_{ii}^{zy},q)\Big]\end{aligned} \tag{A.8}$$

$$\begin{aligned}\frac{\partial}{\partial\psi_i^{11}}(\sum T_k)=&\sum_{j,j\neq i}\Big[(2\phi_j-1)T(A_{ij}^{zz},p)-(2\phi_j-1)T(A_{ij}^{zz},q)\Big]+\\&\sum_{j,j\neq i}\Big[(2\xi_j-1)T(A_{ij}^{yy},p)-(2\xi_j-1)T(A_{ij}^{yy},q)\Big]+\\&\sum_{j,j\neq i}\Big[(2\xi_j-1)T(A_{ij}^{zy},p)-(2\xi_j-1)T(A_{ij}^{zy},q)\Big]+\\&\sum_{j,j\neq i}\Big[(2\phi_j-1)T(A_{ji}^{zy},p)-(2\phi_j-1)T(A_{ji}^{zy},q)\Big]\end{aligned} \tag{A.9}$$

Setting the derivatives to 0 we get the update for θ as (10), (9), (11).

B Proofs of Main Results

To prove Theroem 1, we first need a few lemmas. First we have the following lemma for the parameters p, q and λ.

Lemma A.1. *If $p\asymp q\asymp\rho_n, \rho_n\to 0$ and $p-q=\Omega(\rho_n)$, then*

$$\lambda-q=\Omega(\rho_n)>0, \tag{B.1}$$

$$\frac{p+q}{2} - \lambda = \Omega(\rho_n) > 0. \tag{B.2}$$

Proof. The proof follows from Proposition 2 in (Sarkar et al., 2019). □

In the proof, we utilize the spectral property of the population matrix P and generalize it to the finite sample case by bounding the term related to the residual $R = A - P$. We use Berry-Esseen Theorem to bound the residual terms conditioning on u.

Lemma A.2 (Berry-Esseen bound)**.** *Define*

$$r_i = \sum_{j=1}^{n} (A_{ij} - P_{ij})(u(j) - \frac{1}{2}), \tag{B.3}$$

where u and A are independent.

$$\sup_{x\in\mathbb{R}} |P\left(r_i/\sigma_u \leq x \mid u\right) - \Phi(x)| \leq \frac{C\rho_u}{\sigma_u^3},$$

where C is a general constant, $\Phi(\cdot)$ is the CDF of standard Gaussian, ρ_u and σ_u depend on u.

Proof. Since r_i is the sum of independent, mean zero random variables, the sum of the conditional variances is

$$\sigma_u^2 = \mathrm{Var}(r_i|u) = p(1-p)\sum_{i\in G_1}(u(i) - \frac{1}{2})^2 + q(1-q)\sum_{i\in G_2}(u(i) - \frac{1}{2})^2,$$

and the sum of the conditional absolute third central moments is

$$\rho_u = p(1-p)(1-2p+2p^2)\sum_{i\in G_1}|u(i) - \frac{1}{2}|^3 + q(1-q)(1-2q+2q^2)\sum_{i\in G_2}|u(i) - \frac{1}{2}|^3.$$

The desired bound follows from the Berry-Esseen Theorem. □

The next lemma shows despite the fact that A introduces some dependency among r_i due to its symmetry, we can still treat r_i as almost iid.

Lemma A.3 (McDiarmid's Inequality)**.** *Let r_i be the noise defined in Lemma A.2 and let $h(r_i)$ be a bounded function with $\|h\|_\infty \leq M$. Then*

$$P\left(\left|\frac{2}{n}\sum_{i\in\mathcal{A}} h(r_i) - \mathbb{E}(h(r_i)|u)\right| > w \mid u\right) \leq \exp\left(-\frac{c_0 w^2}{nM}\right)$$

for some general constant c_0, provided $|\mathcal{A}| = \Theta_P(n)$.

Proof. The proof follows from Lemma 20 in (Sarkar et al., 2019). □

Lemma A.4. *Let r_i be defined as in Lemma A.2 and assume A and u are independent, we have $\sup_{i\in\mathcal{A}}|r_i| = O_P(\sqrt{n\rho_n \log n})$ if the index set $|\mathcal{A}| = \Theta_P(n)$.*

Proof. Since r_i is the sum of independent bounded random variables, for all i, $r_i = O_P(\sqrt{n\rho_n})$. By Hoeffding inequility, we know for all $t > 0$

$$P(|r_i| > t) \leq \exp(-\frac{t^2}{2n\rho_n})$$

and by the union bound

$$P(\sup_i |r_i| > t) \leq \exp(C\log n - \frac{t^2}{2n\rho_n})$$

For $\forall\epsilon > 0$, let $t = C_\epsilon\sqrt{n\rho_n \log n}$ with $n^{\frac{C_\epsilon^2}{2}-1} > 1/\epsilon$, then by definition $\sup_i |r_i| = O_P(\sqrt{n\rho_n \log n})$ □

Next we have a lemma ensuring the signal in the first iteration is not too small.

Lemma A.5 (Littlewood-Offord)**.** *Let* $s_1 = (p-\lambda)\sum_{i\in G_1}(u^{(0)}(i)-1/2) + (q-\lambda)\sum_{i\in G_2}(u^{(0)}(i)-1/2)$*,* $s_2 = (q-\lambda)\sum_{i\in G_1}(u^{(0)}(i)-1/2) + (p-\lambda)\sum_{i\in G_2}(u^{(0)}(i)-1/2)$*. Then*

$$P\left(|s_1| \le c\right) \le B\cdot\frac{c}{\rho_n\sqrt{n}}$$

for $c > 0$ *and* B *as constant. The same bound holds for* $|s_2|, |s_1 - s_2|$*.*

Proof. Noting that $2u^{(0)}(i) - 1 \in \{-1, 1\}$ each with probability 1/2, and Lemma A.1, this is a direct consequence of the Littlewood-Offord bound in (Erdös, 1945). □

Finally, we have the following upper and lower bound for some general update ϕ_i.

Lemma A.6. *Assume* ϕ_i *has the update form* $\phi_i = (a+e^{4t(s+r_i)})/(b+e^{4t(s+r_i)})$ *for* $i \in [m]$*,* $b > a > 0$ *and* $b-a$*,* $(b-a)/b$ *are of constant order.* r_i *is defined as in Lemma A.2. Let set* $\mathcal{A} \subset [m]$*, with* $\Delta > 0$*, we have*

$$\begin{aligned}\sum_{i\in\mathcal{A}}\phi_i \ge& |\mathcal{A}| - \frac{b-a}{b}|\mathcal{A}|\Phi(\frac{-s+\Delta}{\sigma_u}) - C'|\mathcal{A}|\frac{\rho_u}{\sigma_u^3} - C''|\mathcal{A}|e^{-4t\Delta} - O_P(\sqrt{|\mathcal{A}|}),\\ \sum_{i\in\mathcal{A}}\phi_i \le& |\mathcal{A}| - \frac{b-a}{b}|\mathcal{A}|\Phi(\frac{-s-\Delta}{\sigma_u}) + C'|\mathcal{A}|\frac{\rho_u}{\sigma_u^3} + |\mathcal{A}|e^{-4t\Delta} + O_P(\sqrt{|\mathcal{A}|}).\end{aligned}$$

Proof. Define the set $J^+ = \{i : r_i > -s+\Delta\}$, $\Delta \ge 0$. For $i \in \mathcal{A}\cap J^+$

$$\phi_i = \frac{a+e^{4t(s+r_i)}}{b+e^{4t(s+r_i)}} \ge \frac{a+e^{4t\Delta}}{b+e^{4t\Delta}} \ge 1-(b-a)e^{-4t\Delta}$$

For $i \in (\mathcal{A}\cap J^+)^c$, $\phi_i \ge a/b$, therefore

$$\begin{aligned}\sum_{i\in\mathcal{A}}\phi_i \ge& |\mathcal{A}\cap J^+|(1-(b-a)e^{-4t\Delta}) + \frac{a}{b}(|\mathcal{A}| - |\mathcal{A}\cap J^+|)\\ =& |\mathcal{A}\cap J^+|(\frac{b-a}{b} - (b-a)e^{-4t\Delta}) + \frac{a}{b}|\mathcal{A}|\end{aligned}$$

By Lemmas A.2 and A.3, we have

$$\begin{aligned}|\mathcal{A}\cap J^+| =& \sum_{i\in\mathcal{A}}\mathbf{1}[r_i > -s+\Delta]\\ =& |\mathcal{A}|\cdot P(r_i > -s+\Delta) + O_P(\sqrt{|\mathcal{A}|})\\ \ge& |\mathcal{A}|\cdot(1-\Phi(\frac{-s+\Delta}{\sigma_u}) - C_0\frac{\rho_u}{\sigma_u^3}) + O_P(\sqrt{|\mathcal{A}|}).\end{aligned}$$

Combining the above,

$$\sum_{i\in\mathcal{A}}\phi_i \ge |\mathcal{A}| - \frac{b-a}{b}|\mathcal{A}|\Phi(\frac{-s+\Delta}{\sigma_u}) - C'|\mathcal{A}|\frac{\rho_u}{\sigma_u^3} - C''|\mathcal{A}|e^{-4t\Delta} - O_P(\sqrt{|\mathcal{A}|})$$

