COMMUNITY EXTRACTION OF NETWORK DATA UNDER STOCHASTIC BLOCK MODELS

Quan Yuan[1], Binghui Liu[1,*], Danning Li[1,*], Yanyuan Ma[2]

[1]*Northeast Normal University and* [2]*Pennsylvania State University*

In Section S1, we propose an accelerated refinement algorithm RACE and indicate that the performance of RACE and RACEn is very similar via some simulation results. In Section S2, we demonstrate in detail how we select the tuning parameters τ and K. In Section S3, we explain and compare the assumptions of the main theorems and corollary imposed. In Section S4, we compare the performance of RACE initialized with the two initialization algorithms INIT and ESC, respectively, with some of their competitors in situation of $K = 3$. In Section S5, we make some additional discussions. Then, in Section S6, we present the proofs of Theorems 1-5, Proposition 1 and Corollary S1. For simplicity, we will abbreviate $I_{t^*}(p, q)$, $\mathbb{P}_{\boldsymbol{P},\boldsymbol{c}}$ and $\mathbb{E}_{\boldsymbol{P},\boldsymbol{c}}$ as I_{t^*}, $\mathbb{P}$ and $\mathbb{E}$, respectively. Let $\delta = p/q - 1$. The constant C and the sequence $\eta = \eta_n$ that tends to 0 may vary case by case.

S1. Supplementary algorithm

Algorithm S1. (RACE)

Input: *The adjacency matrix* $\boldsymbol{A} \in \{0, 1\}^{n\times n}$*, an initialization algorithm and the specific value of* $K \geq 2$*.*

*Corresponding authors.

Output: *An estimator* $\check{\boldsymbol{c}} \in [K]^n$ *of the community label vector* $\boldsymbol{c} \in [K]^n$.

1. *(Initialization) Applying the initialization algorithm to* $\boldsymbol{A}$*, we get the output* $\boldsymbol{c}^0$.

2. *(Refinement) For each* $k, l \in [K]$,

$$P_{kl}^0 = \begin{cases} \dfrac{\sum\limits_{u<v} A_{uv}\mathbb{I}\{\boldsymbol{c}^0(u)=k, \boldsymbol{c}^0(v)=k\}}{\frac{1}{2}n_k^0(n_k^0-1)}, & k = l; \\ \dfrac{\sum\limits_{u,v\in[n]} A_{uv}\mathbb{I}\{\boldsymbol{c}^0(u)=k, \boldsymbol{c}^0(v)=l\}}{n_k^0 n_l^0}, & k \neq l, \end{cases}$$

with $n_k^0 = \sum_{j\in[n]} \mathbb{I}\{\boldsymbol{c}^0(j) = k\}$. *Let* $\hat{q}^0 = \frac{\sum_{k=1}^{K-1} n_k^0 P_{Kk}^0 + (n_K^0-1)P_{KK}^0/2}{\sum_{k=1}^{K-1} n_k^0 + (n_K^0-1)/2}$*, and update* $P_{Kl}^0 = \hat{q}^0$ *for all* $l \in [K]$.

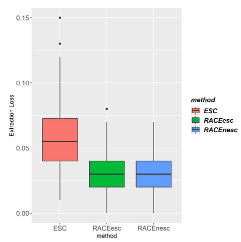
For each $i \in [n]$*, let*

$$\check{\boldsymbol{c}}(i) = \underset{k\in[K]}{\operatorname{argmax}} \sum_{l=1}^{K-1} \sum_{j:\boldsymbol{c}^0(j)=l} \Big[A_{ij}\log P_{kl}^0 + (1-A_{ij})\log(1-P_{kl}^0)\Big].$$

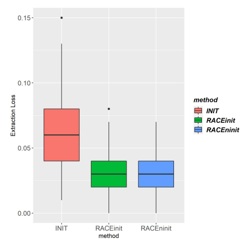
In Algorithm S1, it first assigns node i to the most likely group based on the initialization algorithm $\boldsymbol{c}^0$, for each $i \in [n]$. Note that we only used observations $(A_{ij})_{i\in[n], j:\boldsymbol{c}^0(j)\neq K}$, because from the model (2.1), we can know that the connection probability between each group and the background node is the same, without distinction.

S.1.1 Algorithms 1 and S1 have similar performance

As we mentioned earlier, the community extraction performance of Algorithm 1 and Algorithm S1 is almost the same. Here is an example in the



S.1(a) ESC



S.1(b) INIT

Figure S.1: The almost negligible performance difference between RACE and RACEn.

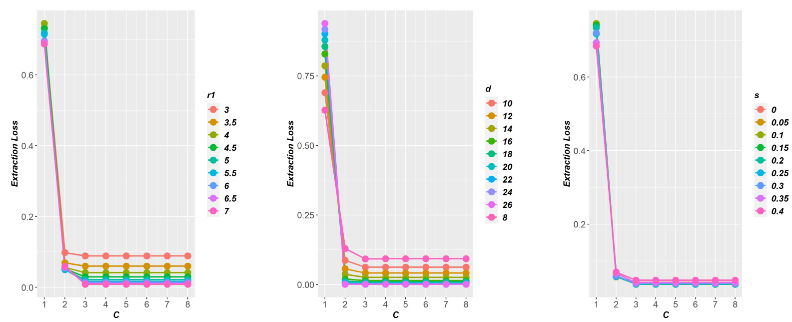
simulation in Figure S.1, with $n = 100$, $p = 4$, $q = 1$, $\lambda = 12$, where RACEn is Algorithm 1 and RACE is Algorithm S1.

S2. Selection of tuning parameters

S.2.1 Selection of τ

Combined with the selection of τ in Gao et al. (2017), i.e, $\tau = 2\bar{d}$, where $\bar{d} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij}/n$ is the average degree of the network $\boldsymbol{A}$, we set $\tau = C\bar{d}$. Suggested by extensive simulation results, we see that $C = 2$, i.e. $\tau = 2\bar{d}$, is also a good choice for our study.

The specific simulation settings are the same as the previous settings (I)-(III) for $K = 2$. The simulation results are summarized in Figures S.2 and S.3. The horizontal axis of Figures S.2 and S.3 are $C \in [8]$. We can see that $\tau = 2\bar{d}$ is indeed a watershed in terms of extraction loss, as shown in Figure S.2 and S.3.



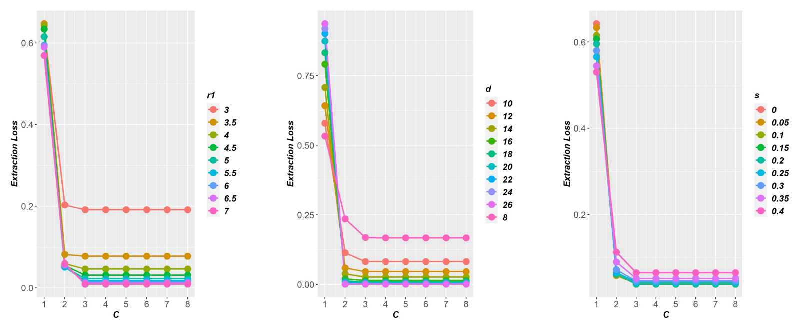
S.2(a) Setting (I) S.2(b) Setting (II) S.2(c) Setting (III)

Figure S.2: The extraction loss with varying C in the case of $K = 2$ for Settings (I)-(III), where the initialization algorithm is ESC.

S.2.2 Selection of K

In addition, the parameter K in the algorithm input is not known in all cases. When K is unknown, we choose an innovative method, "corrected Bayesian information criterion" (CBIC) that is proposed by Hu et al. (2020) to select K. Specifically,

$$L(\tilde{K}; \hat{\boldsymbol{c}}, \boldsymbol{A}) = L_0(\boldsymbol{A}, \tilde{K}, \hat{\boldsymbol{c}}) - \left\{ \kappa n \log \tilde{K} + \frac{1}{2} \tilde{K}(\tilde{K}+1) \log n \right\},$$



S.3(a) Setting (I)   S.3(b) Setting (II)   S.3(c) Setting (III)

Figure S.3: The extraction loss with varying C in the case of $K = 2$ for Settings (I)-(III), where the initialization algorithm is INIT.

where $L_0(\boldsymbol{A}, \tilde{K}, \hat{\boldsymbol{c}}) = \sum_{i=1}^{n} \sum_{j>i} \{A_{ij} \log \hat{P}_{\hat{\mathbf{c}}(i)\hat{\mathbf{c}}(j)} + (1 - A_{ij}) \log(1 - \hat{P}_{\hat{\mathbf{c}}(i)\hat{\mathbf{c}}(j)})\}$ and for each $k, l \in [K]$,

$$\hat{P}_{kl} = \begin{cases} \dfrac{\sum\limits_{i<j} A_{ij}\mathbb{I}\{\hat{\mathbf{c}}(i)=k, \hat{\mathbf{c}}(j)=k\}}{\frac{1}{2}\hat{n}_k(\hat{n}_k - 1)}, & k = l, \\ \dfrac{\sum\limits_{i,j\in[n]} A_{ij}\mathbb{I}\{\hat{\mathbf{c}}(i)=k, \hat{\mathbf{c}}(j)=l\}}{\hat{n}_k \hat{n}_l}, & k \neq l, \end{cases}$$

with $\hat{n}_k = \sum_{j\in[n]} \mathbb{I}\{\hat{\boldsymbol{c}}(j) = k\}$. Then, choose

$$\widehat{K} = \text{argmax}_{\tilde{K}\in\mathcal{K}} L(\tilde{K}; \hat{\boldsymbol{c}}, \boldsymbol{A}),$$

where $\mathcal{K}$ is a candidate set for K. Here, we use RACEinit or RACEesc as the community extraction method for getting $\hat{\boldsymbol{c}}$. By experience, here we set $\kappa = 4.5$.

To see the accuracy of CBIC's selection of K, we consider the following settings with $n = 300$, $d = 32$, $\boldsymbol{\omega} = (1 - s, 1 + s)^\top/2$:

(I) $q_0' = q_0 = 1$, $p_0 = r_1 q_0$, r_1 varies from 5 to 7 and $s = 0$;

(II) d varies from 30 to 38, $p_0 = 5$, $q_0 = q_0' = 1$ and $s = 0$;

(III) s varies from 0 to 0.24, $p_0 = 5$ and $q_0 = q_0' = 1$.

Table S.1: The accuracy of using "corrected Bayesian information criterion" (CBIC) to select K based on RACEesc and RACEinit

		True $K = 2$		True $K = 3$	
Setting	r_1	RACEesc	RACEinit	RACEesc	RACEinit
(I)	5.0	1.00	1.00	1.00	1.00
	5.5	1.00	1.00	1.00	1.00
	6.0	1.00	1.00	1.00	1.00
	6.5	1.00	1.00	1.00	1.00
	7.0	1.00	1.00	1.00	1.00
	d	RACEesc	RACEinit	RACEesc	RACEinit
(II)	30	1.00	1.00	0.89	0.89
	32	1.00	1.00	1.00	1.00
	34	1.00	1.00	1.00	1.00
	36	1.00	1.00	1.00	1.00
	38	1.00	1.00	1.00	1.00
	s	RACEesc	RACEinit	RACEesc	RACEinit
(III)	0.00	1.00	1.00	1.00	1.00
	0.06	1.00	1.00	1.00	1.00
	0.12	1.00	1.00	1.00	1.00
	0.18	1.00	1.00	0.96	0.96
	0.24	1.00	1.00	0.91	0.91

We have repeated all experiments 100 times, and Table S.1 shows the accuracy of selecting the true K from the 100 times. From Table S.1, we can see that the performance of selecting K by using CBIC is good under the above settings.

S3. Comparison of conditions imposed

S.3.1 Comparison of conditions imposed in the main theorems

To better understand the conditions imposed in the main theorems and corollary of this paper, we consider the following setting. Set $p = n^{-a}$ and

$\beta = C_1 n^{-c}$, where $0 \leq a < 1$, $0 \leq c < 1$ and $C_1 \in (0,1)$ is a constant. Besides, when $p \asymp q$, set $p - q = C_2 n^{-b}$ with constants $C_2 \in (0,1)$ and $b \geq a$; when $p \gg q$, set $p/q = C_2 n^b$ with constants $C_2 \in (0,1)$ and $b \geq 0$.

Then, under the above setting, we present the sufficient conditions for Assumption 1 and the conditions imposed in Corollary 1, Theorems 1-4, respectively, in Table S.2, which allow for extremely unbalanced node numbers of community and background group. For example, when $p \asymp q$, i.e. $a = b$, if we set $p = n^{-1/4}$, i.e. $a = 1/4$, then the order of β can be $n^{-(1/8-\epsilon)}$ for any small constant $\epsilon > 0$.

Moreover, it can be seen that in our study the order of the average degree can be much smaller than $\log n$, whereas many existing studies require that the order of the average degree is greater than $\log n$, such as in Zhao et al. (2011). For example, when $p = \log n/n$, $q = 1/(n \log n)$ and $\beta = 1/\log^{1/4} n$, the conditions imposed in Theorem 1 and Theorem 4 still hold, where the order of the average degree is smaller than $2\log^{3/4} n \ll \log n$.

S.3.2 Explanation of the additional conditions on p and q

First, we explain the rationale of the additional assumptions regarding p and q beyond Assumption 1 in Theorems 1-4 and Corollary 1.

Table S.2: Sufficient conditions for Assumption 1 and the conditions imposed in Corollary 1, Theorems 1-4, respectively.

	$p \asymp q$	$p \gg q$
Assumption 1	$a+2c<1$	$a+2c<1$
Corollary 1	$3b+2c-2a<1$	$a+2c<1, 3b\geq a$
Theorem 1	$3b+2c-2a<1$	$a+2c<1$
Theorem 2	$a+2c<1$	$a+2c<1, 3b\geq a$
Theorem 3	$4b+2c-a<1$	$b+3a+2c<1$
Theorem 4	$2b+2c-a<1$	$a+2c<1$

(1) In Theorem 1, when $p \to q$, the additional condition $-\frac{p-q}{p}\frac{\beta^2 n I_{t^*}(p,q)}{\log \beta} \to \infty$ is used to limit the speed at which p tends to q and ensure that the background nodes can be distinguished. As presented in Table S.2, the condition $3b-2a+2c<1$ is a sufficient condition for this condition.

(2) In Theorem 2, when $p \gg q$, the additional conditions are

$$\lim_{n\to\infty}\frac{\log\frac{\log(\frac{p}{q})}{p}}{\log n}<1,\ p\log^3\left(\frac{p}{q}\right)<\infty \text{ and } \lim_{n\to\infty}\frac{\log \beta np}{\log\log\frac{p}{q}}>3.$$

They are used to limit the upper bounds of p/q, $1/p$ and $1/\beta$. Specifically, they require that p/q is not too large, and p and β does not approach 0 too quickly. These additional conditions are used to establish appropriate concentration of the likelihood ratio statistic, specifically in establishing the lower bound of the minimax risk. In Table S.2, $3b>a$ is a sufficient condition for $p\log^3\left(\frac{p}{q}\right)<\infty$, $a<1$ is a sufficient condition for $\lim_{n\to\infty}\frac{\log\frac{\log(\frac{p}{q})}{p}}{\log n}<1$ and $a+c<1$ is a sufficient condition for $\lim_{n\to\infty} \frac{\log \beta n p}{\log\log \frac{p}{q}} > 3$. Therefore, under this setting, these three additional conditions must hold if $a + 2c < 1$ and $3b > a$.

(3) Theorem 3 requires that $\beta n \frac{(p-q)^4}{p} \to \infty$ and the initialization algorithm in Algorithm 1 satisfies Condition 1 with

$$\gamma_n = o\left(-\frac{\beta}{\log\beta}(p-q)\right) \tag{S3.1}$$

when $p \asymp q$, and

$$\gamma_n = o\left(-\frac{\beta}{\log\beta} q \frac{\log\log\frac{p}{q}}{\log\frac{p}{q}}\right) \tag{S3.2}$$

when $p \gg q$. We need these additional conditions because p and q need to be estimated, and obtaining sufficiently good estimates for p and q requires stronger conditions on them. In Table S.2, when $p \asymp q$, $4b + c - a < 1$ is a sufficient condition for $\beta n \frac{(p-q)^4}{p} \to \infty$, and $3b + 2c - a < 1$ is a sufficient condition for the initialization algorithm in Algorithm 1 to satisfy Condition 1 with $\gamma_n = o\left(-\frac{\beta}{\log\beta}(p-q)\right)$. On the other hand, when $p \gg q$, $3a + c < 1$ is a sufficient condition for $\beta n \frac{(p-q)^4}{p} \to \infty$, and $b + 2a + 2c < 1$ is a sufficient condition for the initialization algorithm in Algorithm 1 to satisfy Condition 1 with $\gamma_n = o\left(-\frac{\beta}{\log\beta} q \frac{\log\log\frac{p}{q}}{\log\frac{p}{q}}\right)$.

(4) In Theorem 4, the additional condition is $\frac{-\beta}{\log\beta}\frac{\beta n(p-q)^2}{p} \to \infty$, which is used to establish the consistency of the estimate of $\boldsymbol{c}$. In Table S.2,

when $p \asymp q$, $2b + 2c - a < 1$ is sufficient to ensure this condition to hold. When $p \gg q$, $a + 2c < 1$ is sufficient.

Next, we discuss whether these additional conditions on p and q are satisfied in some typical settings of existing studies, such as Wilson et al. (2017), Yun and Proutiere (2016) and Zhao et al. (2011).

(1) First, we consider the requirements imposed on p and q in Wilson et al. (2017), which are equivalent to that β, p and q with $p > q$ are constants. These requirements are stronger because they automatically lead to the conditions on p and q in Theorems 1-4.

(2) Then, we consider the requirements imposed on p and q in Yun and Proutiere (2016), which are equivalently written as $\beta \gtrsim 1$, $np \to \infty$, $1 < \lim_{n\to\infty} p/q < \infty$ and $nI_{t^*}(p,q) \to \infty$. These requirements are stronger because they automatically lead to the conditions of Theorems 1, 2 and 4. On the other hand, the conditions for p and q in Theorem 3 intersect with the above requirements, but neither implies the other.

(3) Finally, we consider the requirements on p and q in Theorem 2 of Zhao et al. (2011): $\beta \gtrsim 1$, $\lim_{n\to\infty} p/q < \infty$ and $np/\log n \to \infty$. Because their theorem studies the consistency of the estimated community labels, we only compare their requirements with the conditions imposed

in Theorem 4 which also studies the consistency. We find that these requirements are stronger because they automatically lead to the conditions imposed on p and q in Theorem 4.

S.3.3 ESC is a suitable initialization algorithm for RACEn

We use the following corollary to better understand that ESC is a suitable initialization algorithm for RACEn.

Corollary S1. *Let* $(\boldsymbol{P}, \boldsymbol{c}) \in \Theta_n(p, q, \beta)$ *and assume that* $\lim_{n\to\infty} \frac{p\beta}{q(1-\beta)} > 2$*, and* $-\frac{\beta}{\log\beta}\frac{nq^4}{p^2} \to \infty$ *as* $n \to \infty$*. Then, the output of Algorithm 3, i.e.* $\boldsymbol{c}^0_{\mathsf{esc}}$*, satisfies*

$$\inf_{(\boldsymbol{P},\boldsymbol{c})\in\Theta_n(p,q,\beta)} \mathbb{P}_{P,\boldsymbol{c}}\left(\ell\left(\boldsymbol{c}, \boldsymbol{c}^0_{\mathsf{esc}}\right) \le C(1+\xi)\frac{p^2}{nq^4}\right) \ge 1 - n^{-(1+C')},$$

for some constants $C, C' > 0$*, where* ξ *comes from the* $(1+\xi)$*-approximation K-means optimization in step* 3 *of Algorithm 3.*

Corollary S1 indicates that under certain conditions, the combination of ESC and RACEn can output a community extraction result that reaches the asymptotic minimax risk. Under the same setting considered in Table S.2, when $p \asymp q$, due to the assumption $\lim_{n\to\infty} \frac{p\beta}{q(1-\beta)} > 2$, we have $p - q = C_1 n^{-b}$ with $b = a$, and $c = 0$, and the sufficient condition for the conditions imposed in Corollary S1 is $a < 1/2$. Besides, when $p \gg q$, we obtain that the sufficient condition for the conditions imposed in Corollary S1 is $4b + c - 2a < 1$.

S4. Simulation results in the case of $K = 3$

In this section, we investigate the performance of RACE in the case of $K = 3$. We consider the following five settings:

(I) $q_0' = q_0 = 1$, $p_0 = r_1\, q_0$, r_1 varies from 3 to 7 and $s = 0$;

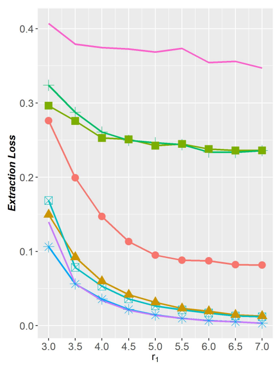
(II) $p_0 = 4$, $q_0 = 1$, $q_0' = r_2\, q_0$, r_2 varies from 0 to 1 and $s = 0$;

(III) $p_0 = 4$, $q_0' = 1$, $q_0 = r_3\, q_0'$, r_3 varies from 0 to 1 and $s = 0$;

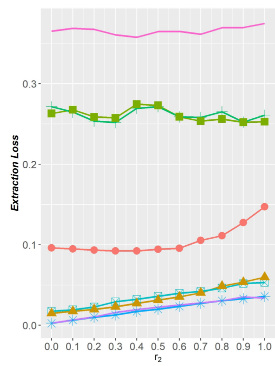
(IV) d varies from 8 to 26, $p_0 = 4$, $q_0 = q_0' = 1$ and $s = 0$;

(V) s varies from 0 to 0.4, $p_0 = 4$ and $q_0 = q_0' = 1$.