Similarly, define the set $J^- = \{i : r_i < -s-\Delta\}$, $\Delta \ge 0$. For $i \in \mathcal{A}\cap J^-$,

$$\phi_i = \frac{a+e^{4t(s+r_i)}}{b+e^{4t(s+r_i)}} \le \frac{a+e^{-4t\Delta}}{b+e^{-4t\Delta}} \le \frac{a}{b} + e^{-4t\Delta}$$

For $i \in (\mathcal{A}\cap J^-)^c$, $\phi_i \le 1$, so

$$\sum_{i\in\mathcal{A}}\phi_i \le |\mathcal{A}\cap J^-|(\frac{a}{b} + e^{-4t\Delta}) + (|\mathcal{A}| - |\mathcal{A}\cap J^-|)$$

$$=|\mathcal{A}| - |\mathcal{A}\cap J^-|(1-\frac{a}{b}-e^{-4t\Delta}) + O_P(\sqrt{|\mathcal{A}|})$$

By Lemmas A.2 and A.3,

$$|\mathcal{A}\cap J^-| \geq |\mathcal{A}|\cdot(\Phi(\frac{-s-\Delta}{\sigma_u}) - C_0\frac{\rho_u}{\sigma_u^3}) - O_P(\sqrt{|\mathcal{A}|})$$

so

$$\sum_{i\in\mathcal{A}}\phi_i \leq |\mathcal{A}| - \frac{b-a}{b}|\mathcal{A}|\Phi(\frac{-s-\Delta}{\sigma_u}) + C'|\mathcal{A}|\frac{\rho_u}{\sigma_u^3} + |\mathcal{A}|e^{-4t\Delta} + O_P(\sqrt{|\mathcal{A}|})$$

□

Proof of Theorem 1. Throughout the proof, we assume A has self-loops for convenience, which does not affect the asymptotic results.

Analysis of the first iteration in the first meta iteration:

For random initialized $u^{(0)}$, the initial signal $|\langle u^{(0)}, v_2\rangle| = O_P(\sqrt{n})$. Using the graph split $A^{(0)}$, we write the update of θ^{10} as

$$\begin{aligned}\theta^{10} =& 4t([6(A^{(0)})^{zz}, 6(A^{(0)})^{zy}] - \lambda J)(u^{(0)} - \frac{1}{2}\mathbf{1}_n)\\ =& \underbrace{4t([P^{zz}, P^{zy}] - \lambda J)(u^{(0)} - \frac{1}{2}\mathbf{1}_n)}_{\text{signal}} + \underbrace{4t[6(A^{(0)})^{zz} - P^{zz}, 6(A^{(0)})^{zy} - P^{zy}](u^{(0)} - \frac{1}{2}\mathbf{1}_n)}_{\text{noise}},\end{aligned} \tag{B.4}$$

where P is the population matrix of A. Denote $R^{(0)} = 6A^{(0)} - P$ and $r^{(0)} = [(R^{(0)})^{zz}, (R^{(0)})^{zy}](u^{(0)} - \frac{1}{2}\mathbf{1})$ Since P has singular value decomposition as $P = \frac{p+q}{2}\mathbf{1}_n\mathbf{1}_n^T + \frac{p-q}{2}v_2v_2^T$, the signal part is blockwise constant and we can write

$$\theta^{10} = 4t(s_1\mathbf{1}_{C_1} + s_2\mathbf{1}_{C_2} + r^{(0)}), \tag{B.5}$$

where

$$\begin{aligned}s_1 =& (\frac{p+q}{2} - \lambda)(\langle u^{(0)}, \mathbf{1}_n\rangle - m) + (\frac{p-q}{2})\langle u^{(0)}, v_2\rangle\\ s_2 =& (\frac{p+q}{2} - \lambda)(\langle u^{(0)}, \mathbf{1}_n\rangle - m) - (\frac{p-q}{2})\langle u^{(0)}, v_2\rangle\end{aligned} \tag{B.6}$$

By (12), since we initialize with $\theta^{01}, \theta^{11} = 0$, the marginal probabilities are updated as

$$\phi_1^{(1)} = \frac{1+e^{\theta^{10}}}{3+e^{\theta^{10}}}, \quad \xi_1^{(1)} = \frac{2}{3+e^{\theta^{10}}} \tag{B.7}$$

Next we show the signal $|\langle u, v_2\rangle|$ increases from $O_P(\sqrt{n})$ to $\Omega_P(n\sqrt{\rho_n})$. (We omit the superscript on logits s,x and y now for simplicity.) Since

$$\langle u_1^{(1)}, v_2\rangle = \langle \phi_{1i}^{(1)}, v_{21}\rangle + \langle \xi_{1i}^{(1)}, v_{22}\rangle = \sum_{i\in C_1}\phi_i^{(1)} - \sum_{i\in C_2}\phi_i^{(1)} + \langle \xi^{(1)}, v_{22}\rangle$$

we use Lemma A.6 to bound $\sum_{i\in C_1}\phi_i^{(1)}$ and $\sum_{i\in C_2}\phi_i^{(1)}$. Since s_1 and s_2 depends on $u^{(0)}$, we consider two cases conditioning on $u^{(0)}$.

Case 1: $s_1 > s_2$. By Lemma A.6, let $\Delta = \frac{1}{4}(s_1 - s_2)$ with $\mathcal{A} = C_1, C_2$, $(a, b) = (1, 3)$, conditioning on $u^{(0)}$,

$$\sum_{i\in C_1}\phi_{1i}^{(1)} \geq \frac{n}{6}(1 - \Phi(-\frac{s_1 - \frac{1}{4}(s_1 - s_2)}{\sigma_u})) + \frac{n}{12} - C'n\frac{\rho_u}{\sigma_u^3} - C''ne^{-t(s_1-s_2)} - O_P(\sqrt{n}),$$

$$\sum_{i\in C_2}\phi_{1i}^{(1)} \le \frac{n}{4} - \frac{n}{6}\Phi(-\frac{s_2+\frac{1}{4}(s_1-s_2)}{\sigma_u}) + C'n\frac{\rho_u}{\sigma_u^3} + C''ne^{-t(s_1-s_2)} + O_P(\sqrt{n}),$$

where the $O_P(\sqrt{n})$ term can be made uniform in $u^{(0)}$. So we have

$$\begin{aligned}\langle \phi_1^{(1)}, v_{21}\rangle \ge & \frac{n}{6}(\Phi(-\frac{s_2+\frac{1}{4}(s_1-s_2)}{\sigma_u}) - \Phi(-\frac{s_1-\frac{1}{4}(s_1-s_2)}{\sigma_u})) \\ & - C'n\frac{\rho_u}{\sigma_u^3} - C''ne^{-t(s_1-s_2)} - O_P(\sqrt{n}) \\ \ge & \frac{n}{6\sqrt{2\pi}}(\frac{s_1-s_2}{2\sigma_u})\exp\big(-\frac{s_1^2\vee s_2^2}{2\sigma_u^2}\big) \\ & - C'n\frac{\rho_u}{\sigma_u^3} - C''ne^{-t(s_1-s_2)} - O_P(\sqrt{n}). \end{aligned} \tag{B.8}$$

Here to approximate the CDF Φ, we have used

$$\begin{aligned}|\Phi(x)-1/2| &= \frac{1}{\sqrt{2\pi}}\int_0^{|x|} e^{-u^2/2}du \\ &\ge \frac{|x|}{\sqrt{2\pi}}e^{-x^2/2}. \end{aligned}\tag{B.9}$$

Case 2: $s_1 < s_2$. The same analysis applies with s_1 and s_2 interchanged.

Combining *Case 1* and *Case 2*, for any given $u^{(0)}$,

$$\begin{aligned}|\langle \phi_1^{(1)}, v_{21}\rangle| \ge & \frac{n}{6\sqrt{2\pi}}(\frac{|s_1-s_2|}{2\sigma_u})\exp\big(-\frac{s_1^2\vee s_2^2}{2\sigma_u^2}\big) \\ & - C'n\frac{\rho_u}{\sigma_u^3} - C''ne^{-t|s_1-s_2|} - O_P(\sqrt{n}). \end{aligned}\tag{B.10}$$

We note that $|s_1|$, $|s_2|$, $|s_1-s_2|$ are of order $\Omega_P(\sqrt{n}\rho_n)$ by Lemma A.5. Also $\sigma_u^2, \rho_u \asymp n\rho_n$, $e^{-4t|s_1-s_2|} = \exp(-\Omega(\rho_n\sqrt{n}))$. We can conclude that $|\langle\phi_1^{(1)}, v_{21}\rangle| = \Omega_P(n\sqrt{\rho_n})$.

For $\langle \xi_1^{(1)}, v_{22}\rangle$ we have

$$\begin{aligned}|\langle \xi_1^{(1)}, v_{22}\rangle| =& \Big|\sum_{i\in C_1'}\xi_i^{(1)} - \sum_{i\in C_2'}\xi_i^{(1)}\Big| = \Big|\sum_{i\in C_2'}\phi_i^{(1)} - \sum_{i\in C_1'}\phi_i^{(1)} + |C_1'| - |C_2'|\Big| \\ =& O_P(\sqrt{n})\end{aligned}$$

Therefore we have $|\langle u_1^{(1)}, v_2\rangle| = \Omega_P(n\sqrt{\rho_n})$. By (B.7), $\langle u_1^{(1)}, \mathbf{1}\rangle - m = 0$.

Due to the symmetry in s_1 and s_2, WLOG in the following analysis, we assume $\langle u_1^{(1)}, v_2\rangle > 0$ (equivalently $s_1 > s_2$).

Analysis of the second iteration in the first meta iteration:

Similar to (B.4), we can write

$$\begin{aligned}\theta^{01} =& 4t([6(A^{(1)})^{yz}, 6(A^{(1)})^{yy}] - \lambda J)(u_1^{(1)} - \frac{1}{2}\mathbf{1}_n) \\ =& \underbrace{4t([P^{yz}, P^{yy}] - \lambda J)(u_1^{(1)} - \frac{1}{2}\mathbf{1}_n)}_{\text{signal}} + \underbrace{4t(R^{(1)})^{yz}(\phi_1^{(1)} - \frac{1}{2}\mathbf{1}_m) + 4t(R^{(1)})^{yy}(\xi_1^{(1)} - \frac{1}{2}\mathbf{1}_m)}_{\text{noise} := 4tr_i^{(1)}}.\end{aligned}$$

Noting the signal part is blockwise constant, we have

$$\theta^{01} = 4t(x_1\mathbf{1}_{C_1'} + x_2\mathbf{1}_{C_2'} + r^{(1)}),$$

where

$$
\begin{aligned}
x_1 =& (\frac{p+q}{2} - \lambda)(\langle u_1^{(1)}, \mathbf{1}_n\rangle - m) + (\frac{p-q}{2})\langle u_1^{(1)}, v_2\rangle \\
x_2 =& (\frac{p+q}{2} - \lambda)(\langle u_1^{(1)}, \mathbf{1}_n\rangle - m) - (\frac{p-q}{2})\langle u_1^{(1)}, v_2\rangle
\end{aligned}
$$

By (B.7), $\langle u_1^{(1)}, \mathbf{1}_n\rangle - m = 0$ and we have

$$
\begin{aligned}
x_1 =& (\frac{p-q}{2})\langle u_1^{(1)}, v_2\rangle, \\
x_2 =& -x_1.
\end{aligned}
$$

It follows then from the first iteration that $x_1, -x_2 = \Omega_P(n\rho_n^{3/2})$. The update for $u_2^{(1)}$ is

$$
\phi_2^{(1)} = \frac{1 + e^{\theta^{10}}}{2 + e^{\theta^{10}} + e^{\theta^{01}}}, \quad \xi_2^{(1)} = \frac{1 + e^{\theta^{01}}}{2 + e^{\theta^{10}} + e^{\theta^{01}}} \tag{B.11}
$$

Since the signal part of θ^{10} and θ^{01} are blockwise constant on C_1, C_2 and C_1', C_2' respectively, $\langle u_2^{(1)}, v_2\rangle$ can be calculated as

$$
\begin{aligned}
\langle \phi_2^{(1)}, v_{21}\rangle =& \sum_{i\in C_{11}} \frac{1 + e^{4t(s_1 + r_i^{(0)})}}{2 + e^{4t(s_1 + r_i^{(0)})} + e^{4t(x_1 + r_i^{(1)})}} + \sum_{i\in C_{12}} \frac{1 + e^{4t(s_1 + r_i^{(0)})}}{2 + e^{4t(s_1 + r_i^{(0)})} + e^{4t(x_2 + r_i^{(1)})}} \\
&- \sum_{i\in C_{21}} \frac{1 + e^{4t(s_2 + r_i^{(0)})}}{2 + e^{4t(s_2 + r_i^{(0)})} + e^{4t(x_1 + r_i^{(1)})}} - \sum_{i\in C_{22}} \frac{1 + e^{4t(s_2 + r_i^{(0)})}}{2 + e^{4t(s_2 + r_i^{(0)})} + e^{4t(x_2 + r_i^{(1)})}}
\end{aligned}
$$

$$
\begin{aligned}
\langle \xi_2^{(1)}, v_{22}\rangle =& \sum_{i\in C_{11}} \frac{1 + e^{4t(x_1 + r_i^{(1)})}}{2 + e^{4t(s_1 + r_i^{(0)})} + e^{4t(x_1 + r_i^{(1)})}} + \sum_{i\in C_{21}} \frac{1 + e^{4t(x_1 + r_i^{(1)})}}{2 + e^{4t(s_2 + r_i^{(0)})} + e^{4t(x_1 + r_i^{(1)})}} \\
&- \sum_{i\in C_{12}} \frac{1 + e^{4t(x_2 + r_i^{(1)})}}{2 + e^{4t(s_1 + r_i^{(0)})} + e^{4t(x_2 + r_i^{(1)})}} - \sum_{i\in C_{22}} \frac{1 + e^{4t(x_2 + r_i^{(1)})}}{2 + e^{4t(s_2 + r_i^{(0)})} + e^{4t(x_2 + r_i^{(1)})}}
\end{aligned}
$$

In the case of $\langle u_1^{(1)}, v_2\rangle > 0$, we know , $s_1 > s_2$ and $x_1 > 0 > x_2$. We first show that $\langle \phi_2^{(1)}, v_{21}\rangle$ is positive by finding a lower bound for the summations over C_{12}, C_{21}, C_{22} (since the sum over C_{11} is always positive).

For the summation over C_{12}, note that $|x_2|$ dominates both s_1 and $r_i^{(0)}$, $r_i^{(1)}$ by Lemma A.4, we have

$$
\sum_{i\in C_{12}} \frac{1 + e^{4t(s_1 + r_i^{(0)})}}{2 + e^{4t(s_1 + r_i^{(0)})} + e^{4t(x_2 + r_i^{(1)})}} = \sum_{i\in C_{12}} \frac{1 + e^{4t(s_1 + r_i^{(0)})}}{2 + e^{4t(s_1 + r_i^{(0)})}} + n\exp(-\Omega_P(n\rho_n^{3/2})).
$$

To lower bound the first term, we use Lemma A.6 by first conditioning on $u^{(0)}$,

$$
\begin{aligned}
&\sum_{i\in C_{12}} \frac{1 + e^{4t(s_1 + r_i^{(0)})}}{2 + e^{4t(s_1 + r_i^{(0)})}} \\
\geq& \frac{n}{8}\left(1 - \frac{1}{2}\Phi(\frac{-s_1 + \Delta}{\sigma_u})\right) - C'n\frac{\rho_u}{\sigma_u^3} - C''ne^{-4t\Delta} - O_P(\sqrt{n})
\end{aligned} \tag{B.12}
$$

For the summation over C_{22},

$$
\sum_{i\in C_{22}} \frac{1 + e^{4t(s_2 + r_i^{(0)})}}{2 + e^{4t(s_2 + r_i^{(0)})} + e^{4t(x_2 + r_i^{(1)})}} \leq \sum_{i\in C_{22}} \frac{1 + e^{4t(s_2 + r_i^{(0)})}}{2 + e^{4t(s_2 + r_i^{(0)})}}
$$

$$\leq \frac{n}{8}\left(1-\frac{1}{2}\Phi(\frac{-s_2-\Delta}{\sigma_u})\right)+C'n\frac{\rho_u}{\sigma_u^3}+C''ne^{-4t\Delta}+O_P(\sqrt{n}) \tag{B.13}$$

For the summation over C_{21}, x_1 dominates s_2 and $r_i^{(0)}, r_i^{(1)}$ by Lemma A.4,

$$\sum_{i\in C_{21}}\frac{1+e^{4t(s_2+r_i^{(0)})}}{2+e^{4t(s_2+r_i^{(0)})}+e^{4t(x_1+r_i^{(1)})}}=n\exp(-\Omega_P(n\rho_n^{3/2})). \tag{B.14}$$