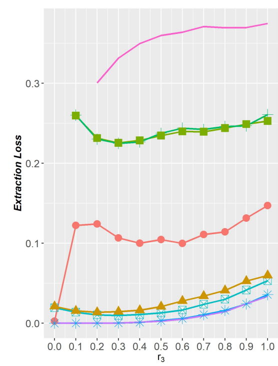
In this situation, we need to demonstrate not only whether the community nodes and background nodes can be well distinguished, but also whether the communities can be well separated. Let $\mathcal{C}_K = \{i \in [n] : \boldsymbol{c}(i) = K\}$, $\check{\mathcal{C}}_K = \{i \in [n] : \check{\boldsymbol{c}}(i) = K\}$, $\underline{\boldsymbol{c}} = \big(\underline{\boldsymbol{c}}(1), \cdots, \underline{\boldsymbol{c}}(n)\big)^\top$ with $\underline{\boldsymbol{c}}(i) = \mathbb{I}\{i \in \mathcal{C}_K\}$ and $\underline{\check{\boldsymbol{c}}} = \big(\underline{\check{\boldsymbol{c}}}(1), \cdots, \underline{\check{\boldsymbol{c}}}(n)\big)^\top$ with $\underline{\check{\boldsymbol{c}}}(i) = \mathbb{I}\{i \in \check{\mathcal{C}}_K\}$. Then, $\ell(\underline{\boldsymbol{c}}, \underline{\check{\boldsymbol{c}}})$ measures the loss of community extraction. On the other hand, we use the normalized mutual information (NMI) between $\boldsymbol{c}$ and



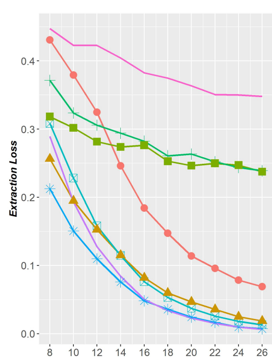
S.4(a) Setting (I)



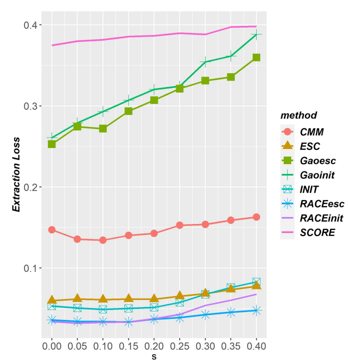
S.4(b) Setting (II)



S.4(c) Setting (III)



S.4(d) Setting (IV)



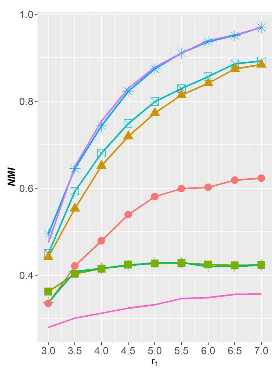
S.4(e) Setting (V)

Figure S.4: The performance of community extraction in the case of $K = 3$ for Settings (I)-(V).

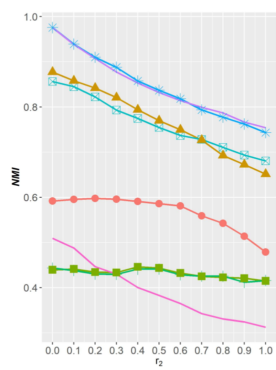
$\check{\boldsymbol{c}}$ on $[n]\backslash\{\mathcal{C}_K \cap \check{\mathcal{C}}_K\}$ to measure the accuracy of separating different communities. Note that NMI is a common criterion for evaluating clustering performance (Wang et al., 2020), which takes a value between 0 and 1.

A larger NMI value reflects a better clustering result. NMI is defined as follows. For any $\boldsymbol{c}_1 \in [K_1]^n$ and $\boldsymbol{c}_2 \in [K_2]^n$, define random variables X and Y with the joint distribution, $\mathbb{P}(X = k, Y = l) = |\{(i,j) : \boldsymbol{c}_1(i) = k, \boldsymbol{c}_2(j) = l\}|/n^2$ for each $k \in [K_1]$ and $l \in [K_2]$. Then the NMI between $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ is $NMI(X,Y) = 2MI(X,Y)/\big(H(X) + H(Y)\big)$, where $H(X) = -\sum_{k\in[K_1]} \mathbb{P}(X = k) \log \mathbb{P}(X = k)$ and $H(Y) = -\sum_{k\in[K_2]} \mathbb{P}(Y = k) \log \mathbb{P}(Y = k)$ are the information entropy of X and Y, respectively. Here, $MI(X,Y) = \sum_{k\in[K_1]} \sum_{k\in[K_2]} \mathbb{P}(X = k, Y = l) \log \frac{\mathbb{P}(X=k,Y=l)}{\mathbb{P}(X=k)\mathbb{P}(Y=l)}$ is the mutual information between X and Y.

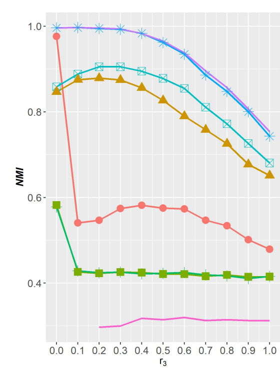
Set $n = 150$, $K = 3$, $d = 18$, $\boldsymbol{\omega} = (1 - s, 1, 1 + s)^\top/3$. Note that because M-E is designed to find overlapping communities, while there is no overlapping in the simulation settings, we exclude M-E from the following comparison. The results for $K = 3$ are summarized in Figures S.4 and S.5, which suggest that RACEinit and RACEesc have very good performance in both community extraction and community discovery (i.e. separating different communities). Note that in setting (III), when r_3 is relatively small, SCORE has missing values because it requires that the network is connected, hence it does not allow isolated nodes to exist in the network. Similarly, the results of Gaoinit and Gaoesc have some missing values.



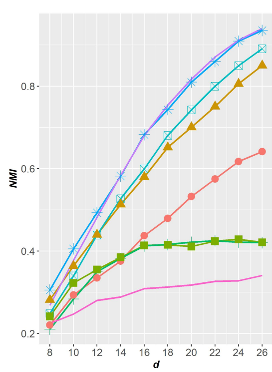
S.5(a) Setting (I)



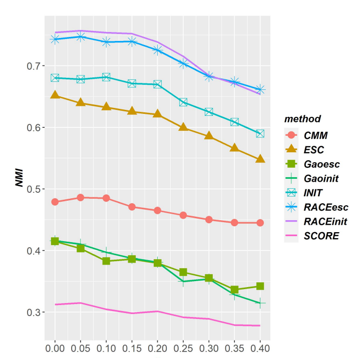
S.5(b) Setting (II)



S.5(c) Setting (III)



S.5(d) Setting (IV)



S.5(e) Setting (V)

Figure S.5: The performance of community discovery (i.e. separating different communities) in case of $K = 3$ for Settings (I)-(V).

S5. Additional discussions

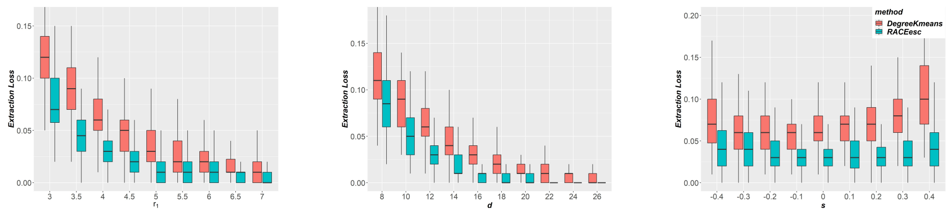
S.5.1 Limitations of degree separability in identifying background nodes

In this subsection, we discuss whether the separability of node degrees is sufficient to identify background nodes. In fact, the ability to distinguish communities and background nodes relies on the separability of edge-probability matrix. Degree separability is one aspect of this separability, which involves information loss and often cannot substitute for the separability of edge-probability matrix. In many cases, methods based solely on degree separability may even completely fail to differentiate between communities and background nodes.

To showcase this point, we consider one example with $K = 3$, $n_1 = n_2 = n_3 = n/3$ and

$$\boldsymbol{P} = \begin{bmatrix} p & 0 & q \\ 0 & p & q \\ q & q & q \end{bmatrix},$$

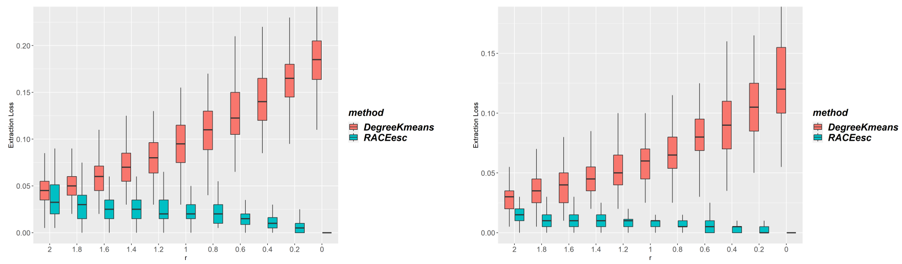
where $p = 2q$. The rows or columns of this matrix exhibit clear separability. However, for each node in one of the two communities and each background node, their expected degrees are approximately equal or equal to $(n-1)q$, based on which the two types of nodes cannot be distinguished.



S.6(a) Setting (I) S.6(b) Setting (II) S.6(c) Setting (III)

Figure S.6: The performance of RACE and DegreeKmeans in case of $K = 2$ for Settings (I)-(III).

Next, we present some numerical results to investigate the performance of a method based on degree separation in comparison with the proposed method. The method based on degree separation is constructed as follows: if $K = 2$, we apply K-means to $\{\sum_{j=1}^{n} A_{ij}\}_{i\in[n]}$ to obtain two clusters, where the cluster with the smaller centroid is considered as the set of background nodes, and the other cluster is considered as the community; if $K > 2$, we apply K-means to $\{\sum_{j=1}^{n} A_{ij}\}_{i\in[n]}$ to obtain $\widetilde{K}$ clusters and the cluster with the smaller centroid is considered as the set of background nodes, where $\widetilde{K}$ is selected from $\{2, \cdots, K\}$ by maximizing the Silhouette coefficient that is a measure of clustering effectiveness (Rousseeuw, 1987). It should be noted that the reason for selecting $\widetilde{K}$ instead of using a fixed K is that the average degrees of nodes in different communities often cannot be distinguished, as seen in the example above. From experience, directly using K-means to get K clusters will result in significantly poorer clustering outcomes. The



S.7(a) $p_1 = p_2 = 4q$

S.7(b) $p_1 = 5q$ and $p_2 = 4q$

Figure S.7: The performance of RACE and DegreeKmeans in case of $K = 3$.

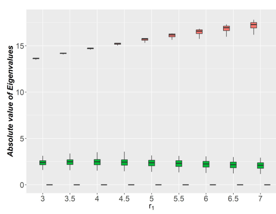
above method based on degree separation is written as DegreeKmeans.

The simulation settings are designed as follows. For the case of $K = 2$, we consider settings (I)-(III) used in Section 4. For the case the $K = 3$, we consider the following setting of $\boldsymbol{P}$:

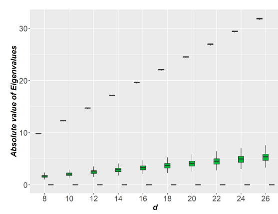
$$\boldsymbol{P} = \begin{bmatrix} p_1 & q_1 & q \\ q_1 & p_2 & q \\ q & q & q \end{bmatrix},$$

with $q_1 = rq$, $r \in [0, 2]$ and $n_1 : n_2 : n_3 = 3 : 5 : 7$. Two settings of p_1 and p_2 are considered: (1) $p_1 = p_2 = 4q$, (2) $p_1 = 5q$, $p_2 = 4q$. The simulation results of DegreeKmeans and RACEesc are presented in Figures $S.6$ and $S.7$ with $K = 2$ and $K = 3$, respectively. These results suggest that RACEesc significantly outperforms DegreeKmeans in identifying background nodes.

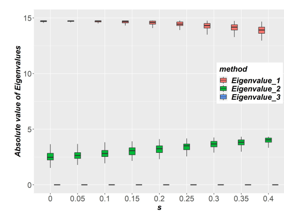
In summary, a simple method based on degree separation, such as DegreeKmeans, has limitations in identifying background nodes, especially when there is sufficient separability in the edge-probability matrix but not enough separability in average degrees. In contrast, our proposed methods make fuller use of the separability of the edge-probability matrix, thus yielding more competitive results.



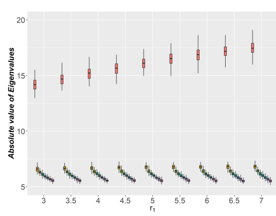
S.8(a) Setting (I) (for $\mathbb{E}(\boldsymbol{A})$)



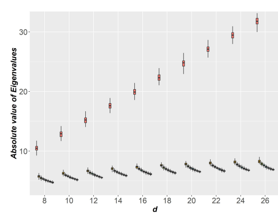
S.8(b) Setting (II) (for $\mathbb{E}(\boldsymbol{A})$)



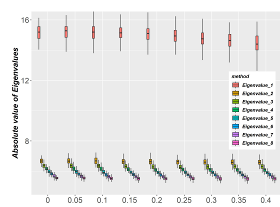
S.8(c) Setting (III) (for $\mathbb{E}(\boldsymbol{A})$)



S.8(d) Setting (I) (for $\boldsymbol{A}$)



S.8(e) Setting (II) (for $\boldsymbol{A}$)

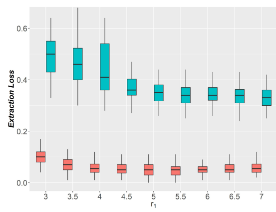


S.8(f) Setting (III) (for $\boldsymbol{A}$)

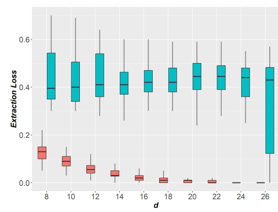
Figure S.8: The absolute values of the eigenvalues of $\boldsymbol{A}$ and $\mathbb{E}(\boldsymbol{A})$ in case of $K = 2$ for Settings (I)-(III).

S.5.2 Suitable initialization method based on spectral clustering

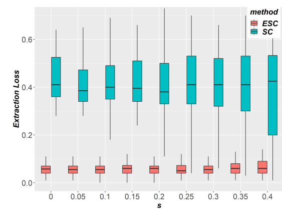
In some related existing studies on community detection without background nodes, such as Lei and Rinaldo (2015) and Gao et al. (2017), K-means is applied to the $n \times K$ matrix constructed by the leading K eigen-



S.9(a) Setting (I)



S.9(b) Setting (II)



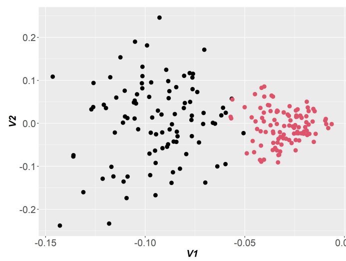
S.9(c) Setting (III)

Figure S.9: The performance of ESC and SC in case of $K = 2$ for Settings (I)-(III).

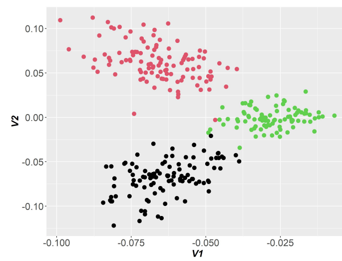
vectors of $\boldsymbol{A}$ to obtain K clusters. In fact, when a network has K communities without background nodes, the edge-probability matrix is generally full rank and the leading K eigenvalues are significantly larger than the other eigenvalues. In this case, selecting K eigenvectors to construct the matrix to be used in K-means is usually the optimal choice.

However, for networks with a certain number of background nodes, using the Kth leading eigenvector is usually not a good choice due to its low signal-to-noise ratio. Below, we will support this argument with an example. Under a network setting with $K - 1$ communities and a set of background nodes, we generated the edge-probability matrix $\boldsymbol{P}$ and the node labels $\boldsymbol{c}$ as in Setting (I)-(III) in Section 4, based on which we randomly generated the adjacency matrix $\boldsymbol{A}$. Then, we investigated the pattern of the eigenvalues of $\mathbb{E}(\boldsymbol{A})$ and $\boldsymbol{A}$, respectively. From Figures S.8(a)-S.8(c), we find that the absolute value of $\mathbb{E}(\boldsymbol{A})$'s Kth eigenvalue is generally moderately larger than 0, i.e. the $(K+1)$th eigenvalue, and much smaller than the $(K-1)$th eigenvalue. Hence, if $\mathbb{E}(\boldsymbol{A})$ is known, applying K-means to the matrix composed of the leading K eigenvectors of $\mathbb{E}(\boldsymbol{A})$ will be more competitive than applying it to the matrix composed of the leading $K-1$ eigenvectors. Since $\mathbb{E}(\boldsymbol{A})$ is unknown, we have to consider the eigenvectors of $\boldsymbol{A}$. From Figures S.8(d)-S.8(f), we find that although the Kth eigenvalue of $\boldsymbol{A}$ is still significantly smaller than the $(K-1)$th eigenvalue, which is similar to the $\mathbb{E}(\boldsymbol{A})$ case, it is no longer well separated from the $(K+1)$th eigenvalue, which is different from the $\mathbb{E}(\boldsymbol{A})$ case. This indicates that the signal-to-noise ratio of the Kth eigenvector of $\boldsymbol{A}$ is very low.

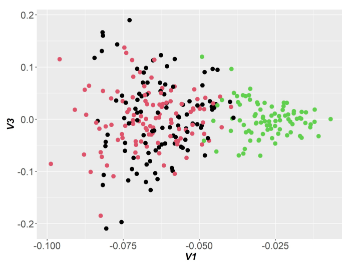
On this ground, using the Kth eigenvector of $\boldsymbol{A}$ may has a negative impact on clustering. Suggested by Figure S.9, the proposed ESC algorithm outperforms SC in terms of identifying background nodes, where the algorithm SC is the same as ESC, except that it applies K-means to the matrix composed of the leading K eigenvectors of $\boldsymbol{A}$. The poor performance of SC is due to the lack of cluster information in the Kth eigenvectors of $\boldsymbol{A}$, which is demonstrated in Figure S.10. Figure S.10(a) presents the scatter plot of the first two eigenvectors of $\boldsymbol{A}$ in case of $K = 2$, which suggests that the first eigenvector can clearly distinguish the clusters, but the second (Kth) eigenvector cannot distinguish the clusters at all. Similar results



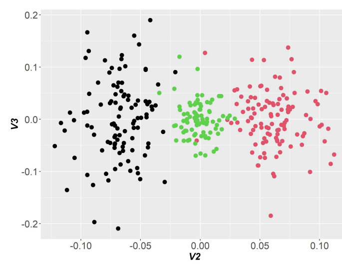
S.10(a) $K = 2$



S.10(b) $K = 3$



S.10(c) $K = 3$



S.10(d) $K = 3$

Figure S.10: Scatter plot of the eigenvectors of $\boldsymbol{A}$.

are presented in Figures S.10(b)-S.10(d) for the case of $K = 3$.

Hence, we ultimately decided to use the first $K - 1$ eigenvectors of $\boldsymbol{A}$ in ESC for clustering to achieve community extraction.

S.5.3 New challenges encountered

In this subsection, we will explain in detail the new challenges that our study has encountered. These challenges are described in terms of the algorithmic and theoretical aspects, respectively.

First, we illustrate the challenges encountered in the algorithmic aspect.

(1) Existing methods often struggle to find a suitable initialization method based on spectral clustering to handle networks with background n-odes. This is because for a network with background nodes, the signal-to-noise ratio of the Kth eigenvector of the adjacency matrix $\boldsymbol{A}$ may be significantly low during the eigen-decomposition process. This occurs despite the fact that the Kth eigenvector of the expected adjacency matrix $\mathbb{E}(\boldsymbol{A})$ does indeed contain a discernible signal; see Section S.5.2. Due to this phenomenon, when K is known, if we use the leading K eigenvectors of $\boldsymbol{A}$ as in Gao et al. (2017), it will be difficult to identify the background nodes, as presented in Figure S.9. In fact, we will likely disperse the background nodes into communities. When K is unknown, if we use the model selection methods such as in Yun and Proutiere (2016) and Ariu et al. (2023), we will generally select a number of clusters smaller than K. This typically leads to unsatisfactory model selection performance, as demonstrated in Table S.3.

To address this challenge, when K is known, the proposed initialization method, ESC, utilizes only the leading $K-1$ eigenvectors of the matrix $\boldsymbol{A}$. Subsequently, the K-means algorithm is applied based on these eigenvectors to obtain K clusters. When K is unknown, to select K clusters including $K-1$ communities and a set of background nodes, we have combined the methods that we proposed with the CBIC criterion.

(2) In the refinement step of many existing two-step methods, such as those proposed by Gao et al. (2017) and Gao et al. (2018), the fundamental concept is to assign a node to the cluster with which it has the closest connection. However, this approach may not effectively handle background nodes. This is primarily because background nodes do not exhibit a tendency to favor any particular community or the set of background nodes in terms of the probability of connection.

To address this challenge, we implement refinement similar to Yun and Proutiere (2016), based on the likelihood information of the community extraction model under study.

Next, we discuss the challenges encountered in the theoretical aspect. In fact, when establishing the asymptotic minimax risk in the network model including background nodes, many conditions on $(n_1, \cdots, n_K)$ and (p, q) used in existing studies are somewhat restrictive. Hence, the asymptotic minimax risk needs to be established under more relaxed conditions.

(1) To obtain the upper bound, Zhang and Zhou (2016) and Gao et al. (2018) used the following condition when $K = 2$: $Kn_k/n \in [\sqrt{3/5}, \sqrt{5/3}]$ for each $k \in [K]$. Further, Yun and Proutiere (2016) relaxed this con-

Table S.3: The performance of model selection by using RACEesc together with CBIC in comparison with the model selection methods in Yun and Proutiere (2016) and Ariu et al. (2023)

		True $K=2$			True $K=3$		
Setting	r_1	RACEesc	Y & P	A, P & Y	RACEesc	Y & P	A, P & Y
(I)	5.0	1.00	0	0	1.00	0	0
	5.5	1.00	0	0	1.00	0	0
	6.0	1.00	0	0	1.00	0	0
	6.5	1.00	0	0	1.00	0	0
	7.0	1.00	0	0	1.00	0	0
	d	RACEesc	Y & P	A, P & Y	RACEesc	Y & P	A, P & Y
(II)	30	1.00	0	0	0.89	0	0
	32	1.00	0	0	1.00	0	0
	34	1.00	0	0	1.00	0	0
	36	1.00	0	0	1.00	0	1.00
	38	1.00	0	1.00	1.00	0	1.00
	s	RACEesc	Y & P	A, P & Y	RACEesc	Y & P	A, P & Y
(III)	0.00	1.00	0	0	1.00	0	0
	0.06	1.00	0	0	1.00	0	0
	0.12	1.00	0	0	1.00	0	0
	0.18	1.00	0	0	0.96	0	0
	0.24	1.00	0	0	0.91	0	0

dition to $n_k \asymp n/K$ for each $k \in [K]$ and used an additional condition $1 < \lim_{n\to\infty} p/q < \infty$. In practical problems, n_K may be much larger or smaller than the number of community nodes. Hence, the conditions of $Kn_k/n \in [\sqrt{3/5}, \sqrt{5/3}]$ and $n_k \asymp n/K$ are unreasonable for background nodes. In addition, the condition $1 < \lim_{n\to\infty} p/q < \infty$ is also unreasonable, because it excludes the case where q is much smaller than p, which is very reasonable for background nodes.

These conditions are crucial for establishing the upper bound in the corresponding studies. Hence, we need to find a new way to establish the upper bound under more relaxed conditions. In fact, we imposed the following conditions on (n_1, n_2) and (p, q): $n_1 \in \Big[\lfloor \beta n \rfloor - 1, \lceil (1 -$

$\beta)n\rceil + 1\Big]$, $-\frac{\beta^2 n I_{t^*}(p,q)}{\log \beta} \to \infty$ and $\lim_{n\to\infty} p/q > 1$, which are much weaker than the above ones.