Combining (B.12) - (B.14), setting $\Delta=\frac{1}{4}(s_1-s_2)$, we have

$$\begin{aligned}\langle\phi_2^{(1)},v_{21}\rangle\geq&\frac{n}{8}\big[\frac{1}{2}\Phi(\frac{-s_2-\Delta}{\sigma_u})-\frac{1}{2}\Phi(\frac{-s_1+\Delta}{\sigma_u})\big]-C'n\frac{\rho_u}{\sigma_u^3}-C''ne^{-4t\Delta}-O_P(\sqrt{n})\\ \geq&\frac{n}{16}\frac{1}{\sqrt{2\pi}}(\frac{s_1-s_2}{\sigma_u})\exp\big(-\frac{s_1^2\vee s_2^2}{2\sigma_u^2}\big)-C'n\frac{\rho_u}{\sigma_u^3}-C''ne^{-t(s_1-s_2)}-O_P(\sqrt{n})\end{aligned}$$

by the same argument as (B.8). As before, we can see that

$$\langle\phi_2^{(1)},v_{21}\rangle=\Omega_P(n\sqrt{\rho_n})$$

For $\langle\xi_2^{(1)},v_{22}\rangle$, since $(1+e^x)/(2+e^x)\leq 1/2+e^x$, we have

$$\begin{aligned}&\sum_{i\in C_{12}}\frac{1+e^{4t(x_2+r_i^{(1)})}}{2+e^{4t(s_1+r_i^{(0)})}+e^{4t(x_2+r_i^{(1)})}}+\sum_{i\in C_{22}}\frac{1+e^{4t(x_2+r_i^{(1)})}}{2+e^{4t(s_2+r_i^{(0)})}+e^{4t(x_2+r_i^{(1)})}}\\ \leq&\frac{n}{8}+\sum_{i\in C_2'}e^{4t(x_2+r_i^{(1)})}+O_P(\sqrt{n})\\ \leq&\frac{n}{8}+O_P(\sqrt{n}).\end{aligned} \tag{B.15}$$

For the other two sums, we have

$$\begin{aligned}\sum_{i\in C_{11}}\frac{1+e^{4t(x_1+r_i^{(1)})}}{2+e^{4t(s_1+r_i^{(0)})}+e^{4t(x_1+r_i^{(1)})}}\geq&\frac{n}{8}-O_P(\sqrt{n})-n\exp(-\Omega_P(n\rho_n^{3/2})),\\ \geq&\frac{n}{8}-O_P(\sqrt{n})\end{aligned} \tag{B.16}$$

and

$$\sum_{i\in C_{21}}\frac{1+e^{4t(x_1+r_i^{(1)})}}{2+e^{4t(s_2+r_i^{(0)})}+e^{4t(x_1+r_i^{(1)})}}\geq\frac{n}{8}-O_P(\sqrt{n}) \tag{B.17}$$

Equations (B.15) - (B.17) imply

$$\langle\xi_2^{(1)},v_{22}\rangle\geq\frac{n}{8}-O_P(\sqrt{n}).$$

Therefore $\langle u_2^{(1)},v_2\rangle\geq n/8-O_P(\sqrt{n})$. Since by (B.11), $\phi_2^{(1)}=\mathbf{1}_m-\xi_2^{(1)}$, the inner product $\langle u_2^{(1)},\mathbf{1}\rangle-m=0$.

Analysis of the third iteration in the first meta iteration:

Similar to the previous two iterations, we can write

$$\theta^{11}=4t(y_1\mathbf{1}_{C_1}+y_2\mathbf{1}_{C_2}+y_1\mathbf{1}_{C_1'}+y_2\mathbf{1}_{C_2'}+r^{(2)}),$$

where

$$y_1=(\frac{p-q}{2})\langle u_2^{(1)},v_2\rangle,\quad y_2=-y_1$$

$$r^{(2)} = [(R^{(2)})^{zz}, (R^{(2)})^{zy}](u_2 - \frac{1}{2}\mathbf{1}_n) + [(R^{(2)})^{yz}, (R^{(2)})^{yy}](u_2^{(1)} - \frac{1}{2}\mathbf{1}_n).$$

It follows from the second iteration that $y_1, -y_2 = \Omega_P(n\rho_n)$.

The update for $u_3^{(1)}$ is

$$\phi_3^{(1)} = \frac{e^{\theta^{11}} + e^{\theta^{10}}}{1 + e^{\theta^{10}} + e^{\theta^{01}} + e^{\theta^{11}}}, \quad \xi_3^{(1)} = \frac{e^{\theta^{11}} + e^{\theta^{01}}}{1 + e^{\theta^{10}} + e^{\theta^{01}} + e^{\theta^{11}}} \tag{B.18}$$

The $\langle u_3^{(1)}, v_2 \rangle$ can be calculated as

$$\begin{aligned} &\langle u_3^{(1)}, v_2 \rangle = \\ &\sum_{i \in C_{11}} \frac{2e^{8t(y_1 + r_i^{(2)})} + e^{4t(x_1 + r_i^{(1)})} + e^{4t(s_1 + r_i^{(0)})}}{1 + e^{4t(s_1 + r_i^{(0)})} + e^{4t(x_1 + r_i^{(1)})} + e^{8t(y_1 + r_i^{(2)})}} + \sum_{i \in C_{12}} \frac{e^{4t(s_1 + r_i^{(0)})} - e^{4t(x_2 + r_i^{(1)})}}{1 + e^{4tr_i^{(2)}} + e^{4t(s_1 + r_i^{(0)})} + e^{4t(x_2 + r_i^{(1)})}} \\ &+ \sum_{i \in C_{21}} \frac{e^{4t(x_1 + r_i^{(1)})} - e^{4t(s_2 + r_i^{(0)})}}{1 + e^{4tr_i^{(2)}} + e^{4t(s_2 + r_i^{(0)})} + e^{4t(x_1 + r_i^{(1)})}} - \sum_{i \in C_{22}} \frac{2e^{8t(y_2 + r_i^{(2)})} + e^{4t(s_2 + r_i^{(0)})} + e^{4t(x_2 + r_i^{(1)})}}{1 + e^{4t(s_2 + r_i^{(0)})} + e^{4t(x_2 + r_i^{(1)})} + e^{8t(y_2 + r_i^{(2)})}} \end{aligned} \tag{B.19}$$

Using the order of the x terms and y terms and Lemma A.4, we can lower bound $\langle u_3^{(1)}, v_2 \rangle$ by

$$\begin{aligned} \langle u_3^{(1)}, v_2 \rangle &\geq \frac{n}{4} + \sum_{i \in C_{12}} \frac{e^{4t(s_1 + r_i^{(0)})}}{1 + e^{4tr_i^{(2)}} + e^{4t(s_1 + r_i^{(0)})}} + \frac{n}{8} - \sum_{i \in C_{22}} \frac{e^{4t(s_2 + r_i^{(0)})}}{1 + e^{4t(s_2 + r_i^{(0)})}} - O_P(\sqrt{n}) \\ &\geq \frac{n}{4} - O_P(\sqrt{n}). \end{aligned} \tag{B.20}$$

Next we bound $\langle u_3^{(1)}, \mathbf{1}_n \rangle - m$.

$$\begin{aligned} &\langle u_3^{(1)}, \mathbf{1}_n \rangle = \\ &\sum_{i \in C_{11}} \frac{2e^{8t(y_1 + r_i^{(2)})} + e^{4t(x_1 + r_i^{(1)})} + e^{4t(s_1 + r_i^{(0)})}}{1 + e^{4t(s_1 + r_i^{(0)})} + e^{4t(x_1 + r_i^{(1)})} + e^{8t(y_1 + r_i^{(2)})}} + \sum_{i \in C_{12}} \frac{2e^{4tr_i^{(2)}} + e^{4t(s_1 + r_i^{(0)})} + e^{4t(x_2 + r_i^{(1)})}}{1 + e^{4tr_i^{(2)}} + e^{4t(s_1 + r_i^{(0)})} + e^{4t(x_2 + r_i^{(1)})}} \\ &+ \sum_{i \in C_{21}} \frac{2e^{4tr_i^{(2)}} + e^{4t(x_1 + r_i^{(1)})} + e^{4t(s_2 + r_i^{(0)})}}{1 + e^{4tr_i^{(2)}} + e^{4t(s_2 + r_i^{(0)})} + e^{4t(x_1 + r_i^{(1)})}} + \sum_{i \in C_{22}} \frac{2e^{8t(y_2 + r_i^{(2)})} + e^{4t(s_2 + r_i^{(0)})} + e^{4t(x_2 + r_i^{(1)})}}{1 + e^{4t(s_2 + r_i^{(0)})} + e^{4t(x_2 + r_i^{(1)})} + e^{8t(y_2 + r_i^{(2)})}}, \end{aligned} \tag{B.21}$$

Then

$$\langle u_3^{(1)}, \mathbf{1}_n \rangle = \frac{3n}{8} + \sum_{i \in C_{12}} \frac{2e^{4tr_i^{(2)}} + e^{4t(s_1 + r_i^{(0)})}}{1 + e^{4tr_i^{(2)}} + e^{4t(s_1 + r_i^{(0)})}} + \sum_{i \in C_{22}} \frac{e^{4t(s_2 + r_i^{(0)})}}{1 + e^{4t(s_2 + r_i^{(0)})}} + O_P(\sqrt{n}),$$

$$\langle u_3^{(1)}, \mathbf{1}_n \rangle \geq \frac{3n}{8} - O_P(\sqrt{n}),$$

and

$$\begin{aligned} \langle u_3^{(1)}, \mathbf{1}_n \rangle \leq \frac{3n}{8} &+ \sum_{i \in C_{12}} \left(\frac{e^{4tr_i^{(2)}}}{1 + e^{4tr_i^{(2)}} + e^{4t(s_1 + r_i^{(0)})}} + \frac{e^{4tr_i^{(2)}} + e^{4t(s_1 + r_i^{(0)})}}{1 + e^{4tr_i^{(2)}} + e^{4t(s_1 + r_i^{(0)})}} \right) \\ &+ \sum_{i \in C_{22}} \frac{e^{4t(s_2 + r_i^{(0)})}}{1 + e^{4t(s_2 + r_i^{(0)})}} + O_P(\sqrt{n}) \\ \leq \frac{3n}{4} &+ O_P(\sqrt{n}) \end{aligned}$$

It follows then

$$-n/8 - O_P(\sqrt{n}) \le \langle u_3^{(1)}, \mathbf{1}_n\rangle - m \le n/4 + O_P(\sqrt{n}). \tag{B.22}$$

Analysis of the second meta iteration:

We first show that from the previous iteration, the signal $\langle u_3, v_2\rangle$ will always dominate $|\langle u_3, \mathbf{1}_n\rangle - m|$ which gives desired sign and magnitude of the logits. Then we show the algorithm converges to the true labels after the second meta iteration.

Using the same decomposition as (B.5),

$$\begin{aligned}
s_1^{(2)} =& (\frac{p+q}{2} - \lambda)(\langle u_3^{(1)}, \mathbf{1}_n\rangle - m) + \frac{p-q}{2}\langle u_3^{(1)}, v_2\rangle \qquad \text{(B.23)}\\
\ge& -\frac{n}{8}(\frac{p+q}{2} - \lambda) + \frac{n}{4}\cdot\frac{p-q}{2} - o_P(n\rho_n)\\
\ge& \frac{n}{8}(\lambda - q) - o_P(n\rho_n)\\
s_2^{(2)} =& (\frac{p+q}{2} - \lambda)(\langle u_3^{(1)}, \mathbf{1}_n\rangle - m) - \frac{p-q}{2}\langle u_3^{(1)}, v_2\rangle \qquad \text{(B.24)}\\
\le& \frac{n}{4}(\frac{p+q}{2} - \lambda) - \frac{n}{4}\cdot\frac{p-q}{2} + o_P(n\rho_n)\\
=& -\frac{n}{4}(\lambda - q) + o_P(n\rho_n), \qquad \text{(B.25)}
\end{aligned}$$

where we have used Lemma A.1.

After the first meta iteration, the logits satisfy

$$\begin{aligned}
&s_1^{(2)}, -s_2^{(2)} = \Omega_P(n\rho_n), \qquad x_1^{(1)}, -x_2^{(1)} = \Omega_P(n\rho_n^{\frac{3}{2}}),\\
&y_1^{(1)}, -y_2^{(1)} = \Omega_P(n\rho_n).
\end{aligned}$$

Here we have added the superscripts for the first meta iteration for clarity.

In the first iteration of the second meta iteration, $\langle u_1^{(2)}, v_2\rangle$ is computed as (B.19) with s_1 and s_2 replaced with $s_1^{(2)}$ and $s_2^{(2)}$ and the noise replaced accordingly. It is easy to see that

$$\langle u_1^{(2)}, v_2\rangle \ge \frac{3n}{8} - o_P(n). \tag{B.26}$$

Similarly from (B.21),

$$-\frac{n}{8} - o_P(n) \le \langle u_1^{(2)}, \mathbf{1}_n\rangle - m \le o_P(n). \tag{B.27}$$

The logits are updated as $(\frac{p+q}{2} - \lambda)(\langle u_1^{(2)}, \mathbf{1}_n\rangle - m) \pm \frac{p-q}{2}\langle u_1^{(2)}, v_2\rangle$, so

$$x_1^{(2)}, -x_2^{(2)} = \Omega_P(n\rho_n), \tag{B.28}$$

The same analysis and results hold for $u_2^{(2)}$ and $(y_1^{(2)}, y_2^{(2)})$. We now show after the second meta iteration, in addition to the condition (B.28), we further have

$$2y_1^{(2)} - s_1^{(2)} = \Omega_P(n\rho_n), \quad 2y_1^{(2)} - x_1^{(2)} = \Omega_P(n\rho_n) \tag{B.29}$$

To simplify notation, let

$$\alpha_i(s_1, x_1, y_1) \coloneqq \frac{2e^{8t(y_1 + r_i^{(y)})} + e^{4t(x_1 + r_i^{(x)})} + e^{4t(s_1 + r_i^{(s)})}}{1 + e^{4t(s_1 + r_i^{(s)})} + e^{4t(x_1 + r_i^{(x)})} + e^{8t(y_1 + r_i^{(y)})}}$$

where r's are the noise associated with each signal and we have Lemma A.4 bounding their order uniformly.

We first provide an upper bound on $\langle u_3^{(1)}, v_2 \rangle$. In (B.19),

$$\begin{aligned}
\langle u_3^{(1)}, v_2 \rangle &\le \frac{n}{4} + \sum_{i \in C_{12}} \frac{e^{4t(s_1^{(1)} + r_i^{(0)})}}{1 + e^{4t(s_1^{(1)} + r_i^{(0)})}} + \frac{n}{8} - \sum_{i \in C_{22}} \frac{e^{4t(s_2^{(1)} + r_i^{(0)})}}{1 + e^{4t(s_2^{(1)} + r_i^{(0)})}} + O_P(\sqrt{n}) \\
&\le \frac{3n}{8} + \frac{n}{8} \left(\Phi(\frac{-s_2^{(1)} + \Delta}{\sigma_u}) - \Phi(\frac{-s_1^{(1)} - \Delta}{\sigma_u}) \right) + C' n \frac{\rho_u}{\sigma_u^3} + C'' n e^{-4t\Delta} + O_P(\sqrt{n}) \\
&\le \frac{3n}{8} + o_P(n). \qquad \text{(B.30)}
\end{aligned}$$

by Lemma A.6.

For $u_1^{(2)}$, based on (B.19) and (B.21),

$$\begin{aligned}
\langle u_1^{(2)}, v_2 \rangle &= \sum_{i \in C_{11}} \alpha_i(s_1^{(2)}, x_1^{(1)}, y_1^{(1)}) + \frac{n}{4} - o_P(n), \\
\langle u_1^{(2)}, \mathbf{1}_n \rangle - m &= \sum_{i \in C_{11}} \alpha_i(s_1^{(2)}, x_1^{(1)}, y_1^{(1)}) - \frac{n}{4} - o_P(n).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\langle u_2^{(2)}, v_2 \rangle &= \sum_{i \in C_{11}} \alpha_i(s_1^{(2)}, x_1^{(2)}, y_1^{(1)}) + \frac{n}{4} - o_P(n), \\
\langle u_2^{(2)}, \mathbf{1}_n \rangle - m &= \sum_{i \in C_{11}} \alpha_i(s_1^{(2)}, x_1^{(2)}, y_1^{(1)}) - \frac{n}{4} - o_P(n).
\end{aligned}$$

For convenience denote $a = \frac{p+q}{2} - \lambda$ and $b = \frac{p-q}{2}$, then we have

$$\begin{aligned}
2y_1^{(2)} - s_1^{(2)} &= a(2\langle u_2^{(2)}, \mathbf{1}_n \rangle - \langle u_3^{(1)}, \mathbf{1}_n \rangle - m) + b(2\langle u_2^{(2)}, v_2 \rangle - \langle u_3^{(1)}, v_2 \rangle) \\
&\ge a \left(2 \sum_{i \in C_{11}} \alpha_i(s_1^{(2)}, x_1^{(2)}, y_1^{(1)}) - \frac{n}{4} - m \right) \\
&\quad + b \left(2 \sum_{i \in C_{11}} \alpha_i(s_1^{(2)}, x_1^{(2)}, y_1^{(1)}) + \frac{n}{2} - \frac{3n}{8} \right) - o_P(n\rho_n) \\
&= 2(a+b) \sum_{i \in C_{11}} \alpha_i(s_1^{(2)}, x_1^{(2)}, y_1^{(1)}) - \frac{3an}{8} + \frac{bn}{8} - o_P(n\rho_n)
\end{aligned}$$

by (B.30) and (B.22). Since $\alpha_i(s_1^{(2)}, x_1^{(2)}, y_1^{(1)}) \ge 1 + o_P(1)$, we can conclude

$$2y_1^{(2)} - s_1^{(2)} \ge \frac{3bn}{8} - \frac{an}{8} - o_P(n\rho_n) = \Omega(n\rho_n).$$

Similarly, we can check that

$$\begin{aligned}
2y_1^{(2)} - x_1^{(2)} &= a(2\langle u_2^{(2)}, \mathbf{1}_n \rangle - \langle u_1^{(2)}, \mathbf{1}_n \rangle - m) + b(2\langle u_2^{(2)}, v_2 \rangle - \langle u_1^{(2)}, v_2 \rangle) \\
&= (a+b) \sum_{i \in C_{11}} [2\alpha_i(s_1^{(2)}, x_1^{(2)}, y_1^{(1)}) - \alpha_i(s_1^{(2)}, x_1^{(1)}, y_1^{(1)})] - \frac{(a-b)n}{4} + o_P(n\rho_n) \\
&\ge \frac{(b-a)n}{4} - o_P(n\rho_n) = \Omega(n\rho_n) \qquad \text{(B.31)}
\end{aligned}$$

as $\alpha_i(s_1^{(2)}, x_1^{(2)}, y_1^{(1)}) > \alpha_i(s_1^{(2)}, x_1^{(1)}, y_1^{(1)})$. Thus condition (B.29) holds.