(2) To obtain the lower bound, Yun and Proutiere (2016) assumed that $1 < \lim_{n\to\infty} p/q = O(1)$ and $n_k \asymp n/K$ for each $k \in [K]$, and Gao et al. (2018) assumed that $1 < \lim_{n\to\infty} p/q = O(1)$. Similarly, these conditions are also unreasonable, so we consider establishing the lower bound under the following more relaxed conditions: $n_1 \in \Big[\lfloor \beta n\rfloor - 1, \lceil (1-\beta)n\rceil + 1\Big]$, $-\frac{\beta^2 n I_{t^*}(p,q)}{\log \beta} \to \infty$ and $p \asymp q$ (or $p \gg q$ together with some additional conditions).

Our approach to establishing the lower bound is partly inspired by Gao et al. (2018)'s proof framework. In their proof, it is relatively direct to verify the Lindeberg condition when using the Lindeberg theorem under the condition $1 < \lim_{n\to\infty} p/q = O(1)$. However, under more relaxed conditions, it becomes much more challenging to determine whether the Lindeberg condition holds. To solve this problem, we have developed a new approach, not based on the Lindeberg theorem, but on the Liapounov theorem.

S6. The Main Proof

We provide some necessary lemmas in the proof, and we will provide proof of all lemmas at the end. Firstly, we provide the following lemmas, which give the properties of t^* and I_{t^*}, respectively.

Lemma 1. *For any $0 < q < p = (1+\delta)q < 1 - \epsilon_0$, where $\delta > 0$ and ϵ_0 is a small positive constant. Then,*

(1) *$t^* \in (0,1)$ defined in (3.11) is the only maximum point of $I_t(p,q)$ on $t \geq 0$, and $1 - t^*$ is the only maximum point of $I_t(q,p)$ on $t \geq 0$;*

(2) *There exists a small constant $C > 0$ such that $t^* \in [C, 1-C]$ when $\delta \lesssim 1$. And $t^* = \big(1 + o(1)\big)\frac{\log(\log(1+\delta))}{\log(1+\delta)}$ when $\delta \to \infty$.*

Lemma 2. *For any $0 < q < p = (1+\delta)q < 1 - \epsilon_0$, where $\delta > 0$ and ϵ_0 is a small positive constant. Then,*

(1) *$I_{t^*} \asymp \delta^2 q$ when $\delta \lesssim 1$;*

(2) *$t^* p \lesssim I_{t^*} \lesssim p$ when $\delta \to \infty$.*

For any $\boldsymbol{c} \in [2]^n$ and its estimation $\hat{\boldsymbol{c}}$, define

$$\tilde{\ell}(\boldsymbol{c}, \hat{\boldsymbol{c}}) = \min_{\pi:[2]\to[2]} \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}\{\pi(\boldsymbol{c}(i)) \neq \hat{\boldsymbol{c}}(i)\},$$

which is a recognized loss in community discovery. Assume that $\tilde{\boldsymbol{c}}^0$ is an initialization algorithm for community discovery.

Condition S1. *For a given positive sequence $\{\gamma_n\}$, there exist a constant $C_0 > 0$, such that*

$$\inf_{(\boldsymbol{P},\boldsymbol{c})\in\Theta_n(p,q,\beta)} \mathbb{P}_{P,\boldsymbol{c}}\left\{\tilde{\ell}\left(\boldsymbol{c},\tilde{\boldsymbol{c}}^0\right) \le \gamma_n\right\} \ge 1 - n^{-(1+C_0)}, \tag{S6.3}$$

where $\tilde{\boldsymbol{c}}^0$ is an initialization algorithm for community discovery.

Then, we discuss the distance between $\tilde{\boldsymbol{P}}^0 = (\tilde{P}^0_{kl})_{K\times K}$ obtained in the initialization step and the real parameter $\boldsymbol{P}$, where

$$\tilde{P}^0_{kl} = \begin{cases} \dfrac{\sum\limits_{u<v} A_{uv}\mathbb{I}\{\tilde{\boldsymbol{c}}^0(u)=k,\tilde{\boldsymbol{c}}^0(v)=k\}}{\frac{1}{2}\tilde{n}^0_k(\tilde{n}^0_k-1)}, & k = l; \\ \dfrac{\sum\limits_{u,v\in[n]} A_{uv}\mathbb{I}\{\tilde{\boldsymbol{c}}^0(u)=k,\tilde{\boldsymbol{c}}^0(v)=l\}}{\tilde{n}^0_k\tilde{n}^0_l}, & k \ne l, \end{cases}$$

where $\tilde{n}^0_k = \sum_{j\in[n]} \mathbb{I}\{\tilde{\boldsymbol{c}}^0(j) = k\}$. Let $p^0 = \max\{P^0_{11}, P^0_{22}\}$, and $q^0 = \frac{\tilde{n}^0_1\tilde{P}^0_{1k^*}+\tilde{n}^0_2\tilde{P}^0_{2k^*}/2}{\tilde{n}^0_1+\tilde{n}^0_2/2}$, where $k^* = \arg\min_{k\in[2]} \tilde{P}^0_{kk}$. The community extraction problem not only provides clustering results $\tilde{\boldsymbol{c}}^0$, but also identifies background nodes. Therefore, we give $\boldsymbol{c}^0$ based on $\tilde{\boldsymbol{c}}^0$. Specifically, $\boldsymbol{c}^0(i) = 2$ if $\tilde{\boldsymbol{c}}^0(i) = k^*$, and $\boldsymbol{c}^0(i) = 1$ if $\tilde{\boldsymbol{c}}^0(i) \in [2]\setminus\{k^*\}$. Based on the above definition, we provide the following lemma, which not only characterizes the accuracy of connection probability matrix estimation, but also describes the relationship between community discovery results and community extraction results.

Lemma 3. *Suppose that as $n \to \infty$, $\beta n\frac{(p-q)^2}{p} \to \infty$ and Condition S1 holds with the requirement for γ_n replaced by $\gamma_n = o(-\frac{\beta}{\log\beta})$. Then there exists a* *constant* $C > 0$ *such that*

$$\inf_{(\boldsymbol{P},\boldsymbol{c})\in\Theta_n(p,q,\beta)} \mathbb{P}_{\boldsymbol{P},\boldsymbol{c}}\left(\max\left\{\left|p^0 - p\right|, \left|q^0 - q\right|\right\} \leq \eta(p-q)\right) \geq 1 - Cn^{-(1+C_0)}, \tag{S6.4}$$

where $\eta = C'(\eta_1 + \eta_2)$ *with* $\eta_1 = \sqrt{\frac{p}{\beta n(p-q)^2}}$ *and* $\eta_2 = -\gamma_n \frac{\log\beta}{\beta}$, *for some large constant* $C' > 0$. *Moreover,* $\ell\left(\boldsymbol{c}, \boldsymbol{c}^0\right) = \tilde{\ell}\left(\boldsymbol{c}, \tilde{\boldsymbol{c}}^0\right)$ *with probability at least* $1 - Cn^{-(1+C_0)}$.

Lemma 3 is similar to Lemma 1 in Gao et al. (2017), we also provided a detailed proof at the end for the completeness of proof.

S.6.1 Proof of Theorem 4

Firstly, because Theorem 4 will be used to prove Theorem 1, we will first provide a proof of Theorem 4.

Proof of Theorem 4. Let $\boldsymbol{c}$ be the true label. $\boldsymbol{A}^{\tau}$, $\tilde{\boldsymbol{c}}^0_{\mathsf{init}}$, $\boldsymbol{c}^0_{\mathsf{init}}$ and $\{\tilde{\boldsymbol{\nu}}_k\}_{k\in[2]}$ come from Algorithm 2, where $\{\tilde{\boldsymbol{\nu}}_k\}_{k\in[2]}$ are taken so that the left side of the inequality of (2.3) reaches the minimum. Let $\boldsymbol{M}' = (P_{\boldsymbol{c}(i)\boldsymbol{c}(j)})_{n\times n}$, whose rank is 2. For all $k \in [2]$, define $T_k = \{i : \boldsymbol{c}(i) = k, \|\tilde{\boldsymbol{V}}_i - \boldsymbol{M}'_i\|_2 < b\}$ as in Lei and Rinaldo (2015) where $\tilde{\boldsymbol{V}}_i = \tilde{\boldsymbol{\nu}}_{\tilde{\boldsymbol{c}}^0_{\mathsf{init}}(i)}$ and $b = \frac{1}{2}\sqrt{\beta n}(p-q)$. Note that $\|\boldsymbol{M}'_i - \boldsymbol{M}'_j\|_2 \geq 2b$ for all i, j with $\boldsymbol{c}(i) \neq \boldsymbol{c}(j)$.

Finally, we will prove that $n\tilde{\ell}\left(\boldsymbol{c}, \tilde{\boldsymbol{c}}^0_{\mathsf{init}}\right) \leq \sum_{k=1}^{2} |T_k^c|$, so let's first the upper bound of $\sum_{k=1}^{2}|T_k^c|$. By the definition of T_k we have

$$\begin{aligned} b^2\sum_{k=1}^{2}|T_k^c| &\leq \|\tilde{\boldsymbol{V}}-\boldsymbol{M}'\|_{\mathrm{F}}^2 \\ &\leq 2\|\tilde{\boldsymbol{V}}-\tilde{\boldsymbol{M}}\|_{\mathrm{F}}^2+2\|\tilde{\boldsymbol{M}}-\boldsymbol{M}'\|_{\mathrm{F}}^2 \\ &\leq 2(2+\xi)\|\tilde{\boldsymbol{M}}-\boldsymbol{M}'\|_{\mathrm{F}}^2 \end{aligned} \tag{S6.5}$$

where $\tilde{\boldsymbol{M}}$ is defined in (2.2). Then, the next task is to present the bound of $\|\tilde{\boldsymbol{M}}-\boldsymbol{M}'\|_{\mathrm{F}}^2$, which can be written as

$$\begin{aligned} &\|\tilde{\boldsymbol{M}}-\boldsymbol{A}^\tau+\boldsymbol{A}^\tau-\boldsymbol{M}'\|_{\mathrm{F}}^2 \\ =&\|\tilde{\boldsymbol{M}}-\boldsymbol{A}^\tau\|_{\mathrm{F}}^2+\|\boldsymbol{A}^\tau-\boldsymbol{M}'\|_{\mathrm{F}}^2-2\langle\boldsymbol{A}^\tau-\tilde{\boldsymbol{M}},\boldsymbol{A}^\tau-\boldsymbol{M}'\rangle \\ \leq&2\langle\tilde{\boldsymbol{M}}-\boldsymbol{M}',\boldsymbol{A}^\tau-\boldsymbol{M}'\rangle \\ \leq&\frac{1}{2}\|\tilde{\boldsymbol{M}}-\boldsymbol{M}'\|_{\mathrm{F}}^2+2\sup_{\substack{rank(\boldsymbol{M}_0)\leq 4\\ \|\boldsymbol{M}_0\|_{\mathrm{F}}=1,\, M_0=M_0^\top}}|\langle\boldsymbol{M}_0,\boldsymbol{A}^\tau-\boldsymbol{M}'\rangle|^2. \end{aligned} \tag{S6.6}$$

The first inequality holds since $\|\boldsymbol{A}^\tau-\tilde{\boldsymbol{M}}\|_{\mathrm{F}}^2\leq\|\boldsymbol{A}^\tau-\boldsymbol{M}'\|_{\mathrm{F}}^2$, and the second inequality holds since $\mathrm{rank}(\tilde{\boldsymbol{M}}-\boldsymbol{M}')\leq 4$. And then, we have

$$\|\tilde{\boldsymbol{M}}-\boldsymbol{M}'\|_{\mathrm{F}}^2\leq 4\sup_{\substack{rank(\boldsymbol{M}_0)\leq 4\\ \|\boldsymbol{M}_0\|_{\mathrm{F}}=1,\, M_0=M_0^\top}}|\langle\boldsymbol{M}_0,\boldsymbol{A}^\tau-\boldsymbol{M}'\rangle|^2. \tag{S6.7}$$

Then, apply singular value decomposition to $\boldsymbol{M}_0$, that is $\boldsymbol{M}_0=\sum_{k=1}^{4}s_k\alpha_k\alpha_k^\top$. Then,

$$\begin{aligned} |\langle\boldsymbol{M}_0,\boldsymbol{A}^\tau-\boldsymbol{M}'\rangle|\leq\sum_{k=1}^{4}s_k|\alpha_k^\top(\boldsymbol{A}^\tau-\boldsymbol{M}')\alpha_k|\leq&\|\boldsymbol{A}^\tau-\boldsymbol{M}'\|_{\mathrm{op}}\sum_{k=1}^{4}s_k \\ \leq&4\|\boldsymbol{A}^\tau-\boldsymbol{M}'\|_{\mathrm{op}}. \end{aligned} \tag{S6.8}$$

The last inequality holds since $s_k \leq \|\boldsymbol{M}_0\|_{\mathrm{op}} = 1$. By Lemma 5 in Gao et al. (2017), for any $C' > 0$, there exists some $C > 0$, such that $\|\boldsymbol{A}^\tau - \boldsymbol{M}'\|_{\mathrm{op}} \leq C\sqrt{np+1}$ with probability at least $1 - n^{-C'}$ uniformly over $\tau \in [C_1(np+1), C_2(np+1)]$ for some sufficiently large constants C_1, C_2. Then, we have $\|\boldsymbol{A}^\tau - \boldsymbol{M}'\|_{\mathrm{op}} \leq C\sqrt{np+1}$, with probability at least $1 - n^{-(1+C')}$.

To sum up, by (S6.5), (S6.6), (S6.7) and (S6.8), we have

$$\sum_{k=1}^{2} |T_k^c| \quad \leq \quad C(1+\xi)\frac{np+1}{\beta n(p-q)^2}, \tag{S6.9}$$

with probability at least $1 - n^{-(1+C')}$.

Next, we will prove that $n\tilde{\ell}\left(\boldsymbol{c}, \tilde{\boldsymbol{c}}^0_{\mathsf{init}}\right) \leq \sum_{k=1}^{2} |T_k^c|$. By the assumption that $\frac{-\beta}{\log \beta}\frac{\beta n(p-q)^2}{p} \to \infty$ as $n \to \infty$, we have $|T_k^c| \leq C(1+\xi)\frac{np+1}{\beta n(p-q)^2} = o(\beta n)$ for each $k \in [2]$ with probability at least $1 - n^{-(1+C')}$, which means based on the event (S6.9), $T_k \neq \emptyset$ for each $k \in [2]$. In such case, T_1 and T_2 have the following properties. First, for any $i \in T_1$ and $j \in T_2$, $\tilde{\boldsymbol{c}}^0_{\mathsf{init}}(i) \neq \tilde{\boldsymbol{c}}^0_{\mathsf{init}}(j)$. Otherwise, if $\tilde{\boldsymbol{c}}^0_{\mathsf{init}}(i) = \tilde{\boldsymbol{c}}^0_{\mathsf{init}}(j)$, we have $\|\boldsymbol{M}'_i - \boldsymbol{M}'_j\|_2 \leq \|\tilde{\boldsymbol{V}}_i - \boldsymbol{M}'_i\|_2 + \|\tilde{\boldsymbol{V}}_j - \boldsymbol{M}'_j\|_2 < 2b$, which contradicts with the fact that $\|\boldsymbol{M}'_i - \boldsymbol{M}'_j\|_2 \geq 2b$ for all i, j with $\boldsymbol{c}(i) \neq \boldsymbol{c}(j)$. Second, fix any $k \in [2]$, and for any $i, j \in T_k$, $\tilde{\boldsymbol{c}}^0_{\mathsf{init}}(i) = \tilde{\boldsymbol{c}}^0_{\mathsf{init}}(j)$, since there are at most two values for $\tilde{\boldsymbol{c}}^0_{\mathsf{init}}(u)$ for each $u \in [n]$.

Combining the above two properties of $\{T_k\}_{k\in[2]}$, we have $n\tilde{\ell}\left(\boldsymbol{c}, \tilde{\boldsymbol{c}}^0_{\mathsf{init}}\right) \leq$

$\sum_{k=1}^{2}|T_k^c|$. Then, we have that for any $(\boldsymbol{P},\boldsymbol{c})\in\Theta_n(p,q,\beta)$,

$$\mathbb{P}_{P,\boldsymbol{c}}\left\{n\tilde{\ell}\left(\boldsymbol{c},\tilde{\boldsymbol{c}}_{\mathsf{init}}^0\right)\le C(1+\xi)\frac{np+1}{\beta n(p-q)^2}\right\}\ge 1-n^{-(1+C')},$$

for constants $C,C'>0$. Based on assumption $\frac{-\beta}{\log\beta}\frac{\beta n(p-q)^2}{p}\to\infty$, we have

$$\mathbb{P}_{P,\boldsymbol{c}}\left\{\tilde{\ell}\left(\boldsymbol{c},\tilde{\boldsymbol{c}}_{\mathsf{init}}^0\right)\le\gamma_n\right\}\ge 1-n^{-(1+C')},$$

for some constant $C'>0$ with $\gamma_n=o(-\beta/\log\beta)$. And by Lemma 3, we have

$$\mathbb{P}_{P,\boldsymbol{c}}\left\{\ell\left(\boldsymbol{c},\boldsymbol{c}_{\mathsf{init}}^0\right)\le C(1+\xi)\frac{np+1}{\beta n(p-q)^2}\right\}\ge 1-3n^{-(1+C')}.$$

Since the results are independent of $(\boldsymbol{P},\boldsymbol{c})$, the proof is completed.

□

S.6.2 Proof of Proposition 1

Next lemma provides the upper bound of the probability of $\hat{\boldsymbol{c}}_{p,q}(i)\neq\boldsymbol{c}(i)$ for each node i.

Lemma 4. *Suppose that as $n\to\infty$, $\beta nI_{t^*}\to\infty$ and the initialization algorithm $\mathbf{c}^0$ satisfies Condition 1 with $\gamma_n=o(\beta)$. If $\lim_{n\to\infty}p/q>1$, then for $\{\hat{\boldsymbol{c}}_{p,q}(i)\}_{i\in[n]}$ from (3.12), there is a sequence $\eta\to 0$ such that*

$$\sup_{(\boldsymbol{P},\boldsymbol{c})\in\Theta_n(p,q,\beta)}\max_{i\in[n]}\mathbb{P}_{\boldsymbol{P},\boldsymbol{c}}\big(\hat{\boldsymbol{c}}_{p,q}(i)\neq\boldsymbol{c}(i)\big)\le\exp\big(-(1-\eta)\beta nI_{t^*}\big)+2n^{-(1+C_0)},\tag{S6.10}$$

where the constant $C_0 > 0$ is from Condition 1. And when $\lim_{n\to\infty} p/q > 1$ is replaced by $\lim_{n\to\infty} p/q = 1$, if $\gamma_n = o(-\beta(p-q)/p)$, then (S6.10) *still holds.*

Proof of Proposition 1. Under the assumption of Proposition 1, by Lemma 4, we have that for $\{\hat{\boldsymbol{c}}_{p,q}(i)\}_{i\in[n]}$ from (3.12), there is a sequence $\eta \to 0$ such that

$$\sup_{(\boldsymbol{P},\boldsymbol{c})\in\Theta_n(p,q,\beta)} \max_{i\in[n]} \mathbb{P}_{\boldsymbol{P},\boldsymbol{c}}\big(\hat{\boldsymbol{c}}_{p,q}(i) \neq \boldsymbol{c}(i)\big) \leq \exp\big(-(1-\eta)\beta n I_{t^*}\big) + 2n^{-(1+C_0)}.$$

And then, we have

$$\begin{aligned}\mathbb{E}\ell(\boldsymbol{c}, \hat{\boldsymbol{c}}_{p,q}) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}\big(\hat{\boldsymbol{c}}_{p,q}(i) \neq \boldsymbol{c}(i)\big) \\ &\leq \exp\big(-(1-\eta)\beta n I_{t^*}\big) + 2n^{-(1+C_0)}.\end{aligned}$$

If $\lim_{n\to\infty} \frac{\beta n I_{t^*}}{\log n} \leq 1$, we have $\exp\big(-(1-\eta)\beta n I_{t^*}\big) \geq n^{-(1-\tilde{\eta})}$ for some $\tilde{\eta} \to 0$. Hence, $\exp\big(-(1-\eta)\beta n I_{t^*}\big) \gg n^{-(1+C_0)}$, which implies that $\mathbb{E}\ell(\boldsymbol{c}, \hat{\boldsymbol{c}}_{p,q}) \leq \exp\big(-(1-\eta)\beta n I_{t^*}\big)$ for some positive sequence $\eta \to 0$.

If $\lim_{n\to\infty} \frac{\beta n I_{t^*}}{\log n} > 1$, there exists a small $\varepsilon > 0$ such that $1 + \varepsilon < \lim_{n\to\infty} \frac{(1-\eta)\beta n I_{t^*}}{\log n}$. And we have $\exp\big(-(1-\eta)\beta n I_{t^*}\big) \leq n^{-(1+\varepsilon/2)}$, which means $\mathbb{E}\ell(\boldsymbol{c}, \hat{\boldsymbol{c}}_{p,q}) \leq n^{-(1+C)}$ for some small positive constant C. This situation means that $\hat{\boldsymbol{c}}_{p,q}$ exactly restored the label $\boldsymbol{c}$ in the expected sense and the specific minimax risk rate is not as important. □

S.6.3 Proof of Theorem 1

Proof of Theorem 1. Firstly, $-\frac{\beta^2 n I_{t^*}}{\log\beta} \to \infty$ can lead to $-\frac{\beta^2 n(p-q)^2}{p\log\beta} \to \infty$. Specifically, by Lemma 2, when $p \asymp q$, we have $I_{t^*} \asymp \frac{(p-q)^2}{p}$ and when $p \gg q$, we have $I_{t^*} \lesssim p$ and $\frac{(p-q)^2}{p} \asymp p$ which can deduce $-\frac{\beta^2 n(p-q)^2}{p\log\beta} \to \infty$, which demonstrate that the conditions of Theorem 1 are sufficient conditions for the condition of Theorem 4.

Moreover, under the assumption of Theorem 1, by Theorem 4, there exist a constant $C_0 > 0$ and a positive sequence $\gamma_n = o(\beta)$ when $\lim_{n\to\infty} \frac{p}{q} > 1$ and $\gamma_n = o(\beta\frac{p-q}{p})$ when $\lim_{n\to\infty} \frac{p}{q} = 1$. such that

$$\begin{aligned}\inf_{(\boldsymbol{P},\boldsymbol{c})\in\Theta_n(p,q,\beta)} \min_{i\in[n]} \mathbb{P}_{P,\boldsymbol{c}}\left\{\ell\left(\boldsymbol{c}_{-i}, \boldsymbol{c}^0_{\mathsf{init},-i}\right) \leq \gamma_n\right\} &\geq 1-(n-1)^{-(1+2C_0)} \\ &\geq 1-n^{-(1+C_0)},\end{aligned}$$

since the result of Theorem 4 are independent of i. □

S.6.4 Proof of Theorem 2

Proof of Theorem 2. For simplicity, $\Theta_n(p,q,\beta)$ is abbreviated as Θ in the following proof. First, we will construct some subspaces of Θ, which can be sufficient for deriving the lower bound.