Now we need to analyze the third iteration in this meta iteration. Since $\alpha_i(s_1^{(2)}, x_1^{(2)}, y_1^{(1)}) \le 2$,

$$y_1^{(2)} + y_2^{(2)} = 2a(\langle u_2^{(2)}, \mathbf{1}_n\rangle - m) = o_P(n\rho_n),$$

then by (B.25)

$$s_1^{(2)} - (y_1^{(2)} + y_2^{(2)}) = \Omega_P(n\rho_n), \quad x_1^{(2)} - (y_1^{(2)} + y_2^{(2)}) = \Omega_P(n\rho_n). \tag{B.32}$$

Now using the update for $u_3^{(2)}$, and defining the noise in the same way as in the first meta iteration,

$$\begin{aligned}
\langle u_3^{(2)}, v_2\rangle = &\sum_{i\in C_{11}} \frac{2e^{8t(y_1^{(2)}+r_i^{(5)})} + e^{4t(x_1^{(2)}+r_i^{(4)})} + e^{4t(s_1^{(2)}+r_i^{(3)})}}{1 + e^{4t(s_1^{(2)}+r_i^{(3)})} + e^{4t(x_1^{(2)}+r_i^{(4)})} + e^{8t(y_1^{(2)}+r_i^{(5)})}} \\
&+ \sum_{i\in C_{12}} \frac{e^{4t(s_1^{(2)}+r_i^{(3)})} - e^{4t(x_2^{(2)}+r_i^{(4)})}}{1 + e^{4t(y_1^{(2)}+y_2^{(2)}+r_i^{(5)})} + e^{4t(s_1^{(2)}+r_i^{(3)})} + e^{4t(x_2^{(2)}+r_i^{(4)})}} \\
&+ \sum_{i\in C_{21}} \frac{e^{4t(x_1^{(2)}+r_i^{(4)})} - e^{4t(s_2^{(2)}+r_i^{(3)})}}{1 + e^{4t(y_1^{(2)}+y_2^{(2)}+r_i^{(5)})} + e^{4t(s_2^{(2)}+r_i^{(3)})} + e^{4t(x_1^{(2)}+r_i^{(4)})}} \\
&- \sum_{i\in C_{22}} \frac{2e^{8t(y_2^{(2)}+r_i^{(5)})} + e^{4t(s_2^{(2)}+r_i^{(3)})} + e^{4t(x_2^{(2)}+r_i^{(4)})}}{1 + e^{4t(s_2^{(2)}+r_i^{(3)})} + e^{4t(x_2^{(2)}+r_i^{(4)})} + e^{8t(y_2^{(2)}+r_i^{(5)})}} \\
\ge &\sum_{i\in C_{11}} \frac{2e^{8t(y_1^{(2)}+r_i^{(5)})}}{1 + e^{4t(s_1^{(2)}+r_i^{(3)})} + e^{4t(x_1^{(2)}+r_i^{(4)})} + e^{8t(y_1^{(2)}+r_i^{(5)})}} \\
&+ \sum_{i\in C_{12}} \frac{e^{4t(s_1^{(2)}+r_i^{(3)})}}{1 + e^{4t(y_1^{(2)}+y_2^{(2)}+r_i^{(5)})} + e^{4t(s_1^{(2)}+r_i^{(3)})} + e^{4t(x_2^{(2)}+r_i^{(4)})}} \\
&+ \sum_{i\in C_{21}} \frac{e^{4t(x_1^{(2)}+r_i^{(4)})}}{1 + e^{4t(y_1^{(2)}+y_2^{(2)}+r_i^{(5)})} + e^{4t(s_2^{(2)}+r_i^{(3)})} + e^{4t(x_1^{(2)}+r_i^{(4)})}} \\
&- n\exp(-\Omega_P(n\rho_n)) \\
\ge &\frac{n}{2} - n\exp(-\Omega_P(n\rho_n)),
\end{aligned}$$

using the conditions (B.28) (B.29) (B.32) and Lemma A.4. Since $\|u - z^*\|_1 = m - |\langle u, v_2\rangle|$, $||u_3^{(2)} - z^*||_1 = n\exp(-\Omega_P(n\rho_n))$ after the second meta iteration.

Finally we show the later iterations conserve strong consistency. Since

$$\begin{aligned}
\langle u_3^{(2)}, \mathbf{1}\rangle - m = &\sum_{i\in C_{11}} \frac{e^{8t(y_1^{(2)}+r_i^{(5)})} - 1}{1 + e^{4t(s_1^{(2)}+r_i^{(3)})} + e^{4t(x_1^{(2)}+r_i^{(4)})} + e^{8t(y_1^{(2)}+r_i^{(5)})}} \\
&+ \sum_{i\in C_{12}} \frac{e^{4t(y_1^{(2)}+y_2^{(2)}+r_i^{(5)})} - 1}{1 + e^{4t(y_1^{(2)}+y_2^{(2)}+r_i^{(5)})} + e^{4t(s_1^{(2)}+r_i^{(3)})} + e^{4t(x_2^{(2)}+r_i^{(4)})}} \\
&+ \sum_{i\in C_{21}} \frac{e^{4t(y_1^{(2)}+y_2^{(2)}+r_i^{(5)})} - 1}{1 + e^{4t(y_1^{(2)}+y_2^{(2)}+r_i^{(5)})} + e^{4t(s_2^{(2)}+r_i^{(3)})} + e^{4t(x_1^{(2)}+r_i^{(4)})}} \\
&+ \sum_{i\in C_{22}} \frac{e^{8t(y_2^{(2)}+r_i^{(5)})} - 1}{1 + e^{4t(s_2^{(2)}+r_i^{(3)})} + e^{4t(x_2^{(2)}+r_i^{(4)})} + e^{8t(y_2^{(2)}+r_i^{(5)})}} \\
= &n\exp(-\Omega_P(n\rho_n))
\end{aligned}$$

by (B.28) (B.29) (B.32) and Lemma A.4, we have

$$s_1^{(3)} = a(\langle u_3^{(2)}, \mathbf{1}\rangle - m) + b\langle u_3^{(2)}, v_2\rangle = \frac{p-q}{4}n + n\rho_n\exp(-\Omega_P(n\rho_n)),$$

$$s_2^{(3)} = a(\langle u_3^{(2)}, \mathbf{1}\rangle - m) - b\langle u_3^{(2)}, v_2\rangle = -\frac{p-q}{4}n + n\rho_n \exp(-\Omega_P(n\rho_n)).$$

Next we note the noise in this iteration now arises from the whole graph A, and can be bounded by

$$\begin{aligned} r_i^{(7)} &= [R^{zz}, R^{zy}]_{i,\cdot}(u_3^{(2)} - \frac{1}{2}\mathbf{1}_n) \\ &= [R^{zz}, R^{zy}]_{i,\cdot}(u_3^{(2)} - z^*) + [R^{zz}, R^{zy}]_{i,\cdot}(z^* - \frac{1}{2}\mathbf{1}_n), \end{aligned}$$

where the second term is $O_P(\sqrt{n\rho_n \log n})$ uniformly for all i, applying Lemma A.4. To bound the first term, note that

$$\begin{aligned} \max_i |[R^{zz}, R^{zy}]_{i,\cdot}(u_3^{(2)} - z^*)| &\le \|[R^{zz}, R^{zy}](u_3^{(2)} - z^*)\|_2 \\ &\le O_P(\sqrt{n\rho_n})\|u_3^{(2)} - z^*\|_1 = o_P(1). \end{aligned}$$

Therefore $r_i^{(7)}$ is uniformly $O_P(\sqrt{n\rho_n \log n})$ for all i. By a similar calculation to (B.31), we can check that condition (B.29) holds for $y_1^{(2)}$ and $s_1^{(3)}$, since when $s_1, x_1, y_1 = \Omega(n\rho_n)$ condition (B.29) and $1 - o_P(1) \le \alpha_i(s_1, x_1, y_1) \le 2 + o_P(1)$ guarantees each other and condition (B.29) is true in the previous iteration. We can check that condition (B.32) also holds. The rest of the argument can be applied to show $\|u_1^{(3)} - z^*\|_1 = n\exp(-\Omega_P(n\rho_n))$. At this point, all the arguments can be repeated for later iterations.

□

Proof of Corollary 1. We first consider $\mu > 0.5$. By (B.6), $s_1 = \Omega_P(n\rho_n)$, $s_2 = \Omega_P(n\rho_n)$. Since $r_i^{(0)} = O_P(\sqrt{n\rho_n \log n})$ uniformly for all i by Lemma A.4, we have

$$\phi_i^{(1)} = \frac{1 + e^{4t(s_1 + r_i^{(0)})}}{3 + e^{4t(s_1 + r_i^{(0)})}} = 1 - \exp(-\Omega_P(n\rho_n))$$

for $i \in C_1$. Similarly for $i \in C_2$, and $\xi_i^{(1)} = \exp(-\Omega_P(n\rho_n))$. Define $u'_i = \mathbf{1}_{[i \in P_1]} + \mathbf{1}_{[i \in P_2]}$. Since the partition into P_1 and P_2 is random, $u'_i \sim$ iid Bernoulli(1/2), and $\|u_1 - u'\|_2 = \sqrt{n}\exp(-\Omega_P(n\rho_n))$.

In the second iteration, we can write

$$\begin{aligned} \theta^{01} &= 4t([A^{yz}, A^{yy}] - \lambda J)(u_1 - \tfrac{1}{2}\mathbf{1}) \\ &= 4t([A^{yz}, A^{yy}] - \lambda J)(u_1 - u') + 4t([A^{yz}, A^{yy}] - \lambda J)(u' - \tfrac{1}{2}\mathbf{1}) \\ &= O_P(n\sqrt{\rho}\exp(-\Omega_P(n\rho_n))) + 4t([A^{yz}, A^{yy}] - \lambda J)(u' - \tfrac{1}{2}\mathbf{1}). \end{aligned}$$

The signal part of the second term is $4t(x_1\mathbf{1}_{C'_1} + x_2\mathbf{1}_{C'_2})$ with x_1 and x_2 having the form of (B.6), with $u^{(0)}$ replaced by u'. Since $x_1, x_2 = \Omega_P(\sqrt{n}\rho_n)$, the rest of the analysis proceeds like that of Theorem 1 restarting from the first iteration.

If $\mu < 0.5$, $s_1 = -\Omega_P(n\rho_n)$, $s_2 = -\Omega_P(n\rho_n)$. We have $\phi_i^{(1)} = \frac{1}{3} + \exp(-\Omega_P(n\rho_n))$, $\xi_i^{(1)} = \frac{2}{3} - \exp(-\Omega_P(n\rho_n))$. This time let $u' = \frac{1}{3}\mathbf{1}_{[i \in P_1]} + \frac{2}{3}\mathbf{1}_{[i \in P_2]}$, then θ^{01} can be written as

$$\theta^{01} = O_P(n\sqrt{\rho}\exp(-\Omega_P(n\rho_n))) + \frac{4t}{3}([A^{yz}, A^{yy}] - \lambda J)(3u' - \frac{3}{2}\mathbf{1}).$$

Noting that $3u'_i - 1 \sim$ iid Bernoulli(1/2), the same argument applies. □

Proof of Proposition 1. (i) We show each point is a stationary point by checking the vector update form of (10), (9), (11). Similar to Theorem 1, we have

$$\theta^{10} = 4t(s_1\mathbf{1}_{C_1} + s_2\mathbf{1}_{C_2} + r_i^{(0)})$$

where $r_i^{(0)} = O_P(\sqrt{n\rho_n \log n})$. Plugging $u^{(0)} = \mathbf{1}_n$ in (9), $s_1 = s_2 = 0.5(\frac{p+q}{2} - \lambda)n$. Similarly

$$\theta^{01} = 4t(x_1 \mathbf{1}_{C_1} + x_2 \mathbf{1}_{C_2} + r_i^{(1)})$$
$$\theta^{11} = 4t(y_1 \mathbf{1}_{C_1} + y_2 \mathbf{1}_{C_2} + r_i^{(1)})$$

where $x_1 = x_2 = 0.5(\frac{p+q}{2} - \lambda)n$, $y_1 = y_2 = (\frac{p+q}{2} - \lambda)n$. Plugging in (12) with $\frac{p+q}{2} - \lambda = \Omega_P(\rho_n)$ by Lemma A.1, we have

$$\phi_i^{(1)} = 1 - \exp(-\Omega_P(n\rho_n)), \quad \xi_i^{(1)} = 1 - \exp(-\Omega_P(n\rho_n))$$

for all $i \in [m]$. Hence for sufficiently large n, $u^{(0)} = \mathbf{1}_n$ is the stationary point. For $u^{(0)} = \mathbf{0}_n$, similarly we have

$$\phi_i^{(1)} = \exp(-\Omega_P(n\rho_n)), \quad \xi_i^{(1)} = \exp(-\Omega_P(n\rho_n))$$

so $u^{(0)} = \mathbf{0}_n$ is also a stationary point for large n.

(ii) The statement for $u^{(0)} = \mathbf{0}_n$ and $u^{(0)} = \mathbf{1}_n$ follows from Corollary 1 by $\mu = 0$ and $\mu = 1$.

□

Proof of Proposition 2. Let $\hat{t}, \hat{\lambda}$ be constants defined in terms of $\hat{p}, \hat{q}$. First we observe using $\hat{p}, \hat{q}$ only replaces t, λ with $\hat{t}, \hat{\lambda}$ everywhere in the updates of Algorithm 1. We can check the analysis in Theorem 1 remains unchanged as long as

$$\text{i) } \frac{p+q}{2} > \hat{\lambda}, \quad \text{ii) } \hat{\lambda} - q = \Omega(\rho_n), \quad \text{iii) } \hat{t} = \Omega(1)$$

□

Proof of Theorem 2. Starting with $p^{(0)}$ and $q^{(0)}$ satisfying the conditions in Corollary 2, after two meta iterations of u updates, we have $\|u_3^{(2)} - z^*\|_1 = n\exp(-\Omega(n\rho_n))$. Updating $p^{(1)}, q^{(1)}$ with (14), we first analyze the population version of the numerator of $p^{(1)}$,

$$\begin{aligned}
&(\mathbf{1}_n - u)^T P(\mathbf{1}_n - u) + u^T P u + 2(\mathbf{1}_m - \psi^{10} - \psi^{01})^T \mathrm{diag}(P^{zy})\mathbf{1}_m \\
=&(\mathbf{1}_n - z^*)^T P(\mathbf{1}_n - z^*) + (z^*)^T P z^* - 2(u - z^*)^T P(\mathbf{1}_n - z^*) + 2(z^*)^T P(u - z^*) \\
&+ (u - z^*)^T P(u - z^*) + O(n\rho_n).
\end{aligned}$$

In the case of $u_3^{(2)}$, the above becomes

$$\frac{n^2}{2}p + O_P(n^{5/2}\rho_n \exp(-\Omega(n\rho_n))) + O(n\rho_n) = \frac{n^2}{2}p + O_P(n\rho_n).$$

Next we can rewrite the noise as

$$\begin{aligned}
&(\mathbf{1}_n - u)^T (A - P)(\mathbf{1}_n - u) + u^T (A - P) u \\
=&(\mathbf{1}_n - z^*)^T (A - P)(\mathbf{1}_n - z^*) + (z^*)^T (A - P) z^* - 2(u - z^*)^T (A - P)(\mathbf{1}_n - z^*) \\
&+ 2(z^*)^T (A - P)(u - z^*) + (u - z^*)^T (A - P)(u - z^*).
\end{aligned}$$

Similarly in the case of $u_3^{(2)}$, the above is $O_P(\sqrt{n^2\rho_n})$. Therefore the numerator of $p^{(1)}$ is $\frac{n^2}{2}p + O_P(\sqrt{n^2\rho_n})$. To lower bound the denominator, note that

$$\begin{aligned}
&u^T (J - I) u + (\mathbf{1} - u)^T (J - I)(\mathbf{1} - u) \\
=&\left(\sum_i u_i\right)^2 + \left(n - \sum_i u_i\right)^2 - u^T u - (\mathbf{1} - u)^T (\mathbf{1} - u) \\
\geq& n^2/2 - 2n,
\end{aligned}$$

then we have $p^{(1)} = p + O_P(\sqrt{\rho_n}/n)$. The same analysis shows $q^{(1)} = q + O_P(\sqrt{\rho_n}/n)$.

Replacing p and q with $p^{(1)}$ and $q^{(1)}$ in the final analysis after the second meta iteration of Theorem 1 does not change the order of the convergence, and the rest of the arguments can be repeated. □

C Generalizations

We present the **update equations for balanced** $K > 2$ models. We will use the notation $a, b \in \{0, \ldots, K-1\}$ to be consistent with the two class case. Let $S_{zy} = 2t(\mathrm{diag}(A^{zy}) - \lambda I)\mathbf{1}_m$.

$$\theta^{ab} = \begin{cases} 2t[A^{zz} - \lambda(J-I)](\phi_a - \phi_0) + 2t[A^{zy} - \lambda(J-I) - \mathrm{diag}(A^{zy})](\xi_a - \xi_0) - S_{zy}, & a \neq 0, b = 0 \\ 2t[A^{zz} - \lambda(J-I)](\phi_b - \phi_0) + 2t[A^{zy} - \lambda(J-I) - \mathrm{diag}(A^{zy})](\xi_b - \xi_0) - S_{zy}, & a = 0, b \neq 0 \\ \theta_{a0} + \theta_{b0} + S_{zy} & a \neq 0, b \neq 0 \end{cases} \tag{C.1}$$

The **update equations for unbalanced two class blockmodels** simply adds an additional term of $\log \pi/(1-\pi)$ to the updates of θ_{10} (Eq. (9)), θ_{01} (Eq. (10)) and $2\log \pi/(1-\pi)$ to θ_{11} (Eq. (11)). We assume that the proportions are known.