1. When $\lim\limits_{n\to\infty}\beta < \frac{1}{2}$, there exists a $\boldsymbol{c}^* \in [2]^n$ that satisfies $n_1(\boldsymbol{c}^*) = \lfloor\beta n\rfloor$. Choose $\eta_0\beta n$ nodes from $\{i : \boldsymbol{c}^*(i) = 2\}$ and put them in set $\mathcal{T}$, where η_0

tends to 0 and satisfies $-\log(\eta_0\beta) = o(\beta n I_{t^*})$. Define $\Theta_n^1(p,q,\beta)$ as

$$\Big\{(P,\boldsymbol{c}) \in \Theta_n(p,q,\beta) : P_{11} = p, P_{12} = P_{22} = q,\ \textit{for all}\ i \in \mathcal{T}^c, \boldsymbol{c}(i) = \boldsymbol{c}^*(i)\Big\}.$$

Similarly, $\Theta_n^1(p,q,\beta)$ is abbreviated as Θ^1. Then, we have

$$\begin{aligned}
\inf_{\hat{\boldsymbol{c}}} \sup_{(\boldsymbol{P},\boldsymbol{c})\in\Theta} \mathbb{E}\ell(\boldsymbol{c},\hat{\boldsymbol{c}}) \geq & \inf_{\hat{\boldsymbol{c}}} \sup_{(\boldsymbol{P},\boldsymbol{c})\in\Theta^1} \mathbb{E}\ell(\boldsymbol{c},\hat{\boldsymbol{c}}) \\
\geq & \inf_{\hat{\boldsymbol{c}}} \frac{1}{|\Theta^1|} \sum_{(\boldsymbol{P},\boldsymbol{c})\in\Theta^1} \mathbb{E}\ell(\boldsymbol{c},\hat{\boldsymbol{c}}) \\
= & \inf_{\hat{\boldsymbol{c}}} \frac{1}{|\Theta^1|} \sum_{(\boldsymbol{P},\boldsymbol{c})\in\Theta^1} \frac{1}{n}\sum_{i=1}^n \mathbb{P}\big(\hat{\boldsymbol{c}}(i) \neq \boldsymbol{c}(i)\big) \\
\geq & \frac{1}{n|\Theta^1|} \sum_{i=1}^n \inf_{\hat{\boldsymbol{c}}(i)} \sum_{(\boldsymbol{P},\boldsymbol{c})\in\Theta^1} \mathbb{P}\big(\hat{\boldsymbol{c}}(i) \neq \boldsymbol{c}(i)\big) \\
= & \frac{1}{n|\Theta^1|} \sum_{i\in\mathcal{T}} \inf_{\hat{\boldsymbol{c}}(i)} \sum_{(\boldsymbol{P},\boldsymbol{c})\in\Theta^1} \mathbb{P}\big(\hat{\boldsymbol{c}}(i) \neq \boldsymbol{c}(i)\big).
\end{aligned} \tag{S6.11}$$

The last inequality holds since for any $i \in \mathcal{T}^c$, $\boldsymbol{c}(i) = \boldsymbol{c}^*(i)$.

Then, for any given $i \in \mathcal{T}$ and $k \in [2]$, let $\Theta^{(i,k)} = \Theta_n^{(i,k)}(p,q,\beta) = \big\{(P,\boldsymbol{c}) \in \Theta_n^1(p,q,\beta) : \boldsymbol{c}(i) = k\big\}$. We can see that $\Theta^{(i,1)} \cup \Theta^{(i,2)} = \Theta^1$ and $\Theta^{(i,1)} \cap \Theta^{(i,2)} = \emptyset$. Moreover, there is a one-to-one correspondence between the elements in $\Theta^{(i,1)}$ and $\Theta^{(i,2)}$. For example, for any $\boldsymbol{c}_1$ in $\mathcal{C}(\Theta^{(i,1)}) = \big\{\boldsymbol{c} : (P,\boldsymbol{c}) \in \Theta^{(i,1)}\big\}$, there is a unique $\boldsymbol{c}_2$ in $\mathcal{C}(\Theta^{(i,2)}) = \big\{\boldsymbol{c} : (P,\boldsymbol{c}) \in \Theta^{(i,2)}\big\}$ with $\boldsymbol{c}_2(j) = \boldsymbol{c}_1(j)$ for all $j \neq i$ corresponding to it, and for any $\boldsymbol{c}_2$ in $\mathcal{C}(\Theta^{(i,2)})$, there is also a unique $\boldsymbol{c}_1$ in $\mathcal{C}(\Theta^{(i,1)})$ with $\boldsymbol{c}_1(j) = \boldsymbol{c}_2(j)$ for all

$j \neq i$ corresponding to it. Then, we have

$$\begin{aligned}
&\inf_{\hat{\boldsymbol{c}}} \sup_{(\boldsymbol{P},\boldsymbol{c})\in\Theta} \mathbb{E}\ell(\boldsymbol{c},\hat{\boldsymbol{c}})\\
\geq& \frac{1}{n|\Theta^1|}\sum_{i\in\mathcal{T}}\inf_{\hat{\boldsymbol{c}}(i)}\sum_{k=1}^{2}\sum_{(\boldsymbol{P},\boldsymbol{c})\in\Theta^{(i,k)}}\mathbb{P}\big(\hat{\boldsymbol{c}}(i)\neq\boldsymbol{c}(i)\big)\\
\geq& \frac{1}{n|\Theta^1|}\sum_{i\in\mathcal{T}}\sum_{(\boldsymbol{P},\boldsymbol{c})\in\Theta^{(i,1)}}\inf_{\hat{\boldsymbol{c}}(i)}\Big[\mathbb{P}_{\boldsymbol{P},\boldsymbol{c}}\left(\hat{\boldsymbol{c}}(i)\neq 1\right)+\mathbb{P}_{\boldsymbol{P},\sigma^i[\boldsymbol{c}]}\big(\hat{\boldsymbol{c}}(i)\neq 2\big)\Big], \qquad \text{(S6.12)}
\end{aligned}$$

where $\sigma^i[\boldsymbol{c}] = \Big(\boldsymbol{c}(1),\cdots,\boldsymbol{c}(i-1),\sigma\big(\boldsymbol{c}(i)\big),\boldsymbol{c}(i+1),\cdots,\boldsymbol{c}(n)\Big)^\top$, and for any x in $\{1,2\}$, $\sigma(x)=1$ if $x=2$, otherwise $\sigma(x)=2$. Then,

$$\begin{aligned}
\inf_{\hat{\boldsymbol{c}}} \sup_{(\boldsymbol{P},\boldsymbol{c})\in\Theta} \mathbb{E}\ell(\boldsymbol{c},\hat{\boldsymbol{c}}) \geq& \frac{\eta_0\beta}{2}\exp\left(-(1+\eta)\beta n I_{t^*}\right)\\
=& \exp\left(-\big(1+o(1)\big)\beta n I_{t^*}\right). \qquad \text{(S6.13)}
\end{aligned}$$

Combining $|\Theta^{(i,1)}| = |\Theta^{(i,2)}| = \frac{1}{2}|\Theta^1|$, Lemma 5 and $|\mathcal{T}| = \eta_0\beta n$, the inequality in (S6.13) is established for some $\eta\to 0$. The equality in (S6.13) holds since $-\log(\eta_0\beta) = o(\beta n I_{t^*})$.

2. When $\lim_{n\to\infty}\beta = \frac{1}{2}$, if n is odd, define

$$\begin{aligned}
\Theta^2 &= \Theta_n^2(p,q,\beta)\\
&= \Bigg\{(P,\boldsymbol{c})\in\Theta_n(p,q,\beta): P_{11}=p, P_{12}=P_{22}=q,\ \textit{for all } k\in[2],\\
&\qquad n_k(\boldsymbol{c})\in\left\{\frac{n-1}{2},\frac{n+1}{2}\right\}\Bigg\}.
\end{aligned}$$

If n is even, define

$$\Theta^2 = \Theta_n^2(p,q,\beta)$$

$$= \Big\{(P,\boldsymbol{c}) \in \Theta_n(p,q,\beta) : P_{11} = p, P_{12} = P_{22} = q,\ \textit{for all } k \in [2],$$
$$n_k(\boldsymbol{c}) \in \Big\{\frac{n}{2}-1, \frac{n}{2}, \frac{n}{2}+1\Big\}\Big\}.$$

In addition, for any $i \in [n]$ and $\boldsymbol{c} \in \Theta^2$, define

$$\psi^i(\boldsymbol{c}) = \begin{cases} \sigma^i[\boldsymbol{c}], & n_{\boldsymbol{c}(i)}(\boldsymbol{c}) \geq \frac{n}{2}, \\ \boldsymbol{c}, & \text{otherwise,} \end{cases}$$

and

$$\widetilde{\Theta}^{(i,1)} = \widetilde{\Theta}_n^{(i,1)}(p,q,\beta) = \Big\{(P,\boldsymbol{c}) : (P,\boldsymbol{c}) \in \Theta^2, \psi^i(\boldsymbol{c}) \neq \boldsymbol{c}\Big\},$$
$$\widetilde{\Theta}^{(i,2)} = \widetilde{\Theta}_n^{(i,2)}(p,q,\beta) = \Big\{(P,\tilde{\boldsymbol{c}}) : (P,\boldsymbol{c}) \in \widetilde{\Theta}^{(i,1)}, \tilde{\boldsymbol{c}} = \psi^i(\boldsymbol{c})\Big\}.$$

Note that $\widetilde{\Theta}^{(i,1)} \subset \Theta^2$, $\widetilde{\Theta}^{(i,2)} \subset \Theta^2$, $|\widetilde{\Theta}^{(i,1)}| = |\widetilde{\Theta}^{(i,2)}|$ and there is a one-to-one mapping between $\widetilde{\Theta}^{(i,1)}$ and $\widetilde{\Theta}^{(i,2)}$. In particular, for any $\boldsymbol{c}_1$ in $\mathcal{C}(\widetilde{\Theta}^{(i,1)}) = \big\{\boldsymbol{c} : (P,\boldsymbol{c}) \in \widetilde{\Theta}^{(i,1)}\big\}$, there is only one $\boldsymbol{c}_2$ in $\mathcal{C}(\widetilde{\Theta}^{(i,2)}) = \big\{\boldsymbol{c} : (P,\boldsymbol{c}) \in \widetilde{\Theta}^{(i,2)}\big\}$ corresponding to it, with $\boldsymbol{c}_1(j) = \boldsymbol{c}_2(j)$ for any $j \neq i$. Besides, there must exist some i in $[n]$ such that $\frac{|\widetilde{\Theta}^{(i,1)}|}{|\Theta^2|} \geq \frac{1}{2}$. Because if for all $i \in [n]$, $\frac{|\widetilde{\Theta}^{(i,1)}|}{|\Theta^2|} < \frac{1}{2}$, we have $n\frac{|\Theta^2|}{2} > \sum_{i=1}^n |\widetilde{\Theta}^{(i,1)}| \geq \frac{n}{2}|\Theta^2|$, which would lead to a contradiction. It is not difficult to see that for any $i \neq j \in [n]$, there is a one-to-one correspondence between $\widetilde{\Theta}^{(i,1)}$ and $\widetilde{\Theta}^{(j,1)}$. In particular, for any $\boldsymbol{c}_1$ in $\mathcal{C}(\widetilde{\Theta}^{(i,1)})$, there is a unique $\boldsymbol{c}_2$ in $\mathcal{C}(\widetilde{\Theta}^{(i,2)})$, with $\boldsymbol{c}_2(l) = \boldsymbol{c}_1(l)$ for any $l \in [n]\backslash\{i,j\}$, $\boldsymbol{c}_2(i) = \boldsymbol{c}_1(j)$, and $\boldsymbol{c}_2(j) = \boldsymbol{c}_1(i)$. Since i and j are arbitrarily selected, we have $|\widetilde{\Theta}^{(i,1)}| = |\widetilde{\Theta}^{(j,1)}|$ for any $i \neq j \in [n]$. Thus, $\frac{|\widetilde{\Theta}^{(i,1)}|}{|\Theta^2|} \geq \frac{1}{2}$

holds for any $i \in [n]$. Then, we have

$$
\begin{aligned}
&\inf_{\hat{\boldsymbol{c}}} \sup_{(\boldsymbol{P},\boldsymbol{c})\in\Theta} \mathbb{E}\ell(\boldsymbol{c},\hat{\boldsymbol{c}}) \geq \inf_{\hat{\boldsymbol{c}}} \sup_{(\boldsymbol{P},\boldsymbol{c})\in\Theta^2} \mathbb{E}\ell(\boldsymbol{c},\hat{\boldsymbol{c}}) \geq \inf_{\hat{\boldsymbol{c}}} \frac{1}{|\Theta^2|} \sum_{(\boldsymbol{P},\boldsymbol{c})\in\Theta^2} \mathbb{E}\ell(\boldsymbol{c},\hat{\boldsymbol{c}}) \\
=&\inf_{\hat{\boldsymbol{c}}} \frac{1}{|\Theta^2|} \sum_{(\boldsymbol{P},\boldsymbol{c})\in\Theta^2} \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}\big(\hat{\boldsymbol{c}}(i) \neq \boldsymbol{c}(i)\big) \geq \frac{1}{n|\Theta^2|} \sum_{i=1}^{n} \inf_{\hat{\boldsymbol{c}}(i)} \sum_{(\boldsymbol{P},\boldsymbol{c})\in\Theta^2} \mathbb{P}\big(\hat{\boldsymbol{c}}(i) \neq \boldsymbol{c}(i)\big) \\
\geq& \frac{1}{2n|\Theta^2|} \sum_{i=1}^{n} \sum_{(\boldsymbol{P},\boldsymbol{c})\in\widetilde{\Theta}^{(i,1)}} \inf_{\hat{\boldsymbol{c}}(i)} \Big[\mathbb{P}_{\boldsymbol{P},\boldsymbol{c}}\big(\hat{\boldsymbol{c}}(i) \neq \boldsymbol{c}(i)\big) + \mathbb{P}_{\boldsymbol{P},\psi^i(\boldsymbol{c})}\Big(\hat{\boldsymbol{c}}(i) \neq \psi^i(\boldsymbol{c})(i)\Big)\Big] \\
\geq& \exp\left(-\frac{1}{2}(1+\eta)nI_{t^*}\right),
\end{aligned}
\tag{S6.14}
$$

for some $\eta \to 0$. The fourth equation holds because $\widetilde{\Theta}^{(i,1)} \subset \Theta^2$, $\widetilde{\Theta}^{(i,2)} \subset \Theta^2$ and ψ^i is a one-to-one correspondence between $\widetilde{\Theta}^{(i,1)}$ and $\widetilde{\Theta}^{(i,2)}$. Combining $\frac{|\widetilde{\Theta}^{(i,1)}|}{|\Theta^2|} \geq \frac{1}{2}$, Lemma 5 and $nI_{t^*} \to \infty$, the last inequality holds. □

Assume that $\{X_i\}_{i=1}^{n_0}$ is a sequence of independent random variables. Let $\boldsymbol{X} = (X_1, \cdots, X_{n_0})^\top$, and then we consider the following hypothesis testing problem

$$
H_0 : \boldsymbol{X} \sim \bigotimes_{i=1}^{n_0} \text{Bern}\,(p)\,,\;\; H_1 : \boldsymbol{X} \sim \bigotimes_{i=1}^{n_0} \text{Bern}\,(q)\,,
\tag{S6.15}
$$

where $0 < q < p = (1+\delta)q < 1 - \epsilon_0$, $\delta > 0$, and $\epsilon_0 > 0$ is a small constant. Then the following lemma presents the minimum possible $Type\ I + II$ error of the above hypothesis testing.

Lemma 5. *Assume that as $n \to \infty$, $n_0 I_{t^*}(p,q) \to \infty$ and $p \asymp q$. Then, we* *have*

$$\inf_{\phi}\left(\mathbb{P}_{H_0}\phi+\mathbb{P}_{H_1}(1-\phi)\right)\geq \exp\Big(-(1+\eta)n_0 I_{t^*}(p,q)\Big), \qquad \text{(S6.16)}$$

for some $\eta\to 0$. *If* $n_0 I_{t^*}(p,q)\lesssim 1$, *then* $\inf_{\phi}\left(\mathbb{P}_{H_0}\phi+\mathbb{P}_{H_1}(1-\phi)\right)\geq C$ *for some constant* $C>0$. *When* $p\asymp q$ *is replaced by* $p\gg q$, *if additional conditions*

$$p\log^3\left(\frac{p}{q}\right)<\infty,\quad \lim_{n\to\infty}\frac{\log\frac{\log(\frac{p}{q})}{p}}{\log n}<1 \quad \textit{and} \quad \lim_{n\to\infty}\frac{\log\beta np}{\log\log\frac{p}{q}}>3$$

are added, (S6.16) *still holds.*

S.6.5 Proof of Theorem 3

We give some necessary lemmas, i.e. Lemmas 6 and 7. Lemma 6 is similar to Lemma 1 in Gao et al. (2017), discussing the distance between $\boldsymbol{P}^{0i}$ obtained in the initialization step and the real parameter $\boldsymbol{P}$. Lemmas 7 characterizes the loss of the refinement step in RACEn.

Lemma 6. *Suppose that as* $n\to\infty$, $\beta n\frac{(p-q)^2}{p}\to\infty$ *and Condition 1 holds with the requirement for* γ_n *replaced by* $\gamma_n=o(-\frac{\beta}{\log\beta})$. *Then there exists a constant* $C>0$ *such that*

$$\inf_{(\boldsymbol{P},\boldsymbol{c})\in\Theta_n(p,q,\beta)}\min_{i\in[n]}\mathbb{P}_{\boldsymbol{P},\boldsymbol{c}}\left(\max\left\{\left|\hat{p}^{0i}-p\right|,\left|\hat{q}^{0i}-q\right|\right\}\leq\eta(p-q)\right)\geq 1-Cn^{-(1+C_0)}, \qquad \text{(S6.17)}$$

where $\hat{p}^{0i}$ and $\hat{q}^{0i}$ are obtained in the refinement step of Algorithm 1, the constant $C_0 > 0$ is from Condition 1 and $\eta = C'(\eta_1+\eta_2)$ with $\eta_1 = \left(\frac{p}{\beta n(p-q)^2}\right)^{1/2}$ and $\eta_2 = -\gamma_n \frac{\log \beta}{\beta}$, for some large constant $C' > 0$.

The proof of Lemma 6 can be directly obtained from Lemma 3. Since $\tilde{\ell}\left(\boldsymbol{c}_{-i}, \boldsymbol{c}^0_{-i}\right) \leq \ell\left(\boldsymbol{c}_{-i}, \boldsymbol{c}^0_{-i}\right)$, by Lemma 3, we have

$$\begin{aligned}
&\inf_{(\boldsymbol{P},\boldsymbol{c})\in\Theta_n(p,q,\beta)} \min_{i\in[n]} \mathbb{P}_{\boldsymbol{P},\boldsymbol{c}}\left(\max\left\{\left|\hat{p}^{0i} - p\right|, \left|\hat{q}^{0i} - q\right|\right\} \leq \eta(p-q)\right) \\
&\geq 1 - \frac{1}{2}C(n-1)^{-(1+C_0)} \geq 1 - Cn^{-(1+C_0)},
\end{aligned}$$

for some constant $C > 0$. Because the result are independent of i, which means the proof is complete.

Lemma 7. *Suppose that as $n \to \infty$, $\beta n \frac{(p-q)^4}{p} \to \infty$, $\beta n I_{t^*} \to \infty$ and Condition 1 holds with $\gamma_n = o\left(-\frac{\beta(p-q)}{\log \beta}\right)$ when $p \asymp q$, and $\gamma_n = o\left(-\frac{\beta q}{\log \beta}\frac{\log\log\frac{p}{q}}{\log\frac{p}{q}}\right)$ when $p \gg q$. Then for $\{\check{\boldsymbol{c}}(i)\}_{i\in[n]}$ in Algorithm 1, there is a sequence $\eta \to 0$ and a constant $C > 0$, such that*

$$\sup_{(\boldsymbol{P},\boldsymbol{c})\in\Theta_n(p,q,\beta)} \max_{i\in[n]} \mathbb{P}_{\boldsymbol{P},\boldsymbol{c}}\big(\check{\boldsymbol{c}}(i) \neq \boldsymbol{c}(i)\big) \leq \exp\big(-(1-\eta)\beta n I_{t^*}\big) + Cn^{-(1+C_0)}, \tag{S6.18}$$

where the constant $C_0 > 0$ is from Condition 1.

Proof of Theorem 3. By Lemma 7, for any $(\boldsymbol{P}, \boldsymbol{c}) \in \Theta$ and $i \in [n]$, we have $\mathbb{P}\big(\check{\boldsymbol{c}}(i) \neq \boldsymbol{c}(i)\big) \leq \exp\big(-(1-\eta')\beta n I_{t^*}\big) + Cn^{-(1+C_0)}$, for constants $C, C_0 > 0$

and some $\eta' \to 0$. Set $\eta = \eta' + \frac{1}{\sqrt{\beta n I_{t^*}}}$, and by Markov's inequality, we have

$$\begin{aligned}
&\mathbb{P}\Big\{\ell(\boldsymbol{c}, \check{\boldsymbol{c}}) > \exp\big(-(1-\eta)\beta n I_{t^*}\big)\Big\} \\
&\le \frac{\frac{1}{n}\sum_{i=1}^{n}\mathbb{P}\big(\check{\boldsymbol{c}}(i) \neq \boldsymbol{c}(i)\big)}{\exp\big(-(1-\eta)\beta n I_{t^*}\big)} \\
&\le \exp\big((\eta' - \eta)\beta n I_{t^*}\big) + \frac{Cn^{-(1+C_0')}}{\exp\big(-(1-\eta)\beta n I_{t^*}\big)} \\
&= \exp(-\sqrt{\beta n I_{t^*}}) + \frac{Cn^{-(1+C_0')}}{\exp\big(-(1-\eta)\beta n I_{t^*}\big)}.
\end{aligned} \tag{S6.19}$$

If $\exp\big(-(1-\eta)\beta n I_{t^*}\big) > Cn^{-(1+C_0'/2)}$, then

$$\mathbb{P}\Big\{\ell(\boldsymbol{c}, \check{\boldsymbol{c}}) > \exp\big(-(1-\eta)\beta n I_{t^*}\big)\Big\} \le \exp\Big(-\sqrt{\beta n I_{t^*}}\Big) + Cn^{-C_0'/2} \to 0.$$

If $\exp\big(-(1-\eta)\beta n I_{t^*}\big) \le Cn^{-(1+C_0'/2)}$, then

$$\begin{aligned}
&\mathbb{P}\Big\{\ell(\boldsymbol{c}, \check{\boldsymbol{c}}) > \exp\big(-(1-\eta)\beta n I_{t^*}\big)\Big\} \\
&\le \mathbb{P}\big(\ell(\boldsymbol{c}, \check{\boldsymbol{c}}) > 0\big) = \mathbb{P}\left(\bigcup_{i=1}^{n}\big\{\check{\boldsymbol{c}}(i) \neq \boldsymbol{c}(i)\big\}\right) \\
&\le \sum_{i=1}^{n}\mathbb{P}\Big(\check{\boldsymbol{c}}(i) \neq \boldsymbol{c}(i)\Big) \\
&\le \sum_{i=1}^{n}\Big[\exp\big(-(1-\eta')\beta n I_{t^*}\big) + Cn^{-(1+C_0')}\Big] \\
&\le Cn^{-C_0'/2} + Cn^{-C_0'} \to 0.
\end{aligned} \tag{S6.20}$$

The last inequality in (S6.20) holds since $\eta \ge \eta'$, which implies $\exp\big(-(1-\eta')\beta n I_{t^*}\big) \le \exp\big(-(1-\eta)\beta n I_{t^*}\big) \le Cn^{-(1+C_0'/2)}$. Therefore, we can obtain that

$$\mathbb{P}\Big\{\ell(\boldsymbol{c}, \check{\boldsymbol{c}}) > \exp\big(-(1-\eta)\beta n I_{t^*}\big)\Big\} \to 0,$$

which completes the proof.