In Figure A.1, we show the heatmap for mis-specified parameters for VIPS on unbalanced SBM ($\pi = .3$) and balanced SBM with $K = 3$. For each starting point of $\hat{p}, \hat{q}$ the average NMI is shown. We see that in both cases the VIPS algorithm converges to the correct labels for a wide range of initial parameter settings.

(a) Unbalanced with $\pi = 0.3$ (b) Balanced with $K = 3$

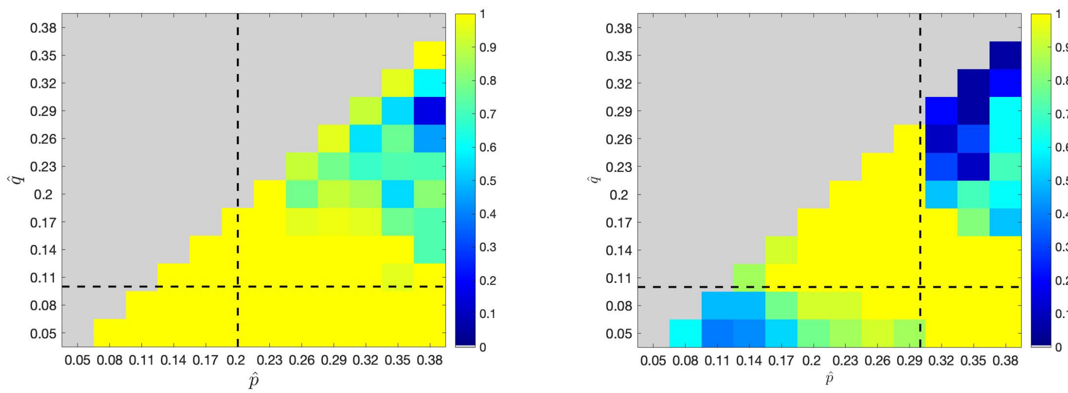

Figure A.1: NMI with different estimation of $\hat{p}$, $\hat{q}$ with $\hat{p} > \hat{q}$, averaged over 20 random initializations for each $\hat{p}$, $\hat{q}$. The left figure has $\pi = 0.3, K = 2$ and the right figure has balanced clusters with $K = 3$. The true $(p_0, q_0) = (0.2, 0.1)$ and $n = 2000$.

For $K = 3$, we also show Figure A.2, where each row represents the estimated membership of one random trial and both MFVI and VIPS are run with the true p_0, q_0. We show VIPS can recover true membership with higher probability than MFVI.

(a) MFVI (b) VIPS

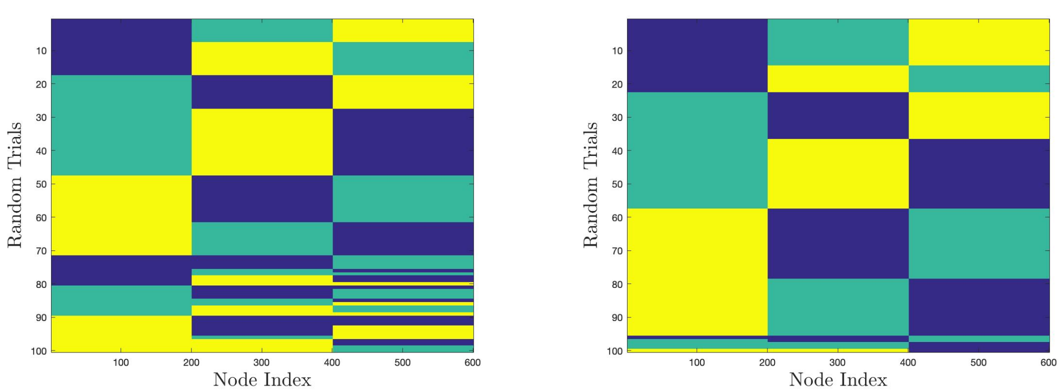

Figure A.2: Compare VIPS and MFVI when K=3, equal sized communities, for known p_0, q_0 in 100 random trials. $p_0 = 0.5, q_0 = 0.01$. Rows permuted for visual clarity.

D Additional Experimental Results

In Figure A.3, we compare different update rules in VIPS with (i) parameters p, q fixed at the true values (True), (ii) (p, q) estimated using $(\sum_{i \neq j} A_{ij}/(n(n-1)), \sum_{i \neq j} A_{ij}/(2n(n-1)))$ but fixed (Estimate), and (iii) (p, q) initialized as in (ii) and updated in the algorithm (Update) using Eq. (14). In all settings, VIPS successfully converges to the ground truth, which is consistent with our theoretical results and shows robustness of the parameter setting.

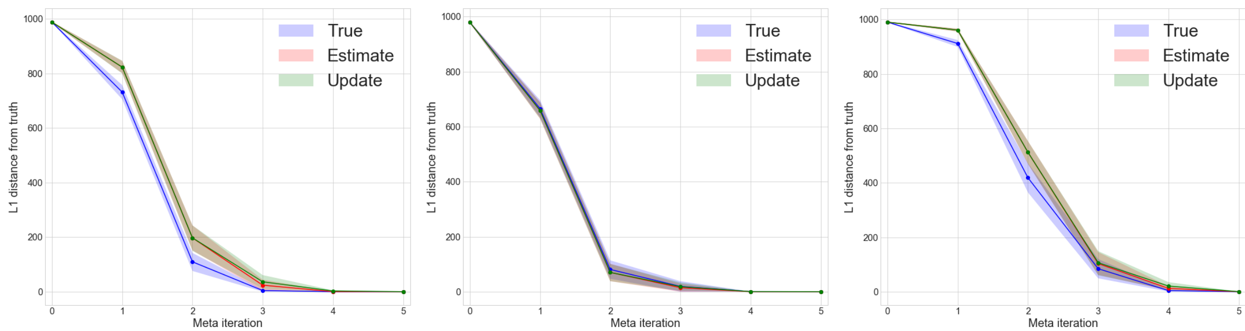

Figure A.3: Values of $\|u - z^*\|_1$ as the number of meta iterations increases. Each line is the mean curve of 50 random trials and the shaded area is the standard deviation. Here $n = 2000$ and $p_0 = 0.1, q_0 = 0.02$. u is initialized by Bernoulli distribution with mean $\mu = 0.1, 0.5, 0.9$ from the left to right.

In Figure A.4, we compare VIPS and MFVI with and without parameter updates. For VIPS, we do parameter updates from 3rd meta iteration onward, and for fairness, we start parameter updates 9 iterations onward for MFVI. In both schemes, the VIPS performs better than MFVI.

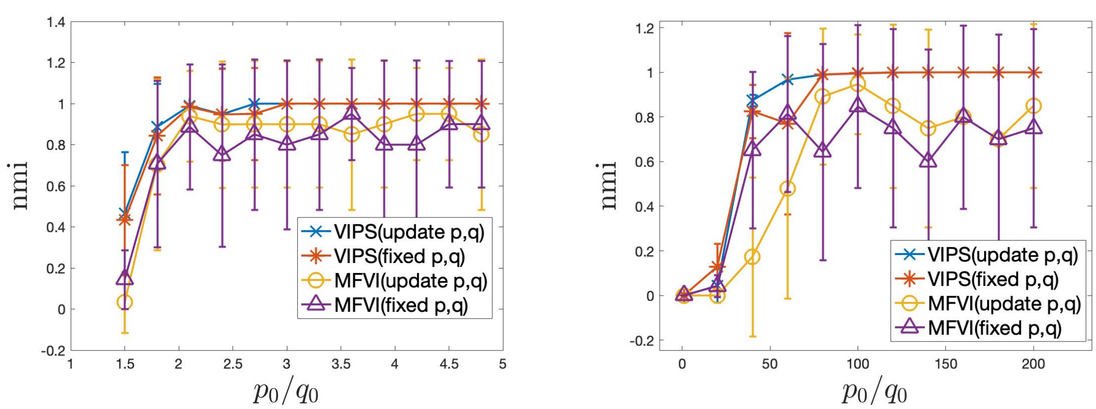

Figure A.4: Two schemes for estimating model parameters for VIPS and MFVI. Both use the initial $\hat{p}$ and $\hat{q}$ as described in Figure 4 in the main paper. The first scheme starts updating $\hat{p}$ and $\hat{q}$ after 3 meta iterations for VIPS and 9 iterations for MFVI. The other scheme has $\hat{p}, \hat{q}$ held fixed.

References

Paul Erdös. On a lemma of littlewood and offord. *Bulletin of the American Mathematical Society*, 51(12):898–902, 1945.

Purnamrita Sarkar, YX Wang, and Soumendu Sunder Mukherjee. When random initializations help: a study of variational inference for community detection. *arXiv preprint arXiv:1905.06661*, 2019.