□

S.6.6 Proof of Theorem 5

Let's first take a look at the separability of a node with $\boldsymbol{c}(i) = 1$ and a node with $\boldsymbol{c}(i) = 2$ based on parameter matrix $(P_{\boldsymbol{c}(i)\boldsymbol{c}(j)})_{n\times n}$, which is shown in the following lemma.

Lemma 8. *For* $(\boldsymbol{P}, \boldsymbol{c}) \in \Theta_n(p, q, \beta)$, *set* $\boldsymbol{M}' = (P_{\boldsymbol{c}(i)\boldsymbol{c}(j)})_{n\times n}$, *whose rank is* 2. *Let* $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ *be the eigen-decomposition of* $\boldsymbol{M}'$.

Then, $\boldsymbol{\Lambda} = \mathrm{diag}\,(\lambda_1, \lambda_2)$ *with* $\lambda_1 = x + \sqrt{\tilde{x}^2 + y^2}$ *and* $\lambda_2 = x - \sqrt{\tilde{x}^2 + y^2}$, *where*

$$x = \frac{n_1(\boldsymbol{c})p + n_2(\boldsymbol{c})q}{2},\ \tilde{x} = \frac{n_1(\boldsymbol{c})p - n_2(\boldsymbol{c})q}{2},\ \textit{and}\ y = \sqrt{n_1(\boldsymbol{c})n_2(\boldsymbol{c})}q.$$

Besides, $\boldsymbol{c}(i) = \boldsymbol{c}(j)$ *if and only if* $U_{i1} = U_{j1}$ *and if* $\boldsymbol{c}(i) \neq \boldsymbol{c}(j)$, $(U_{i1} - U_{j1})^2 = \Delta^2$ *as long as* $\Delta > 0$, *where* U_{ij} *is the* (i, j)*-th element of matrix* $\boldsymbol{U}$, *and*

$$\Delta^2 = \frac{1}{(\tilde{x}+z)^2 + y^2}\left\{\frac{1}{\sqrt{n_1(\boldsymbol{c})}}(\tilde{x}+z) - \frac{1}{\sqrt{n_2(\boldsymbol{c})}}y\right\}^2, \tag{S6.21}$$

with $z = \sqrt{\tilde{x}^2 + y^2}$.

Proof of Theorem 5. Let $\boldsymbol{c}$ be the true label. $\boldsymbol{A}^\tau$, $\hat{\boldsymbol{U}}^{K-1} = \hat{\boldsymbol{U}}^1 = (\hat{U}^1) \in \mathbb{R}^n$, $\tilde{\boldsymbol{c}}^0_{\mathsf{esc}}$, $\boldsymbol{c}^0_{\mathsf{esc}}$ and $\{\tilde{\nu}_k\}_{k\in[2]}$ with $\tilde{\nu}_k \in \mathbb{R}$ come from Algorithm 2, where

$\{\tilde{\nu}_k\}_{k\in[2]}$ are taken so that the left side of the inequality of (2.5) reaches the minimum. Let $\boldsymbol{M}' = (P_{\mathbf{c}(i)\mathbf{c}(j)})_{n\times n}$, whose rank is 2. Let $\boldsymbol{U\Lambda U}^\top$ be the eigen-decomposition of $\boldsymbol{M}'$, and for each i in $[n]$, let $\bar{U}_{i1} = \vartheta U_{i1}$ where $\vartheta \in \{-1, 1\}$, which will be determined later.

For each $k \in [2]$, define $\tilde{T}_k = \{i : \boldsymbol{c}(i) = k, |\tilde{V}_i - \bar{U}_{i1}| < \Delta/2\}$ where $\tilde{V}_i = \tilde{\nu}_{\tilde{\boldsymbol{c}}^0_{\mathsf{esc}}(i)}$ and Δ is defined in (S6.21). Note that $|\bar{U}_{i1} - \bar{U}_{j1}| = \Delta$ for all i, j with $\boldsymbol{c}(i) \neq \boldsymbol{c}(j)$.

Finally, we will prove that $n\tilde{\ell}\left(\boldsymbol{c}, \tilde{\boldsymbol{c}}^0_{\mathsf{esc}}\right) \leq \sum_{k=1}^2 |\tilde{T}_k^c|$, so let's first the upper bound of $\sum_{k=1}^2 |\tilde{T}_k^c|$. By the definition of $\tilde{T}_k$ we have

$$\begin{aligned}
\frac{1}{4}\Delta^2 \sum_{k=1}^2 |\tilde{T}_k^c| &\leq \sum_{i=1}^n (\tilde{V}_i - \bar{U}_{i1})^2 \\
&\leq 2\sum_{i=1}^n (\tilde{V}_i - \hat{U}_i^1)^2 + 2\sum_{i=1}^n (\hat{U}_i^1 - \bar{U}_{i1})^2 \\
&\leq 2(2+\xi)\sum_{i=1}^n (\hat{U}_i^1 - \bar{U}_{i1})^2, \qquad \text{(S6.22)}
\end{aligned}$$

Next, the next task is to present the bound of $\sum_{i=1}^n (\hat{U}_i^1 - \bar{U}_{i1})^2$, which can be obtained through the Davis-Kahan theorem. The version of Davis-Kahan theorem used here is Theorem 4.5.5 in Vershynin (2018). Specifically, let $\boldsymbol{S}$ and $\boldsymbol{T}$ be symmetric matrices with the same dimensions. Fix i and assume that the i-th largest eigenvalue of $\boldsymbol{S}$ is well separated from the rest of the spectrum:

$$\min_{j:j\neq i} |\lambda_i(\boldsymbol{S}) - \lambda_j(\boldsymbol{S})| = \tilde{\delta} > 0,$$

Then the unit eigenvectors $\boldsymbol{v}_i(\boldsymbol{S})$ and $\boldsymbol{v}_i(\boldsymbol{T})$ are close to each other up to a sign, namely

$$\exists \tilde{\vartheta} \in \{-1,1\}: \quad \left\| \boldsymbol{v}_i(\boldsymbol{S}) - \tilde{\vartheta}\boldsymbol{v}_i(\boldsymbol{T}) \right\|_2 \leq \frac{2^{3/2}\|\boldsymbol{S}-\boldsymbol{T}\|_{\mathrm{op}}}{\tilde{\delta}}.$$

Hence, setting $\boldsymbol{S} = \boldsymbol{A}^{\tau}$, $\boldsymbol{T} = \boldsymbol{M}'$ and $\vartheta = \arg\min_{\tilde{\vartheta}\in\{-1,1\}} \|v_i(\boldsymbol{S}) - \tilde{\vartheta} v_i(\boldsymbol{T})\|_2$, we have $\tilde{\delta} = |\lambda_1 - \lambda_2|$ where λ_1 and λ_2 are eigenvalues of matrix $\boldsymbol{M}'$, whose specific forms are given in Lemma 8. Then, we have

$$\sum_{i=1}^{n} (\hat{U}_i^1 - \bar{U}_{i1})^2 \leq \frac{2^3\|\boldsymbol{A}^{\tau} - \boldsymbol{M}'\|_{\mathrm{op}}^2}{(\lambda_1-\lambda_2)^2}. \tag{S6.23}$$

By Lemma 5 in Gao et al. (2017), for any $C' > 0$, there exists some $C > 0$, such that $\|\boldsymbol{A}^{\tau} - \boldsymbol{M}'\|_{\mathrm{op}} \leq C\sqrt{np+1}$ with probability at least $1 - n^{-C'}$ uniformly over $\tau \in [C_1(np+1), C_2(np+1)]$ for some sufficiently large constants C_1, C_2.

To sum up, by (S6.22) and (S6.23), we have

$$\sum_{k=1}^{2} |\tilde{T}_k^c| \quad \leq \quad C(1+\xi)\frac{np+1}{\Delta^2(\lambda_1-\lambda_2)^2}, \tag{S6.24}$$

with probability at least $1 - n^{-(1+C')}$.

Next, we will prove that $n\tilde{\ell}\left(\boldsymbol{c}, \tilde{\boldsymbol{c}}^0_{\mathsf{esc}}\right) \leq \sum_{k=1}^2 |\tilde{T}_k^c|$. By the assumption that $-\frac{\beta\Delta^2(\lambda_1-\lambda_2)^2}{p\log\beta} \to \infty$ as $n \to \infty$, we have $|\tilde{T}_k^c| \leq C(1+\xi)\frac{np+1}{\Delta^2(\lambda_1-\lambda_2)^2} = o(\beta n)$ for each $k \in [2]$ with probability at least $1 - n^{-(1+C')}$, which means based on the event (S6.24), $\tilde{T}_k \neq \emptyset$ for each $k \in [2]$. In such case, $\tilde{T}_1$ and $\tilde{T}_2$ have the following properties. First, for any $i \in \tilde{T}_1$ and $j \in \tilde{T}_2$, $\tilde{\boldsymbol{c}}^0_{\mathsf{esc}}(i) \neq$

$\tilde{\boldsymbol{c}}^0_{\mathsf{esc}}(j)$. Otherwise, if $\tilde{\boldsymbol{c}}^0_{\mathsf{esc}}(i) = \tilde{\boldsymbol{c}}^0_{\mathsf{esc}}(j)$, we have $\|\bar{U}_{i1} - \bar{U}_{j1}\|_2 \leq \|\tilde{V}_i - \bar{U}_{i1}\|_2 + \|\tilde{V}_j - \bar{U}_{j1}\|_2 < \Delta$, which contradicts with the fact that $\|\bar{U}_{i1} - \bar{U}_{j1}\|_2 = \Delta$ for all i, j with $\boldsymbol{c}(i) \neq \boldsymbol{c}(j)$. Second, fix any $k \in [2]$, and for any $i, j \in \tilde{T}_k$, $\tilde{\boldsymbol{c}}^0_{\mathsf{esc}}(i) = \tilde{\boldsymbol{c}}^0_{\mathsf{esc}}(j)$, since there are at most two values for $\tilde{\boldsymbol{c}}^0_{\mathsf{esc}}(u)$ for each $u \in [n]$.

Combining the above two properties of $\{\tilde{T}_k\}_{k\in[2]}$, we have $n\tilde{\ell}\left(\boldsymbol{c}, \tilde{\boldsymbol{c}}^0_{\mathsf{esc}}\right) \leq \sum_{k=1}^{2} |\tilde{T}_k^c|$. Then, we have that for any $(\boldsymbol{P}, \boldsymbol{c}) \in \Theta_n(p, q, \beta)$,

$$\mathbb{P}_{P,\boldsymbol{c}}\left\{n\tilde{\ell}\left(\boldsymbol{c}, \tilde{\boldsymbol{c}}^0_{\mathsf{esc}}\right) \leq C(1+\xi)\frac{np+1}{\Delta^2(\lambda_1 - \lambda_2)^2}\right\} \geq 1 - n^{-(1+C')},$$

for constants $C, C' > 0$. Based on assumption $\frac{-\beta}{\log\beta}\frac{\Delta^2(\lambda_1-\lambda_2)^2}{p} \to \infty$, we have

$$\mathbb{P}_{P,\boldsymbol{c}}\left\{\tilde{\ell}\left(\boldsymbol{c}, \tilde{\boldsymbol{c}}^0_{\mathsf{esc}}\right) \leq \gamma_n\right\} \geq 1 - n^{-(1+C')},$$

for some constant $C' > 0$ with $\gamma_n = o(-\beta/\log\beta)$. And by Lemma 3, we have

$$\mathbb{P}_{P,\boldsymbol{c}}\left\{n\ell\left(\boldsymbol{c}, \boldsymbol{c}^0_{\mathsf{esc}}\right) \leq C(1+\xi)\frac{np+1}{\Delta^2(\lambda_1 - \lambda_2)^2}\right\} \geq 1 - 3n^{-(1+C')}.$$

for constants $C, C' > 0$ and since $p \gg 1/n$, the proof is completed. □

S.6.7 Proof of Corollary 2

Proof of Corollary S1. Firstly, we give the lower bound of $(\lambda_1 - \lambda_2)^2$. For simplicity, we abbreviate $n_k(\boldsymbol{c})$ as n_k for each $k \in [2]$. By assumption

$\lim_{n\to\infty} \frac{p\beta}{q(1-\beta)} > 2$, there is a small constant $\epsilon > 0$ such that $\lim_{n\to\infty} \frac{p\beta}{q(1-\beta)} > 2 + \epsilon$. Then,

$$\begin{aligned}
(\lambda_1 - \lambda_2)^2 &= (n_1 p - n_2 q)^2 + 4 n_1 n_2 q^2 \\
&= n_1 n_2 \left\{ \left(\sqrt{\frac{n_1}{n_2}} p - \sqrt{\frac{n_2}{n_1}} q \right)^2 + 4q^2 \right\} \\
&\geq n_1 n_2 \left\{ (1+\epsilon)^2 \frac{1-\beta}{\beta} q^2 + 4q^2 \right\} \\
&\geq \beta(1-\beta) n^2 q^2 \left\{ (1+\epsilon)^2 \frac{1-\beta}{\beta} + 4 \right\} \\
&\geq (1-\beta)^2 n^2 q^2 (1+\epsilon)^2 \gtrsim n^2 q^2. \qquad \text{(S6.25)}
\end{aligned}$$

The first inequality holds by the fact that $\sqrt{\frac{n_1}{n_2}} p - \sqrt{\frac{n_2}{n_1}} q \geq (1+\epsilon)(1-\beta)q/\beta$ when $\lim_{n\to\infty} \frac{p\beta}{q(1-\beta)} > 2 + \epsilon$. Next, we calculate the upper bound of $(\tilde{x} + z)^2 + y^2$ and the lower bound of $\left\{ \frac{1}{\sqrt{n_1}} (\tilde{x} + z) - \frac{1}{\sqrt{n_2}} y \right\}^2$ respectively, and then give the lower bound of Δ^2.

$$\begin{aligned}
& (\tilde{x} + z)^2 + y^2 \\
= \;& \left\{ \frac{1}{2}(n_1 p - n_2 q) + \left(\frac{1}{2^2}(n_1 p - n_2 q)^2 + n_1 n_2 q^2 \right)^{\frac{1}{2}} \right\}^2 + n_1 n_2 q^2 \\
\leq \;& 2 \left\{ \frac{1}{2}(n_1 p - n_2 q) + \left(\frac{1}{2^2}(n_1 p - n_2 q)^2 + n_1 n_2 q^2 \right)^{\frac{1}{2}} \right\}^2 \\
\leq \;& 2 n_1 n_2 \left\{ \frac{1}{2} \left(\frac{1-\beta}{\beta} \right)^{\frac{1}{2}} p + \left(\frac{1}{4} \frac{1-\beta}{\beta} p^2 + q^2 \right)^{\frac{1}{2}} \right\}^2 \\
\leq \;& 2 n_1 n_2 p^2 \left\{ 4 + \frac{1-\beta}{\beta} \right\}. \qquad \text{(S6.26)}
\end{aligned}$$

The second inequality holds since $(n_1/n_2)^{1/2}\, p - (n_2/n_1)^{1/2}\, q \le \Big\{ \left(\frac{1-\beta}{\beta}\right)^{1/2} p - \left(\frac{\beta}{1-\beta}\right)^{1/2} q \Big\}/2 \le \left(\frac{1-\beta}{\beta}\right)^{1/2} p/2$. Besides,

$$\begin{aligned}
&\left\{\frac{1}{\sqrt{n_1}}(\tilde{x}+z) - \frac{1}{\sqrt{n_2}}y\right\}^2 \\
&= \Bigg[\frac{1}{\sqrt{n_1}}\left\{\frac{1}{2}(n_1p - n_2q) + \left(\frac{1}{2^2}(n_1p-n_2q)^2 + n_1n_2q^2\right)^{\frac{1}{2}}\right\} \\
&\quad -\frac{1}{\sqrt{n_2}}\sqrt{n_1n_2}q\Bigg]^2 \\
&\ge \left[\sqrt{n_2}\left\{\frac{1}{2}(1+\epsilon)\left(\frac{1-\beta}{\beta}\right)^{\frac{1}{2}} + \left(\frac{1}{2^2}(1+\epsilon)^2\frac{1-\beta}{\beta}+1\right)^{\frac{1}{2}}\right\} - \sqrt{n_1}\right]^2 q^2 \\
&\ge \left\{\sqrt{n_2}(1+\epsilon)\left(\frac{1-\beta}{\beta}\right)^{\frac{1}{2}} - \sqrt{n_1}\right\}^2 q^2. \qquad \text{(S6.27)}
\end{aligned}$$

Combining (S6.26) and (S6.27), we can obtain that,

$$\begin{aligned}
\Delta^2 \;&\ge\; \frac{1}{2n_1n_2p^2\left\{4+\frac{1-\beta}{\beta}\right\}}\left\{\sqrt{n_2}(1+\epsilon)\left(\frac{1-\beta}{\beta}\right)^{\frac{1}{2}} - \sqrt{n_1}\right\}^2 q^2 \\
&\ge\; \frac{1}{2p^2\frac{1+3\beta}{\beta}}\frac{\epsilon^2q^2}{\beta n} = \frac{\epsilon^2q^2}{2(1+3\beta)np^2} \gtrsim \frac{q^2}{np^2}. \qquad \text{(S6.28)}
\end{aligned}$$

Then, combining (S6.25) and (S6.28), and the results are independent of $(\boldsymbol{P}, \boldsymbol{c})$, our proof is complete.

□

Proof of Lemma 1. Let $\frac{\mathrm{d}I_t(p,q)}{\mathrm{d}t} = 0$, i.e.

$$-\frac{(\frac{q}{p})^t p\log\frac{q}{p} + (\frac{1-q}{1-p})^t(1-p)\log\frac{1-q}{1-p}}{q^tp^{1-t}+(1-q)^t(1-p)^{1-t}} = 0,$$

and then, we can get $\frac{\mathrm{d}I_t(p,q)}{\mathrm{d}t}\Big|_{t=t^*} = 0$. Besides, for any $t > 0$, the second derivative of $I_t(p,q)$ is as follows

$$-\frac{(\frac{q}{p})^t p(\log\frac{q}{p})^2 + (\frac{1-q}{1-p})^t(1-p)(\log\frac{1-q}{1-p})^2}{(q^t p^{1-t} + (1-q)^t(1-p)^{1-t})^2},$$

which is strictly less than 0. Thus, we have that t^* is the only maximum point of $I_t(p,q)$ on $[0,+\infty)$. Combined with $I_0(p,q) = I_1(p,q)$, we have $t^* \in (0,1)$. In addition, $t'^* = 1 - t^*$ is the only maximum point of $I_t(q,p)$ on $(0,1)$, since for any $t > 0$, the second derivative of $I_t(q,p)$ is strictly less than 0, $I_0(q,p) = I_1(q,p)$ and $I_t(q,p) = I_{1-t}(p,q)$ for any $t \in [0,1]$. Thus, we get (1).

Next, we investigate the properties of t^* in (2). Firstly, considering the situation where $\delta \asymp 1$, in retrospect, t^* has the following form

$$t^* = \frac{\log\left[\frac{(p-1)\left(\log(1-p)-\log(1-q)\right)}{p(\log p - \log q)}\right]}{\log\left(\frac{q(1-p)}{p(1-q)}\right)} = \frac{\log\left[\frac{(1+\delta)q-1}{(1+\delta)q}\frac{\log\frac{1-(1+\delta)q}{1-q}}{\log(1+\delta)}\right]}{\log\left(\frac{1-(1+\delta)q}{(1-q)(1+\delta)}\right)}.$$

Then, we rewrite t^* as

$$t^* = 1 - \frac{\log\frac{1-q}{q} + \log\Big(-\log\big(1-\frac{\delta q}{1-q}\big)\Big) - \log\log(1+\delta)}{\log(1+\delta) - \log\big(1-\frac{\delta q}{1-q}\big)} \tag{S6.29}$$

$$= 1 - \frac{\log\frac{1-q}{\delta q} + \log\Big(-\log\big(1-\frac{\delta q}{1-q}\big)\Big) - \log\frac{\log(1+\delta)}{\delta}}{\log(1+\delta) - \log\big(1-\frac{\delta q}{1-q}\big)}. \tag{S6.30}$$

From (S6.30), we can see that when δ or δq tends to 0, we can see t^* more clearly. First, we analyze the case when $\delta q \gtrsim 1$. We will first present that t^* stays away from 0 and 1 by using reduction to absurdity.

From (S6.29), we know that the statement that there is a certain distance between t^* and 1 is equivalent to the statement that there is a certain distance between $-\frac{1-q}{q}\frac{\log\left(1-\frac{\delta q}{1-q}\right)}{\log(1+\delta)}$ and 1. Let $y = \frac{q}{1-q}$ and assume $-\frac{1}{y}\frac{\log(1-\delta y)}{\log(1+\delta)} = 1$. Note that $\delta y < 1$ since $p < 1 - \epsilon_0$ for some small constant $\epsilon_0 \in (0, 1)$. Then, due to $\log(1 - x) < -x$ for any $x < 1$, we have $\delta < \log(1 + \delta)$, and there is a contradiction.

Similarly, to illustrate that t^* is away from 0, assume

$$-\frac{1}{y}\frac{\log(1-\delta y)}{\log(1+\delta)} = \frac{1+\delta}{1-\delta y}.$$

Due to $\log(1-x) > -\frac{x}{1-x}$ for any $x \in (0, 1)$, we have $\delta > (1+\delta)\log(1+\delta)$. And there is a contradiction, since $f(\delta) = \delta - (1 + \delta)\log(1 + \delta)$ decreases monotonically on $(0, +\infty)$.

Next, we consider the case of $\delta q = o(1)$. When $\delta = o(1)$, from (S6.29) and by the Taylor expansion of $\log(1 + x)$ at $x = 0$, we can get

$$t^* = 1 - \frac{\log\left[1 + \frac{1}{2}\frac{\delta q}{1-q} + \frac{1}{3}\left(\frac{\delta q}{1-q}\right)^2 + o(\delta^2 q^2)\right] - \log\left[1 - \frac{1}{2}\delta + \frac{1}{3}\delta^2 + o(\delta^2)\right]}{\log(1+\delta) - \log\left(1 - \frac{\delta q}{1-q}\right)}.$$

Using the Taylor expansion of $\log(1 + x)$ at $x = 0$ again, we have

$$t^* = 1 - \frac{1}{2}\frac{\frac{q}{1-q}\left(1 + \frac{5}{12}\frac{\delta q}{1-q}\right) + 1 - \frac{5}{12}\delta + o(\delta)}{\frac{q}{1-q}\left(1 + \frac{1}{2}\frac{\delta q}{1-q}\right) + 1 - \frac{1}{2}\delta + o(\delta)}.$$

Simplifying the numerator, we can get

$$t^* = \frac{1}{2} - \delta\frac{1-2q}{24(1-q)} + o(\delta).$$

The rest case in $\delta \asymp 1$, is $q = o(1)$. From (S6.29) and by the Taylor expansion of $\log(1+x)$ at $x = 0$, and through simple calculations, we have

$$t^* = \big(1 + o(1)\big)\frac{\log(1+\delta) - \log\delta + \log\log(1+\delta)}{\log(1+\delta)}.$$

Thus, due to the fact that for $\delta > 0$,

$$\log(1+\delta) > \log\delta - \log\log(1+\delta) > \frac{1}{2}\log(1+\delta),$$

we have $t^* \in [\epsilon, \frac{1}{2} - \epsilon]$ for a constant $\epsilon > 0$.

Finally, we consider the case where $\delta \to \infty$. We rewrite t^* as

$$t^* = \frac{-\log\log(1+\delta) + \log\Big(-\frac{1-p}{p}\log\frac{1-p}{1-q}\Big)}{-\log(1+\delta) + \log\frac{1-p}{1-q}}.$$

Since $p < 1 - \epsilon_0$ for some small constant $\epsilon_0 \in (0,1)$, when $p \gg q$, we have $-\log\frac{1-p}{1-q} \asymp p$ and $\log(1-p) \asymp -p$, which means the dominant terms of numerator and denominator are $-\log\log(1+\delta)$ and $-\log(1+\delta)$. Hence,

$$t^* = \big(1 + o(1)\big)\frac{\log\log(1+\delta)}{\log(1+\delta)},$$

when $\delta \to \infty$.

□

Proof of Lemma 2. For any $t, p, q \in (0,1)$, define $J_t(p,q) = 2\Big(tp + (1-t)q - p^t q^{1-t}\Big)$, which is nonnegative. Then,

$$I_t(p,q) \;=\; -\log\Big(q^t p^{1-t} + (1-q)^t(1-p)^{1-t}\Big)$$

$$= \quad -\log\left(1 - \frac{1}{2}J_t(q,p) - \frac{1}{2}J_t(1-q,1-p)\right).$$

Due to Lemma 11 in Gao et al. (2018), we have $2\min(t^*,1-t^*)(\sqrt{p}-\sqrt{q})^2 \leq J_{t^*}(q,p) \leq 2(\sqrt{p}-\sqrt{q})^2$ and $J_{t^*}(1-q,1-p) \lesssim (p-q)^2$ since $1-p > \epsilon_0$ for some small constant $\epsilon_0 \in (0,1)$.

The first task is to give the proof of (1). From Lemma 1, when $\delta \lesssim 1$, t^* stays away from 0 and 1, hence $J_{t^*}(q,p) \asymp \delta^2 q$ and $J_{t^*}(1-q,1-p) \lesssim \delta^2 q^2$. First, we consider the case of $\delta^2 q = o(1)$. When $\delta^2 q = o(1)$, $J_{t^*}(q,p)$ and $J_{t^*}(1-q,1-p)$ both tend to 0. Hence,

$$\begin{aligned} I_{t^*}(p,q) &= -\log\left(1 - \frac{1}{2}J_{t^*}(q,p) - \frac{1}{2}J_{t^*}(1-q,1-p)\right) \\ &= \frac{1}{2}\big(1+o(1)\big)\Big(J_{t^*}(q,p) + J_{t^*}(1-q,1-p)\Big), \end{aligned}$$

which indicates that $I_{t^*} \asymp \delta^2 q$.

Next, we consider the case of $\delta^2 q \asymp 1$. Since $\delta \lesssim 1$ and $q < 1$, when $\delta^2 q \asymp 1$, we have $\delta, q \asymp 1$. And then, we have $I_{1/2}(p,q)$ stays away from 0. In addition, from the proof of Lemma 1, we have for any $t \in [0,1]$, the second derivative of $I_t(p,q)$ is strictly less than 0, and $I_0(p,q) = I_1(p,q) = 0$. Besides, from Lemma 1, when $\delta \asymp 1$, t^* stays away from 0 and 1. And based on the above information, we have $I_{t^*}(p,q) \asymp 1$. And the proof of (1) is completed.

The following task is to give the proof of (2). Let's first give the lower bound of $I_{t^*}(p,q)$. From the proof of (1),

$$\begin{aligned} I_{t^*}(p,q) &\geq \frac{1}{2}J_{t^*}(q,p)+\frac{1}{2}J_{t^*}(1-q,1-p) \\ &\geq t^*\left\{(\sqrt{p}-\sqrt{q})^2+\left(\sqrt{1-p}-\sqrt{1-q}\right)^2\right\} \\ &\gtrsim t^*p, \end{aligned}$$

where the second inequality holds since $2\min(t^*,1-t^*)(\sqrt{p}-\sqrt{q})^2 \leq J_{t^*}(q,p) \leq 2(\sqrt{p}-\sqrt{q})^2$ and the last inequality holds since when $p \gg q$, $t^* \to 0$, and $1-p > \epsilon_0$ for some small constant $\epsilon_0 \in (0,1)$.

Next, we give the upper bound of $I_{t^*}(p,q)$. When $p=o(1)$, we have

$$\begin{aligned} I_{t^*}(p,q) &= \frac{1}{2}\big(1+o(1)\big)\Big(J_{t^*}(q,p)+J_{t^*}(1-q,1-p)\Big) \\ &\leq \big(1+o(1)\big)\left\{(\sqrt{p}-\sqrt{q})^2+\left(\sqrt{1-p}-\sqrt{1-q}\right)^2\right\} \\ &\lesssim p. \end{aligned}$$

When $p \gtrsim 1$, by the definition of $I_t(p,q)$, we have

$$\begin{aligned} I_{t^*}(p,q) &= -\log\left(q^{t^*}p^{1-t^*}+(1-q)^{t^*}(1-p)^{1-t^*}\right) \\ &= -\log\left\{p\left(\frac{q}{p}\right)^{t^*}+(1-p)\left(\frac{1-q}{1-p}\right)^{t^*}\right\} \\ &= -\log\left\{\big(1+o(1)\big)(1-p)\right\} \\ &\lesssim p. \end{aligned}$$

When $p \gg q$, $\left(\frac{q}{p}\right)^{t^*} \to 0$ and $\left(\frac{1-q}{1-p}\right)^{t^*} \to 1$ since when $p \gg q$, $t^* \to 0$, and $1-p > \epsilon_0$ for some small constant $\epsilon_0 \in (0,1)$, and thus the third equality holds.

□

Proof of Lemma 3. For any $(\boldsymbol{P}, \boldsymbol{c}) \in \Theta_n(p,q,\beta)$, define event $E = \{\tilde{\ell}\left(\boldsymbol{c}, \tilde{\boldsymbol{c}}^0\right) \leq \gamma_n\}$. Because $\gamma_n = o(-\frac{\beta}{\log \beta})$, there is a unique $\pi : [2] \to [2]$ that makes $\tilde{\ell}\left(\boldsymbol{c}, \tilde{\boldsymbol{c}}^0\right) = \ell\left(\pi[\boldsymbol{c}], \tilde{\boldsymbol{c}}^0\right)$, where $\pi[\boldsymbol{c}] = \left(\pi(\boldsymbol{c}(1)), \cdots, \pi(\boldsymbol{c}(n))\right)^\top$. Without loss of generality, assume that $\pi(1) = 1$ and $\pi(2) = 2$. In addition, for any $k, l \in [2]$, let $\mathcal{C}_k = \{i : \boldsymbol{c}(i) = k\}$, $\widetilde{\mathcal{C}}_k = \{i : \tilde{\boldsymbol{c}}^0(i) = k\}$, $n_k = |\mathcal{C}_k|$, $\widetilde{n}_k = \left|\widetilde{\mathcal{C}}_k\right|$, $\mathcal{E}_{kl} = \left\{(i,j) : i < j, i \in \mathcal{C}_k, j \in \mathcal{C}_l, A_{ij} = 1\right\}$ and $\widetilde{\mathcal{E}}_{kl} = \left\{(i,j) : i < j, i \in \widetilde{\mathcal{C}}_k, j \in \widetilde{\mathcal{C}}_l, A_{ij} = 1\right\}$. Fix any $k \in [2]$. On E,

$$\exists \gamma_1, \gamma_2 \geq 0, \gamma_1 + \gamma_2 \leq \gamma_n, \quad \text{s.t.} \quad n_1 \geq \left|\widetilde{\mathcal{C}}_1 \cap \mathcal{C}_1\right| \geq n_1 - \gamma_1 n$$
$$\text{and} \quad n_2 \geq \left|\widetilde{\mathcal{C}}_2 \cap \mathcal{C}_2\right| \geq n_2 - \gamma_2 n. \tag{S6.31}$$

Let $\mathcal{C}'_k$ be any deterministic subset of $[n]$ such that (S6.31) holds with $\widetilde{\mathcal{C}}_k$ replaced by $\mathcal{C}'_k$, i.e. $\exists\ \gamma_1(\mathcal{C}'_k), \gamma_2(\mathcal{C}'_k) \geq 0, \gamma_1(\mathcal{C}'_k) + \gamma_2(\mathcal{C}'_k) \leq \gamma_n$, such that $n_1 \geq \left|\widetilde{\mathcal{C}}_1 \cap \mathcal{C}_1\right| \geq n_1 - \gamma_1 n$ and $n_2 \geq \left|\widetilde{\mathcal{C}}_2 \cap \mathcal{C}_2\right| \geq n_2 - \gamma_2 n$. Then, there are at most

$$\begin{aligned}
\sum_{l=0}^{\gamma_n n} \begin{pmatrix} n_1 \\ l \end{pmatrix} \sum_{m=0}^{\gamma_n n - l} \begin{pmatrix} n_2 \\ m \end{pmatrix} &\leq (\gamma_n n + 1)^2 \left(\frac{e n_1}{\gamma_n n}\right)^{\gamma_n n} \left(\frac{e n_2}{\gamma_n n}\right)^{\gamma_n n} \\
&\leq \exp\left\{2 \log(\gamma_n n + 1) + 2\gamma_n n \log \frac{e}{\gamma_n}\right\} \\
&\leq \exp\left\{C_1 \gamma_n n \log \frac{1}{\gamma_n}\right\}
\end{aligned}$$

different subsets satisfying this property, where C_1 is a positive constant. The first inequality holds since $\gamma_n n = o(\beta n)$. For any $k, l \in [2]$, let $\mathcal{E}'_{kl} = \{(i,j) : i < j, i \in \mathcal{C}'_k, j \in \mathcal{C}'_l, A_{ij} = 1\}$, and $n'_k = |\mathcal{C}'_k|$. By definition of $\mathcal{C}'_k$, we can get $n_k - \gamma_k(\mathcal{C}'_k)n \le n'_k \le n_k + \gamma_k(\mathcal{C}'_k)n$ for each $k \in [2]$.

Hence, we can get the fact that $|\mathcal{E}'_{11}|$ consists of $n'_1(n'_1-1)/2$ independent Bernoulli random variables, where at most $\gamma_n n(n_1 + \gamma_n n)$ of them follow Bern (q), Due to the property of $\mathcal{E}'_{11}$, we have

$$\frac{\gamma_n n(n_1 + \gamma_n n)q + \{\frac{1}{2}n'_1(n'_1 - 1) - \gamma_n n(n_1 + \gamma_n n)\}p}{\frac{1}{2}n'_1(n'_1 - 1)} \le \mathbb{E}\frac{|\mathcal{E}'_{11}|}{\frac{1}{2}n'_1(n'_1 - 1)} \le p.$$

Under the assumption $\gamma_n = \eta_2(-\frac{\beta}{\log \beta})$, we have

$$\left|\mathbb{E}\frac{|\mathcal{E}'_{11}|}{\frac{1}{2}n'_1(n'_1 - 1)} - P_{11}\right| \le C\frac{\gamma_n}{\beta}(p - q) = -C\frac{\eta_2}{\log \beta}(p - q), \qquad \text{(S6.32)}$$

for a constant $C > 0$. Similarly, we can obtain that $|\mathcal{E}'_{22}|$ consists of $n'_2(n'_2 - 1)/2$ independent Bernoulli random variables, where at most $\gamma_n n(\gamma_n n - 1)/2$ of them follow Bern (p), Due to the property of $\mathcal{E}'_{22}$, we have

$$q \le \mathbb{E}\frac{|\mathcal{E}'_{11}|}{\frac{1}{2}n'_1(n'_1 - 1)} \le \frac{\frac{1}{2}\gamma_n n(\gamma_n n - 1)p + \{\frac{1}{2}n'_2(n'_2 - 1) - \frac{1}{2}\gamma_n n(\gamma_n n - 1)\}q}{\frac{1}{2}n'_2(n'_2 - 1)}.$$

Under the assumption $\gamma_n = \eta_2(-\frac{\beta}{\log \beta})$, we have

$$\left|\mathbb{E}\frac{|\mathcal{E}'_{22}|}{\frac{1}{2}n'_1(n'_1 - 1)} - P_{22}\right| \le C\frac{\gamma_n^2}{\beta^2}(p - q) = -C\frac{\eta_2^2}{\log^2 \beta}(p - q), \qquad \text{(S6.33)}$$

for a constant $C > 0$. Besides, since $|\mathcal{E}'_{kk}| = \sum_{u<v\in\mathcal{C}'_k} A_{uv}$, we can get

$$\mathrm{Var}(|\mathcal{E}'_{kk}|) = \sum_{u<v\in\mathcal{C}'_k} \mathrm{Var}\, A_{uv} \quad = \sum_{u<v\in\mathcal{C}'_k} P_{\boldsymbol{c}(u)\boldsymbol{c}(v)}(1 - P_{\boldsymbol{c}(u)\boldsymbol{c}(v)})$$

$$\leq \frac{1}{2}n'_k(n'_k-1)p \leq \frac{1}{2}(n_k+\gamma_n n)^2 p.$$

By Bernstein's inequality, for any $t > 0$,

$$\mathbb{P}\left\{\left||\mathcal{E}'_{kk}| - \mathbb{E}|\mathcal{E}'_{kk}|\right| > t\right\} \leq 2\exp\left\{-\frac{t^2}{2\left(\frac{1}{2}\left(n_k+\gamma_n n\right)^2 p + \frac{1}{3}t\right)}\right\}.$$

Let

$$\begin{aligned} t^2 &= 2\left(n_k+\gamma_n n\right)^2 p\Big(C_1\gamma_n n\log\gamma_n^{-1} + (1+C_0)\log n\Big) \\ &\quad \vee \Big(2C_1\gamma_n n\log\gamma_n^{-1} + 2(1+C_0)\log n\Big)^2 \\ &\lesssim \Big(n_k\left(np\gamma_n\log\gamma_n^{-1}\right)^{\frac{1}{2}} + \gamma_n n\log\gamma_n^{-1}\Big)^2, \end{aligned}$$

where the second equality holds since $\gamma_n\log\gamma_n^{-1} \geq \frac{1}{n}\log n$ for any $\frac{1}{\gamma_n} \leq n$ because $\frac{\log x}{x}$ decreases monotonically on $[e,+\infty)$. And $\gamma_n < \frac{1}{n}$ means the initialization is already perfect. When $\gamma_n < \frac{1}{n}$, we can still continue to the following arguments by replacing every γ_n with $\frac{1}{n}$ and all the steps still hold. Then, there exists a constant $C(C_0,C_1) > 0$ that depends only on C_0 and C_1, such that

$$\begin{aligned} &\mathbb{P}\left\{\left||\mathcal{E}'_{kk}| - \mathbb{E}|\mathcal{E}'_{kk}|\right| > C(C_0,C_1)\left(n_k\left(np\gamma_n\log\gamma_n^{-1}\right)^{\frac{1}{2}} + \gamma_n n\log\gamma_n^{-1}\right)\right\} \\ &\qquad \leq 2\exp\left\{-C_1\gamma_n n\log\gamma_n^{-1}\right\} n^{-(1+C_0)}. \end{aligned}$$

Under the assumption $\frac{1}{\eta_1^2} = \frac{\beta n(p-q)^2}{p} \to \infty$, we have $(p/(\beta n))^{1/2} = \eta_1(p-q)$ and $\frac{1}{\beta n} = \eta_1^2\frac{(p-q)^2}{p} \leq \eta_1^2(p-q)$. Besides, since $\gamma_n = o(-\frac{\beta}{\log\beta})$, we have $\gamma_n \leq -\frac{\beta}{\log\beta}$ as $n\to\infty$, which implies that $\frac{1}{\beta}\gamma_n\log\gamma_n^{-1} \lesssim 1$. Thus, we have

$$\begin{aligned}
&\left|\frac{|\mathcal{E}'_{kk}|}{\frac{1}{2}n'_k(n'_k-1)} - \mathbb{E}\frac{|\mathcal{E}'_{kk}|}{\frac{1}{2}n'_k(n'_k-1)}\right| \\
&\leq C(C_0, C_1)\left(\frac{1}{n_k}\left(np\gamma_n \log \gamma_n^{-1}\right)^{\frac{1}{2}} + \frac{n}{n_k^2}\gamma_n \log \gamma_n^{-1}\right) \\
&\leq C\left(\left(\frac{p}{\beta n}\frac{1}{\beta}\gamma_n \log \gamma_n^{-1}\right)^{\frac{1}{2}} + \frac{1}{\beta n}\frac{1}{\beta}\gamma_n \log \gamma_n^{-1}\right) \\
&\leq C\eta_1(p-q),
\end{aligned} \tag{S6.34}$$

for a positive constant $C > 0$, with probability at least $1 - 2n^{-(1+C_0)}$ $\times \exp\{-C_1\gamma_n n \log \gamma_n^{-1}\}$. Using the triangle inequality to (S6.32), (S6.33) and (S6.34), we have

$$\left|\frac{|\mathcal{E}'_{kk}|}{\frac{1}{2}n'_k(n'_k-1)} - P_{kk}\right| \leq \eta(p-q),$$

with probability at least $1 - 2\exp\{-C_1\gamma_n n \log \gamma_n^{-1}\} n^{-(1+C_0)}$, and here $\eta = C(\eta_1 + \eta_2)$ for some large constant $C > 0$. Next, apply the union bound to obtain the upper bound of $\left|\frac{|\widetilde{\mathcal{E}}_{kk}|}{\frac{1}{2}\widetilde{n}_k(\widetilde{n}_k-1)} - P_{kk}\right|$. For the sake of simplicity, define event

$$\mathcal{F}_{\mathcal{C}'_k} = \left\{\left|\frac{|\mathcal{E}'_{kk}|}{\frac{1}{2}n'_k(n'_k-1)} - P_{kk}\right| \leq \eta(p-q)\right\},$$

and let $\{\mathcal{C}'_k\}$ denote the set that contains all $\mathcal{C}'_k$ that satisfies (S6.31). Hence, $|\{\mathcal{C}'_k\}| \leq \exp\{C_1\gamma_n n \log \gamma_n^{-1}\}$. Then, we have

$$\mathbb{P}\left(\bigcap_{\mathcal{C}'_k \in \{\mathcal{C}'_k\}} \mathcal{F}_{\mathcal{C}'_k} \middle| E\right) = 1 - \mathbb{P}\left(\bigcup_{\mathcal{C}'_k \in \{\mathcal{C}'_k\}} \mathcal{F}^c_{\mathcal{C}'_k} \middle| E\right) \geq 1 - 2n^{-(1+C_0)}.$$

Similar to $\mathcal{F}_{\mathcal{C}'_k}$, define event

$$\mathcal{F}_{\widetilde{\mathcal{C}}_k} = \left\{ \left| \frac{\left|\widetilde{\mathcal{E}}_{kk}\right|}{\frac{1}{2}\widetilde{n}_k(\widetilde{n}_k - 1)} - P_{kk} \right| \leq \eta(p-q) \right\},$$

and then, we have

$$\mathbb{P}\left(\mathcal{F}_{\widetilde{\mathcal{C}}_k} | E\right) \geq \mathbb{P}\left(\bigcap_{\mathcal{C}'_k \in \{\mathcal{C}'_k\}} \mathcal{F}_{\mathcal{C}'_k} \middle| E \right) \geq 1 - 2n^{-(1+C_0)}.$$

The above completes the proof of the upper bound of $\left|\tilde{P}^0_{kk} - P_{kk}\right|$ for any k in [2]. The proof for $\left|\tilde{P}^0_{kl} - P_{kl}\right|$ for $k \neq l \in [2]$, are analogous and hence are omitted. Based on these results, a final union bound over $\{(k,l) : k,l \in [2]\}$ can be obtained, i.e.

$$\mathbb{P}\left\{ \max_{k,l\in[2]} \left|\tilde{P}^0_{kl} - P_{kl}\right| \leq \eta(p-q) \middle| E \right\} \geq 1 - 9n^{-(1+C_0)}.$$

Then,

$$\begin{aligned} & \mathbb{P}\left\{ \max_{k,l\in[2]} \left|\tilde{P}^0_{kl} - P_{kl}\right| > \eta(p-q) \right\} \\ = \;& \mathbb{P}(E)\mathbb{P}\left\{ \max_{k,l\in[2]} \left|\tilde{P}^0_{kl} - P_{kl}\right| > \eta(p-q) \middle| E \right\} \\ & + \mathbb{P}(E^c)\mathbb{P}\left\{ \max_{k,l\in[2]} \left|\tilde{P}^0_{kl} - P_{kl}\right| > \eta(p-q) \middle| E^c \right\} \\ \leq \;& 9n^{-(1+C_0)} + n^{-(1+C_0)} = 10n^{-(1+C_0)}. \end{aligned}$$

Review that $\tilde{\ell}\left(\boldsymbol{c}, \tilde{\boldsymbol{c}}^0\right) = \min_{\pi:[2]\to[2]} \ell\left(\pi[\boldsymbol{c}], \tilde{\boldsymbol{c}}^0\right)$. From the above proof, if $\ell\left(\pi[\boldsymbol{c}], \tilde{\boldsymbol{c}}^0\right) = \tilde{\ell}\left(\boldsymbol{c}, \tilde{\boldsymbol{c}}^0\right)$ with $\pi(1) = 1$ and $\pi(2) = 2$, we have $\eta(p-q) \geq \max_{k,l\in[2]} \left|\tilde{P}^0_{kl} - P_{kl}\right|$ holds with probability at least $1 - 10n^{-(1+C_0)}$. On

$\eta(p-q) \geq \max_{k,l\in[2]}\left|\tilde{P}^0_{kl} - P_{kl}\right|$, we have $\tilde{P}^0_{11} > p - \eta(p-q)$ and $\tilde{P}^0_{22} < q + \eta(p-q)$, and then, $\tilde{P}^0_{11} - \tilde{P}^0_{22} > p - q - 2\eta(p-q) > 0$, which implies $p^0 = \max\{\tilde{P}^0_{11}, \tilde{P}^0_{22}\} = \tilde{P}^0_{11}$ and $k^* = 2$. And easy to obtain $|p^0 - p| \leq \eta(p-q)$, $|q^0 - q| \leq \eta(p-q)$ and $\boldsymbol{c}^0 = \tilde{\boldsymbol{c}}^0$. If $\ell\left(\pi[\boldsymbol{c}], \tilde{\boldsymbol{c}}^0\right) = \tilde{\ell}\left(\boldsymbol{c}, \tilde{\boldsymbol{c}}^0\right)$ with $\pi(1) = 2$ and $\pi(2) = 1$, we have $\max_{k,l\in[2]}\left|\tilde{P}^0_{kl} - P_{\pi(k)\pi(l)}\right| \leq \eta(p-q)$ holds with probability at least $1 - 10n^{-(1+C_0)}$. On $\max_{k,l\in[2]}\left|\tilde{P}^0_{kl} - P_{\pi(k)\pi(l)}\right| \leq \eta(p-q)$, we have $\tilde{P}^0_{22} > p - \eta(p-q)$ and $\tilde{P}^0_{11} < q + \eta(p-q)$, and then, $\tilde{P}^0_{22} - \tilde{P}^0_{11} > p - q - 2\eta(p-q) > 0$, which implies $p^0 = \max\{\tilde{P}^0_{11}, \tilde{P}^0_{22}\} = \tilde{P}^0_{22}$ and $k^* = 1$. And easy to obtain $|p^0 - p| \leq \eta(p-q)$, $|q^0 - q| \leq \eta(p-q)$ and $\boldsymbol{c}^0 = \pi[\tilde{\boldsymbol{c}}^0]$.

Then, the proof of this lemma is completed since $C = 10$ and η do not rely on $(P, \boldsymbol{c})$.

□

Proof of Lemma 4. For any $(\boldsymbol{P}, \boldsymbol{c}) \in \Theta_n(p, q, \beta)$ and $i \in [n]$, define event $D_i = \left\{\ell\left(\boldsymbol{c}_{-i}, \boldsymbol{c}^0_{-i}\right) \leq \gamma_n\right\}$. And then, we have

$$
\begin{aligned}
& \mathbb{P}\big(\hat{\boldsymbol{c}}_{p,q}(i) \neq \boldsymbol{c}(i)\big) \\
= \;& \mathbb{P}_{\boldsymbol{c}(i)=1}\big(\hat{\boldsymbol{c}}_{p,q}(i) = 2\big) + \mathbb{P}_{\boldsymbol{c}(i)=2}\big(\hat{\boldsymbol{c}}_{p,q}(i) = 1\big) \\
= \;& \mathbb{P}_{\boldsymbol{c}(i)=1}\big\{(\hat{\boldsymbol{c}}_{p,q}(i) = 2) \cap D_i\big\} + \mathbb{P}_{\boldsymbol{c}(i)=1}\big\{(\hat{\boldsymbol{c}}_{p,q}(i) = 2) \cap D_i^c\big\} \\
& + \mathbb{P}_{\boldsymbol{c}(i)=2}\big\{(\hat{\boldsymbol{c}}_{p,q}(i) = 1) \cap D_i\big\} + \mathbb{P}_{\boldsymbol{c}(i)=2}\big\{(\hat{\boldsymbol{c}}_{p,q}(i) = 1) \cap D_i^c\big\} \\
\leq \;& \mathbb{P}_{\boldsymbol{c}(i)=1}\big\{(\hat{\boldsymbol{c}}_{p,q}(i) = 2) \cap D_i\big\} + \mathbb{P}_{\boldsymbol{c}(i)=1}\big\{D_i^c\big\}
\end{aligned}
$$

$$+ \mathbb{P}_{\boldsymbol{c}(i)=2}\big\{(\hat{\boldsymbol{c}}_{p,q}(i)=1)\cap D_i\big\} + \mathbb{P}_{\boldsymbol{c}(i)=2}\big\{D_i^c\big\}.$$

Next, we just need to provide the upper bounds of $\mathbb{P}_{\boldsymbol{c}(i)=1}\big\{(\hat{\boldsymbol{c}}_{p,q}(i)=2)\cap D_i\big\}$ and $\mathbb{P}_{\boldsymbol{c}(i)=2}\big\{(\hat{\boldsymbol{c}}_{p,q}(i)=1)\cap D_i\big\}$ separately.

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{c}(i)=1}\big\{(\hat{\boldsymbol{c}}_{p,q}(i)=2)\cap D_i\big\}\\
&= \mathbb{P}_{\boldsymbol{c}(i)=1}\Big[\Big\{ \sum_{j\neq i:\boldsymbol{c}^{0i}(j)=1} (A_{ij}\log q + (1-A_{ij})\log(1-q))\\
&\qquad\qquad > \sum_{j\neq i:\boldsymbol{c}^{0i}(j)=1} (A_{ij}\log p + (1-A_{ij})\log(1-p))\Big\}\cap D_i\Big]\\
&= \mathbb{P}_{\boldsymbol{c}(i)=1}\left[\left\{\sum_{j\neq i:\boldsymbol{c}^{0i}(j)=1}\left(A_{ij}\log\frac{q}{p} + (1-A_{ij})\log\frac{1-q}{1-p}\right)>0\right\}\cap D_i\right]\\
&= \mathbb{E}\left(\mathbb{E}_{\boldsymbol{c}(i)=1}\left[\mathbb{I}\left\{\sum_{\substack{j\neq i:\\ \boldsymbol{c}^{0i}(j)=1}}\left(A_{ij}\log\frac{q}{p} + (1-A_{ij})\log\frac{1-q}{1-p}\right)>0\right\}\right.\right.\\
&\qquad\qquad \left.\left.\times\mathbb{I}\{D_i\}\,\middle|\,\boldsymbol{A}^{-i}\right]\right).
\end{aligned}$$

Besides, note that $\boldsymbol{A}^{-i}$ and $\boldsymbol{A}_i$ are independent of each other. And then, using the Chernoff bound, choose $t=t^*$, we have

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{c}(i)=1}\big\{(\hat{\boldsymbol{c}}_{p,q}(i)=2)\cap D_i\big\}\\
&\le \mathbb{E}\Bigg[\prod_{\substack{j\neq i:\boldsymbol{c}^{0i}(j)=1\\ \boldsymbol{c}(j)=1}}\big\{p^{1-t^*}q^{t^*} + (1-p)^{1-t^*}(1-q)^{t^*}\big\}\\
&\qquad\times\prod_{\substack{j\neq i:\boldsymbol{c}^{0i}(j)=1\\ \boldsymbol{c}(j)=2}}\big\{p^{-t^*}q^{1+t^*} + (1-p)^{-t^*}(1-q)^{1+t^*}\big\}\mathbb{I}\{D_i\}\Bigg]\\
&\le \mathbb{E}\Bigg[\exp\Big\{-(n_1(\boldsymbol{c})-\gamma_n n)\,I_{t^*}\Big\}
\end{aligned}$$

$$\times\Big\{ \big(p^{-t^*}q^{1+t^*} + (1-p)^{-t^*}(1-q)^{1+t^*}\big)^{\gamma_n n} \Big\}\mathbb{I}\{D_i\}\Bigg]$$

$$\leq \exp\big(-(1-\eta)\beta n I_{t^*}\big)\mathbb{P}\{D_i\}$$

$$\leq \exp\big(-(1-\eta)\beta n I_{t^*}\big),$$

for some positive sequence $\eta \to 0$. The third inequality holds because $p^{-t^*}q^{1+t^*} + (1-p)^{-t^*}(1-q)^{1+t^*} \leq \{(1-q)/(1-p)\}^{t^*} = \exp[t^* \log\{1+(p-q)/(1-p)\}] \leq \exp\{t^*(p-q)/(1-p)\}$, and by Lemma 1 and Lemma 2, when $\lim_{n\to\infty} p/q > 1$, $\gamma_n n t^*(p-q)/(1-p) \asymp \gamma_n n t^*(p-q)^2/p \lesssim \gamma_n n I_{t^*} \ll \beta n I_{t^*}$ if $\gamma_n = o(\beta)$ and when $\lim_{n\to\infty} p/q = 1$, $\gamma_n n t^*(p-q)/(1-p) \ll \beta n(p-q)^2/p \asymp \beta n I_{t^*}$ if $\gamma_n = o(\beta(p-q)/p)$. Similarly, we have

$$\mathbb{P}_{\boldsymbol{c}(i)=2}\big\{(\hat{\boldsymbol{c}}_{p,q}(i)=1)\cap D_i\big\}$$

$$= \mathbb{P}_{\boldsymbol{c}(i)=2}\Big[\Big\{ \sum_{j\neq i:\boldsymbol{c}^{0i}(j)=1} (A_{ij}\log p + (1-A_{ij})\log(1-p))$$

$$> \sum_{j\neq i:\boldsymbol{c}^{0i}(j)=1} (A_{ij}\log q + (1-A_{ij})\log(1-q)) \Big\} \cap D_i\Big]$$

$$= \mathbb{P}_{\boldsymbol{c}(i)=2}\left[\left\{ \sum_{j\neq i:\boldsymbol{c}^{0i}(j)=1} \left(A_{ij}\log\frac{p}{q} + (1-A_{ij})\log\frac{1-p}{1-q}\right) > 0 \right\} \cap D_i\right]$$

$$= \mathbb{E}\left(\mathbb{E}_{\boldsymbol{c}(i)=2}\left[\mathbb{I}\left\{ \sum_{\substack{j\neq i:\\ \boldsymbol{c}^{0i}(j)=1}} \left(A_{ij}\log\frac{p}{q} + (1-A_{ij})\log\frac{1-p}{1-q}\right) > 0 \right\}\right.\right.$$

$$\left.\left.\times\mathbb{I}\{D_i\}\middle|\boldsymbol{A}^{-i}\right]\right).$$

And then, using the Chernoff bound, choose $t = 1-t^*$, we have

$$\mathbb{P}_{\boldsymbol{c}(i)=2}\big\{(\hat{\boldsymbol{c}}_{p,q}(i)=1)\cap D_i\big\}$$

$$
\begin{aligned}
&\leq \mathbb{E}\left[\prod_{j\neq i:\boldsymbol{c}^{0i}(j)=1}\{p^{1-t^*}q^{t^*}+(1-p)^{1-t^*}(1-q)^{t^*}\}\mathbb{I}\{D_i\}\right] \\
&\leq \mathbb{E}\left[\exp\Big\{-\left(n_1(\boldsymbol{c})-\gamma_n n\right)I_{t^*}\Big\}\mathbb{I}\{D_i\}\right] \\
&\leq \exp\big(-(1-\eta)\beta n I_{t^*}\big)\mathbb{P}\{D_i\} \\
&\leq \exp\big(-(1-\eta)\beta n I_{t^*}\big),
\end{aligned}
$$

for some positive sequence $\eta \to 0$. Hence, we have that

$$
\mathbb{P}\big(\hat{\boldsymbol{c}}_{p,q}(i)\neq \boldsymbol{c}(i)\big) \leq \exp\big(-(1-\eta)\beta n I_{t^*}\big) + 2n^{-(1+C_0)},
$$

for some positive sequence $\eta \to 0$. Note that the results are independent of $(\boldsymbol{P}, \boldsymbol{c})$ and i, thereby proving completion.

□

Proof of Lemma 5. The proof of Lemma 5 is similar to the proof of Lemma 2 in Gao et al. (2018). Since (S6.15) is a a simple-versus-simple hypothesis testing problem, by the Neyman-Pearson lemma, the optimal test is the likelihood ratio test ϕ^*, which rejects H_0 if

$$
\prod_{i=1}^{n_0} p^{X_i}(1-p)^{1-X_i} < \prod_{i=1}^{n_0} q^{X_i}(1-q)^{1-X_i}.
$$

Therefore,

$$
\mathbb{P}_{H_0}\phi^* = \mathbb{P}\left(\sum_{i=1}^{n_0}\left[X_i\log\frac{q}{p}+(1-X_i)\log\frac{1-q}{1-p}\right]>0\right). \tag{S6.35}
$$

Let $Y_1, \cdots, Y_{n_0}$ be independent and identically distributed random variables with the distribution given by

$$\mathbb{P}\left(Y = \log\frac{q}{p}\right) = p, \quad \mathbb{P}\left(Y = \log\frac{1-q}{1-p}\right) = 1-p.$$

Let $\boldsymbol{Y} = (Y_1, \cdots, Y_{n_0})^\top$, $\boldsymbol{y} = (y_1, \cdots, y_{n_0})^\top \in \mathbb{R}^{n_0}$ and for any fixed $t > 0$, let $K_n(t) = \log \mathbb{E}(e^{tY_1})$. Then, we have

$$\begin{aligned}
\mathbb{P}_{H_0}\phi^* &= \mathbb{P}\left(\sum_{i=1}^{n_0} Y_i > 0\right) \\
&= \sum_{\sum_{i=1}^{n_0} y_i > 0} \mathbb{P}\left(\boldsymbol{Y} = \boldsymbol{y}\right) = \sum_{\sum_{i=1}^{n_0} y_i > 0} \prod_{i=1}^{n_0} \mathbb{P}\left(Y_i = y_i\right) \\
&= e^{n_0 K_n(t^*)} \sum_{\sum_{i=1}^{n_0} y_i > 0} e^{-t^* \sum_{i=1}^{n_0} y_i} \prod_{i=1}^{n_0} \frac{\mathbb{P}\left(Y_i = y_i\right) e^{t^* y_i}}{e^{K_n(t^*)}},
\end{aligned}$$

where t^* is defined in (3.11). In addition, let $Z_1, \cdots, Z_{n_0}$ be independent and identically distributed random variables with the distribution given by

$$\begin{aligned}
\mathbb{P}_Z\left(Z = \log\frac{q}{p}\right) &= \frac{p^{1-t^*} q^{t^*}}{e^{K_n(t^*)}}, \\
\mathbb{P}_Z\left(Z = \log\frac{1-q}{1-p}\right) &= \frac{(1-p)^{1-t^*}\left(1-q\right)^{t^*}}{e^{K_n(t^*)}}.
\end{aligned}$$

Note that

$$e^{n_0 K_n(t^*)} = e^{n_0 \log\left(p^{1-t^*} q^{t^*} + (1-p)^{1-t^*}(1-q)^{t^*}\right)} = e^{-n_0 I_{t^*}},$$

and due to Lemma 1, t^* is the only maximum point of $e^{n_0 K_n(t)}$ on $t > 0$. Let $f(t) = p^{1-t}q^t + (1-p)^{1-t}(1-q)^t$. We have

$$f'(t) = p^{1-t} q^t \log\frac{q^t}{p^t} + (1-p)^{1-t}\left(1-q\right)^t \log\frac{(1-q)^t}{(1-p)^t}.$$

Thus, we have $\mathbb{E}_Z(Z) = \frac{f'(t^*)}{e^{K_n(t^*)}} = 0$. In addition, for some $L > 0$,

$$\begin{aligned}
&\sum_{\sum_{i=1}^{n_0} y_i > 0} e^{-t^* \sum_{i=1}^{n_0} y_i} \prod_{i=1}^{n_0} \frac{\mathbb{P}\left(Y_i = y_i\right) e^{t^* y_i}}{e^{K_n(t^*)}} \\
\geq & \sum_{0 < \sum_{i=1}^{n_0} y_i < L} e^{-\sum_{i=1}^{n_0} y_i} \prod_{i=1}^{n_0} \frac{\mathbb{P}\left(Y_i = y_i\right) e^{t^* y_i}}{e^{K_n(t^*)}} \\
\geq & e^{-t^* L} \sum_{0 < \sum_{i=1}^{n_0} y_i < L} \prod_{i=1}^{n_0} \frac{\mathbb{P}\left(Y_i = y_i\right) e^{t^* y_i}}{e^{K_n(t^*)}} \\
= & e^{-t^* L} \sum_{0 < \sum_{i=1}^{n_0} y_i < L} \prod_{i=1}^{n_0} \mathbb{P}_Z\left(Z_i = y_i\right) \\
= & e^{-t^* L} \mathbb{P}_{\boldsymbol{Z}}\left(0 < \sum_{i=1}^{n_0} Z_i < L\right),
\end{aligned}$$

where $\boldsymbol{Z} = (Z_1, \cdots, Z_{n_0})^\top$.

Then, we consider the lower bound of $e^{-t^* L} \mathbb{P}_{\boldsymbol{Z}}\left(0 < \sum_{i=1}^{n_0} Z_i < L\right)$. First, we calculate the expectation of Z_i^2. For any $i \in [n_0]$,

$$\mathbb{E}_{\boldsymbol{Z}}(Z_i^2) = \frac{p^{1-t^*} q^{t^*}}{e^{K_n(t^*)}} \left(\log \frac{q}{p}\right)^2 + \frac{(1-p)^{1-t^*}(1-q)^{t^*}}{e^{K_n(t^*)}} \left(\log \frac{1-q}{1-p}\right)^2.$$

When $p \asymp q$, by Lemma 2, $I_{t^*} = -K_n(t^*) \asymp \delta^2 q$, and then we have $e^{K_n(t^*)} = e^{-I_{t^*}} \asymp 1$. Moreover, we have $t^* \in [\epsilon, 1-\epsilon]$, for a small enough constant $\epsilon > 0$ by Lemma 1. Thus,

$$\mathbb{E}_{\boldsymbol{Z}}(Z_i^2) \asymp \tfrac{(p-q)^2}{p}, \ \ \textstyle\sum_{i=1}^{n_0} \mathrm{Var}_{\boldsymbol{Z}}\, Z_i \asymp n_0 \tfrac{(p-q)^2}{p}.$$

Note that the value of Z_i is bounded by some constant, for any $i \in [n_0]$. Under the assumption that $n_0 I_{t^*}(p,q) \to \infty$ and $p \asymp q$, we have $\frac{n_0(p-q)^2}{p} \to$

∞ by Lemma 2. Hence, $\mathbb{I}\left\{|Z_i - \mathbb{E}_{\boldsymbol{Z}}(Z_i)| > \varepsilon\left(\sum_{i=1}^{n_0} \mathrm{Var}_{\boldsymbol{Z}}(Z_i)\right)^{1/2}\right\} \to 0$ for every i and for any constant $\varepsilon > 0$. Thus

$$\lim_{n_0\to\infty} \sum_{i=1}^{n_0} \mathbb{E}_{\boldsymbol{Z}}\left(Z_i - \mathbb{E}_{\boldsymbol{Z}}(Z_i)\right)^2 \mathbb{I}\left\{|Z_i - \mathbb{E}_{\boldsymbol{Z}}(Z_i)| > \varepsilon\left(\sum_{i=1}^{n_0} \mathrm{Var}_{\boldsymbol{Z}}(Z_i)\right)^{\frac{1}{2}}\right\} = 0,$$

for any constant $\varepsilon > 0$. Together with $\sum_{i=1}^{n_0} \mathbb{E}_{\boldsymbol{Z}}(Z_i) = 0$, the Lindeberg condition implies that under $\mathbb{P}_{\boldsymbol{Z}}$, $\frac{\sum_{i=1}^{n_0} Z_i}{\left(\sum_{i=1}^{n_0} \mathrm{Var}_{\boldsymbol{Z}}(Z_i)\right)^{1/2}}$ converges to $N(0,1)$. Then set $L = \left(\sum_{i=1}^{n_0} \mathrm{Var}_{\boldsymbol{Z}}(Z_i)\right)^{1/2} \asymp \sqrt{n_0 I_{t^*}}$. Hence, $L = o(n_0 I_{t^*})$, and combined with $t^* \in [\epsilon, 1-\epsilon]$, for a small enough constant $\epsilon > 0$, we have

$$\mathbb{P}_{H_0}\phi^* \geq \exp\Big(-(1+\eta)n_0 I_{t^*}\Big),$$

for some $\eta \to 0$.

When $p \gg q$, by Lemma 2, $t^* p \lesssim I_{t^*} \lesssim p$, and we have $e^{K_n(t^*)} = e^{-I_{t^*}} \asymp 1$. Then,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{Z}}(Z_i^2) &\asymp p^{1-t^*} q^{t^*} \left(\log\frac{q}{p}\right)^2 + (1-p)^{1-t^*}(1-q)^{t^*}\left(\log\frac{1-q}{1-p}\right)^2 \\
&= p\exp\left\{\big(1+o(1)\big)\log\log(1+\delta)\right\}.
\end{aligned}$$

The last equality holds since $t^* = \big(1+o(1)\big)\frac{\log(\log(1+\delta))}{\log(1+\delta)}$ when $\delta \to \infty$ by Lemma 1, where $\delta = \frac{p-q}{q}$. Similarly, we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{Z}}(|Z_i|^3) &\asymp p^{1-t^*} q^{t^*} \left|\log\frac{q}{p}\right|^3 + (1-p)^{1-t^*}(1-q)^{t^*}\left|\log\frac{1-q}{1-p}\right|^3 \\
&= p\exp\left\{2\big(1+o(1)\big)\log\log(1+\delta)\right\}.
\end{aligned}$$

Under the assumption $p\log^3(p/q) < \infty$, we have $\mathbb{E}_{\boldsymbol{Z}}(|Z_i|^3) < \infty$. Then, we have

$$\frac{\mathbb{E}_{\boldsymbol{Z}}(|Z_i|^3)}{(\mathbb{E}_{\boldsymbol{Z}}(Z_i^2))^{\frac{3}{2}}} = p^{-\frac{1}{2}} \exp\left\{\frac{1}{2}\big(1+o(1)\big)\log\log(1+\delta)\right\}.$$

By the assumption that when $p \gg q$,

$$\lim_{n\to\infty} \frac{\log \frac{\log(\frac{p}{q})}{p}}{\log n} < 1,$$

we have $\frac{\mathbb{E}_{\boldsymbol{Z}}(|Z_i|^3)}{\left(\mathbb{E}_{\boldsymbol{Z}}(Z_i^2)\right)^{\frac{3}{2}}} = o(\sqrt{n})$. And then, by Corollary 2.7.2 of Lehmann (1999), i.e. Liapounov theorem, we have that under $\mathbb{P}_{\boldsymbol{Z}}$, $\frac{\sum_{i=1}^{n_0} Z_i}{\left(\sum_{i=1}^{n_0} \mathrm{Var}_{\boldsymbol{Z}}(Z_i)\right)^{1/2}}$ converges to $N(0,1)$. Then set $L = (\sum_{i=1}^{n_0} \mathrm{Var}_{\boldsymbol{Z}}(Z_i))^{1/2}$. By the assumption that when $p \gg q$,

$$\lim_{n\to\infty} \frac{\log \beta n p}{\log\log \frac{p}{q}} > 3,$$

we have $L \ll n_0 p t^* \lesssim n_0 I_{t^*}$, which means

$$\mathbb{P}_{H_0}\phi^* \geq \exp\Big(-(1+\eta)n_0 I_{t^*}\Big),$$

for some $\eta \to 0$.

By the same way, it can be proved that

$$\mathbb{P}_{H_1}\big(1-\phi^*\big) \geq \exp\Big(-(1+\eta)n_0 I_{t^*}\Big),$$

for some $\eta \to 0$.

□

Proof of Lemma 7. Fix any $(\boldsymbol{P}, \boldsymbol{c}) \in \Theta_n(p,q,\beta)$ and $i \in [n]$. $\boldsymbol{c}^{0i}$, $\hat{p}^{0i}$, $\hat{q}^{0i}$ and $\check{\boldsymbol{c}}$ are obtained in the initialization step and the refinement step of Algorithm 1.

Then, for some $\eta_0 \to 0$, which will be determined later. For $i \in [n]$, define events

$$F_i = \Big\{\ell\left(\boldsymbol{c}, \boldsymbol{c}^{0i}\right) \leq \gamma_n, \left|\hat{p}^{0i} - p\right| \leq \eta_0(p-q), \left|\hat{q}^{0i} - q\right| \leq \eta_0(p-q)\Big\}.$$

Combined the assumption of Lemma 7, by Lemma 6, we have

$$\inf_{(\boldsymbol{P},\boldsymbol{c})\in\Theta_n(p,q,\beta)} \min_{i\in[n]} \mathbb{P}_{\boldsymbol{P},\boldsymbol{c}}\left(F_i\right) \geq 1 - Cn^{-(1+C_0)},$$

for constants $C, C_0 > 0$, with $\eta_0 = o(p-q)$, and $\eta_0 = o\left(q\frac{\log\log\frac{p}{q}}{\log\frac{p}{q}}\right)$ when $p \gg q$. Moreover, for simplicity, we write $\hat{p}$ as $\hat{p}^{0i}$ and $\hat{q}$ as $\hat{q}^{0i}$. Then, we have

$$\mathbb{P}\Big(\check{\boldsymbol{c}}(i) \neq \boldsymbol{c}(i), F_i\Big) \leq \mathbb{P}_{\boldsymbol{c}(i)=1}\Big(\check{\boldsymbol{c}}(i) = 2, F_i\Big) + \mathbb{P}_{\boldsymbol{c}(i)=2}\Big(\check{\boldsymbol{c}}(i) = 1, F_i\Big).$$

The next step is to get the upper bounds of the two terms on the right-hand side of the above inequality. Firstly, we have

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{c}(i)=1}\Big(\check{\boldsymbol{c}}(i) = 2 \text{ and } F_i\Big)\\
=&\mathbb{P}_{\boldsymbol{c}(i)=1}\left[\sum_{j:\check{\boldsymbol{c}}(j)=1}\left(A_{ij}\log\frac{\hat{q}}{\hat{p}} + (1-A_{ij})\log\frac{1-\hat{q}}{1-\hat{p}}\right) > 0 \text{ and } F_i\right]\\
=&\mathbb{P}_{\boldsymbol{c}(i)=1}\Bigg[\sum_{\substack{j:\check{\boldsymbol{c}}(j)=1\\ \boldsymbol{c}(j)=1}}\left(A_{ij}\log\frac{\hat{q}}{\hat{p}} + (1-A_{ij})\log\frac{1-\hat{q}}{1-\hat{p}}\right)\\
&\quad + \sum_{\substack{j:\check{\boldsymbol{c}}(j)=1\\ \boldsymbol{c}(j)=2}}\left(A_{ij}\log\frac{\hat{q}}{\hat{p}} + (1-A_{ij})\log\frac{1-\hat{q}}{1-\hat{p}}\right) > 0 \,\Bigg|\, F_i\Bigg]\mathbb{P}(F_i). \qquad \text{(S6.36)}
\end{aligned}$$

Next, we focus on the conditional distribution in (S6.36). On $\boldsymbol{c}(i) = 1$, using the Chernoff bound, for any $\lambda > 0$,

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{c}(i)=1}\Bigg[\sum_{\substack{j:\check{\boldsymbol{c}}(j)=1\\ \boldsymbol{c}(j)=1}} \left(A_{ij}\log\frac{\hat{q}}{\hat{p}} + (1-A_{ij})\log\frac{1-\hat{q}}{1-\hat{p}} \right) \\
&\quad + \sum_{\substack{j:\check{\boldsymbol{c}}(j)=1\\ \boldsymbol{c}(j)=2}} \left(A_{ij}\log\frac{\hat{q}}{\hat{p}} + (1-A_{ij})\log\frac{1-\hat{q}}{1-\hat{p}} \right) > 0 \Bigg| F_i \Bigg] \\
&\le \prod_{\substack{j:\check{\boldsymbol{c}}(j)=1\\ \boldsymbol{c}(j)=1}} \mathbb{E}_{\cdot|F_i,\boldsymbol{c}(i)=1} \exp\left[\lambda \left(A_{ij}\log\frac{\hat{q}}{\hat{p}} + (1-A_{ij})\log\frac{1-\hat{q}}{1-\hat{p}} \right) \right] \\
&\quad \times \prod_{\substack{j:\check{\boldsymbol{c}}(j)=1\\ \boldsymbol{c}(j)=2}} \mathbb{E}_{\cdot|F_i,\boldsymbol{c}(i)=1} \exp\left[\lambda \left(A_{ij}\log\frac{\hat{q}}{\hat{p}} + (1-A_{ij})\log\frac{1-\hat{q}}{1-\hat{p}} \right) \right].
\end{aligned} \tag{S6.37}$$

Note that for any $j \in \{j' : \check{\boldsymbol{c}}(j') = 1, \boldsymbol{c}(j') = 1\}$, $P(A_{ij} = 1) = p$ when $\boldsymbol{c}(i) = 1$. Then,

$$\begin{aligned}
&\mathbb{E}_{\cdot|F_i,\boldsymbol{c}(i)=1} \exp\left[\lambda \left(A_{ij}\log\frac{\hat{q}}{\hat{p}} + (1-A_{ij})\log\frac{1-\hat{q}}{1-\hat{p}} \right) \right] \\
= \;& p\left(\frac{\hat{q}}{\hat{p}}\right)^{\lambda} + (1-p)\left(\frac{1-\hat{q}}{1-\hat{p}}\right)^{\lambda} \\
= \;& q^{\lambda}p^{1-\lambda}\left[\left(\frac{p\hat{q}}{\hat{p}q}\right)^{\lambda} - 1 \right] + (1-q)^{\lambda}(1-p)^{1-\lambda}\left[\left(\frac{(1-p)(1-\hat{q})}{(1-\hat{p})(1-q)}\right)^{\lambda} - 1 \right] \\
&+ q^{\lambda}p^{1-\lambda} + (1-q)^{\lambda}(1-p)^{1-\lambda}.
\end{aligned}$$

Since p keeps away from 1, by simple manipulation, we have $\frac{p\hat{q}}{\hat{p}q} \le 1 + C\eta_0\delta$ and $\frac{(1-p)(1-\hat{q})}{(1-\hat{p})(1-q)} \le 1 + C\eta_0\delta$, for some constant $C > 0$. Consider the fact that $(1+Cx)^{\lambda} - 1 \le 2C\lambda x$ as $x \to 0$ and $\eta_0 = o(p-q)$, and $\eta_0 = o\left(q\frac{\log\log\frac{p}{q}}{\log\frac{p}{q}}\right)$

when $p \gg q$. Then, set $\lambda = t^*$, where $t^* \in (0,1)$ by Lemma 1, and we have

$$\begin{aligned} & \mathbb{E}_{\cdot|F_i,\boldsymbol{c}(i)=1} \exp\left[t^*\left(A_{ij}\log\frac{\hat{q}}{\hat{p}} + (1-A_{ij})\log\frac{1-\hat{q}}{1-\hat{p}}\right)\right] \\ \leq\ & \left(q^{t^*}p^{1-t^*} + (1-q)^{t^*}(1-p)^{1-t^*}\right)(1+2Ct^*\eta_0\delta) \\ \leq\ & \exp\big(-(1-\eta)I_{t^*}\big), \end{aligned}$$

for a sequence $\eta \to 0$. Here, the last inequality holds since when $p \asymp q$ $\log(1+2Ct^*\eta_0\delta) = o(\delta^2 p)$ and $I_{t^*} \asymp \delta^2 q$ by (1) of Lemma 2 and when $p \gg q$ $\log(1+2Ct^*\eta_0\delta) = o(t^*\delta q)$ and $I_{t^*} \gtrsim t^* p = t^*(1+\delta)q$ by (2) of Lemma 2. Besides, for any $j \in \{j' : \check{\boldsymbol{c}}(j') = 1, \boldsymbol{c}(j') = 2\}$, $P(A_{ij} = 1) = q$. Then,

$$\begin{aligned} & \mathbb{E}_{\cdot|F_i,\boldsymbol{c}(i)=1} \exp\left[\lambda\left(A_{ij}\log\frac{\hat{q}}{\hat{p}} + (1-A_{ij})\log\frac{1-\hat{q}}{1-\hat{p}}\right)\right] \\ =\ & q\left(\frac{\hat{q}}{\hat{p}}\right)^{\lambda} + (1-q)\left(\frac{1-\hat{q}}{1-\hat{p}}\right)^{\lambda} \\ =\ & q^{\lambda}p^{1-\lambda}\frac{q}{p} + (1-q)^{\lambda}(1-p)^{1-\lambda}\frac{1-q}{1-p} + q^{\lambda}p^{1-\lambda}\frac{q}{p}\left[\left(\frac{p\hat{q}}{\hat{p}q}\right)^{\lambda} - 1\right] \\ & +(1-q)^{\lambda}(1-p)^{1-\lambda}\frac{1-q}{1-p}\left[\left(\frac{(1-p)(1-\hat{q})}{(1-\hat{p})(1-q)}\right)^{\lambda} - 1\right]. \end{aligned}$$

Besides, we also have

$$\begin{aligned} & \mathbb{E}_{\cdot|F_i,\boldsymbol{c}(i)=1} \exp\left[t^*\left(A_{ij}\log\frac{\hat{q}}{\hat{p}} + (1-A_{ij})\log\frac{1-\hat{q}}{1-\hat{p}}\right)\right] \\ \leq\ & \left(q^{t^*}p^{1-t^*}\frac{q}{p} + (1-q)^{t^*}(1-p)^{1-t^*}\frac{1-q}{1-p}\right)(1+2Ct^*\eta_0\delta) \\ \leq\ & 1 + C\delta q, \end{aligned}$$

for a large constant $C > 0$, and although it is somewhat ambiguous, in order to reduce unnecessary symbols, C may have different values in different places, but it will not affect the results. The last inequality holds since $q^{t^*}p^{1-t^*}\frac{q}{p}+(1-q)^{t^*}(1-p)^{1-t^*}\frac{1-q}{1-p}\le q+\frac{1-q}{1-p}(1-q)=1+(1-q)\frac{\delta q}{1-p}$ and $\eta_0\delta=o(\delta q)$. Denote $\check{m}_1=\check{m}_1(\boldsymbol{c},\check{\boldsymbol{c}})=|\{j:\boldsymbol{c}(j)=1,\check{\boldsymbol{c}}(j)=2\}|$, $\check{m}_2=\check{m}_2(\boldsymbol{c},\check{\boldsymbol{c}})=|\{j:\boldsymbol{c}(j)=2,\check{\boldsymbol{c}}(j)=1\}|$ and $n_1=n_1(\boldsymbol{c})=|\{j:\boldsymbol{c}(j)=1\}|$. We have,

$$\begin{aligned}\mathbb{P}_{\boldsymbol{c}(i)=1}\Big(\check{\boldsymbol{c}}(i)=2,F_i\Big)\le&(1+C\delta q)^{\check{m}_2}\exp\big(-(1-\eta)I_{t^*}(n_1-\check{m}_1)\big)\\ \le&\exp\big(-(1-\eta)\beta nI_{t^*}\big),\end{aligned}\tag{S6.38}$$

for a sequence $\eta\to 0$, and in order to reduce unnecessary symbols, η may be different in different places, but it will not affect the results. On F_i, we know that $\check{m}_1\le\gamma_n n$ and $\check{m}_2\le\gamma_n n$. And under the assumption that γ_n satisfied $\gamma_n=o\left(-\frac{\beta(p-q)}{\log\beta}\right)$ and $\gamma_n=o\left(-\frac{\beta q}{\log\beta}\right)$ when $p\gg q$, we can get $(1+C\delta q)^{\check{m}_2}\le\exp\big(\gamma_n n\log(1+C\delta q)\big)=\exp\big(o(1)\beta nI_{t^*}\big)$, which establishes the last inequality. On the other hand,

$$\begin{aligned}&\mathbb{P}_{\boldsymbol{c}(i)=2}\Big(\check{\boldsymbol{c}}(i)=1,F_i\Big)\\ =&\ \mathbb{P}_{\boldsymbol{c}(i)=2}\left[-\sum_{j:\check{\boldsymbol{c}}(j)=1}\left(A_{ij}\log\frac{\hat{q}}{\hat{p}}+(1-A_{ij})\log\frac{1-\hat{q}}{1-\hat{p}}\right)>0,F_i\right]\\ =&\ \mathbb{P}_{\boldsymbol{c}(i)=2}\left[-\sum_{j:\check{\boldsymbol{c}}(j)=1}\left(A_{ij}\log\frac{\hat{q}}{\hat{p}}+(1-A_{ij})\log\frac{1-\hat{q}}{1-\hat{p}}\right)>0\,\Bigg|F_i\right]\mathbb{P}(F_i).\end{aligned}$$

Next, we focus on the conditional distribution in the last equation above.

On $\boldsymbol{c}(i) = 2$, using Chernoff bound again, for any $\lambda > 0$,

$$\mathbb{P}_{\boldsymbol{c}(i)=2}\left[-\sum_{j:\check{\boldsymbol{c}}(j)=1} \left(A_{ij} \log \frac{\hat{q}}{\hat{p}} + (1-A_{ij}) \log \frac{1-\hat{q}}{1-\hat{p}} \right) > 0 \,\middle|\, F_i \right]$$

$$\leq \prod_{j:\check{\boldsymbol{c}}(j)=1} \mathbb{E}_{\cdot|F_i,\boldsymbol{c}(i)=2} \exp\left[-\lambda \left(A_{ij} \log \frac{\hat{q}}{\hat{p}} + (1-A_{ij}) \log \frac{1-\hat{q}}{1-\hat{p}} \right) \right]. \quad \text{(S6.39)}$$

And for any $j \in [n]$, $P(A_{ij} = 1) = q$. Then,

$$\begin{aligned} & \mathbb{E}_{\cdot|F_i,\boldsymbol{c}(i)=2} \exp\left[-\lambda \left(A_{ij} \log \frac{\hat{q}}{\hat{p}} + (1-A_{ij}) \log \frac{1-\hat{q}}{1-\hat{p}} \right) \right] \\ = \; & q \left(\frac{\hat{q}}{\hat{p}} \right)^{-\lambda} + (1-q) \left(\frac{1-\hat{q}}{1-\hat{p}} \right)^{-\lambda} \\ = \; & p^{\lambda} q^{1-\lambda} + (1-p)^{\lambda} (1-q)^{1-\lambda} + p^{\lambda} q^{1-\lambda} \left[\left(\frac{p\hat{q}}{\hat{p}q} \right)^{-\lambda} - 1 \right] \\ & + (1-p)^{\lambda} (1-q)^{1-\lambda} \left[\left(\frac{(1-p)(1-\hat{q})}{(1-\hat{p})(1-q)} \right)^{-\lambda} - 1 \right]. \end{aligned}$$

Let $\lambda = 1 - t^*$. By simple manipulation, we have,

$$\begin{aligned} & \mathbb{E}_{\cdot|F_i,\boldsymbol{c}(i)=2} \exp\left[-(1-t^*) \left(A_{ij} \log \frac{\hat{q}}{\hat{p}} + (1-A_{ij}) \log \frac{1-\hat{q}}{1-\hat{p}} \right) \right] \\ \leq & \Big(q^{t^*} p^{1-t^*} + (1-q)^{t^*} (1-p)^{1-t^*} \Big) \big(1 + 2C(1-t^*)\eta_0 \delta \big) \\ \leq & (1 + C\eta_0\delta) \exp(-I_{t^*}), \end{aligned} \quad \text{(S6.40)}$$

where $C > 0$ is a large constant. Denote $\check{n}_1 = \check{n}_1(\check{\boldsymbol{c}}) = |\{j : \check{\boldsymbol{c}}(j) = 1\}|$. And then, we have,

$$\begin{aligned} \mathbb{P}_{\boldsymbol{c}(i)=2}\Big(\check{\boldsymbol{c}}(i) = 1 \text{ and } F_i \Big) \leq & (1 + C\eta_0\delta q)^{\check{n}_1} \exp(-\check{n}_1 I_{t^*}) \\ \leq & \exp\big(-(1-\eta)\beta n I_{t^*} \big), \end{aligned} \quad \text{(S6.41)}$$

for a sequence $\eta \to 0$. Since $\eta_0 = o(p-q)$, and $\eta_0 = o\left(q\frac{\log\log\frac{p}{q}}{\log\frac{p}{q}}\right)$ when $p \gg q$, we can get $(1+C\eta_0\delta)^{\check{n}_1} = \exp\left(\check{n}_1 \log(1+C\eta_0\delta)\right) = \exp\left(o(1)\check{n}_1 I_{t^*}\right)$. Thus, the last inequality in (S6.41) holds. And then,

$$\begin{aligned}
\mathbb{P}\big(\check{\boldsymbol{c}}(i) \neq \boldsymbol{c}(i)\big) &\leq \mathbb{P}\Big(\check{\boldsymbol{c}}(i) \neq \boldsymbol{c}(i), F_i\Big) + Cn^{-(1+C_0)} \\
&\leq \mathbb{P}_{\boldsymbol{c}(i)=1}\Big(\check{\boldsymbol{c}}(i) \neq 2, F_i\Big) + \mathbb{P}_{\boldsymbol{c}(i)=2}\Big(\check{\boldsymbol{c}}(i) \neq 1, F_i\Big) + Cn^{-(1+C_0)} \\
&\leq \exp\big(-(1-\eta)\beta n I_{t^*}\big) + Cn^{-(1+C_0)},
\end{aligned}$$

for a sequence $\eta \to 0$ since $\beta n I_{t^*} \to \infty$ as $n \to \infty$. The proof is then completed since C and η do not rely on i or $(\boldsymbol{P}, \boldsymbol{c})$.

□

Proof of Lemma 8. Firstly, we abbreviate $n_k(\boldsymbol{c})$ as n_k for each $k \in [2]$. Let $\boldsymbol{\Gamma} = \operatorname{diag}\left(\sqrt{n_1}, \sqrt{n_2}\right)$ and $\boldsymbol{Z} = (Z_{ik}) \in \{0,1\}^{n\times 2}$ with $Z_{ik} = 1$ if and only if $\boldsymbol{c}(i) = k$. Let $\boldsymbol{U}_0\boldsymbol{\Lambda}\boldsymbol{U}_0^\top$ be the eigen-decomposition of $\boldsymbol{\Gamma}\boldsymbol{P}\boldsymbol{\Gamma}$ and we can calculate $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2)$. And then $\boldsymbol{M}' = \boldsymbol{Z}\boldsymbol{P}\boldsymbol{Z}^\top = \boldsymbol{Z}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}\boldsymbol{P}\boldsymbol{\Gamma}\left(\boldsymbol{Z}\boldsymbol{\Gamma}^{-1}\right)^\top = \boldsymbol{Z}\boldsymbol{\Gamma}^{-1}\boldsymbol{U}_0\boldsymbol{\Lambda}\boldsymbol{U}_0^\top\left(\boldsymbol{Z}\boldsymbol{\Gamma}^{-1}\right)^\top$. Since $\left(\boldsymbol{Z}\boldsymbol{\Gamma}^{-1}\right)^\top \boldsymbol{Z}\boldsymbol{\Gamma}^{-1} = \operatorname{diag}(1,1)$, we have $\boldsymbol{M}' = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ where $\boldsymbol{U} = \boldsymbol{Z}\boldsymbol{\Gamma}^{-1}\boldsymbol{U}_0$. Hence, we can obtain that $\boldsymbol{U}_i = \boldsymbol{U}_j$ if and only if $\boldsymbol{c}(i) = \boldsymbol{c}(j)$ where $\boldsymbol{U}_i$ is the i-th row of $\boldsymbol{U}$, and we can calculate

$$U_{i1} = \frac{1}{\sqrt{n_1(\boldsymbol{c})}} \frac{\tilde{x}+z}{\sqrt{(\tilde{x}+z)^2 + y^2}}, \quad \text{if } \boldsymbol{c}(i) = 1,$$

and

$$U_{i2} = \frac{1}{\sqrt{n_2(\boldsymbol{c})}} \frac{y}{\sqrt{(\tilde{x}+z)^2+y^2}}, \quad \text{if } \boldsymbol{c}(i) = 2,$$

which completes the proof.

□

Bibliography

Ariu, K., A. Proutiere, and S.-Y. Yun (2023). Instance-optimal cluster recovery in the labeled stochastic block model. *arXiv*, 2306.12968.

Gao, C., Z. Ma, A. Zhang, and H. Zhou (2017). Achieving optimal misclassification proportion in stochastic block model. *Journal of Machine Learning Research 18*, 1–45.

Gao, C., Z. Ma, A. Y. Zhang, and H. H. Zhou (2018). Community detection in degree-corrected block models. *The Annals of Statistics 46*(5), 2153–2185.

Hu, J., H. Qin, T. Yan, and Y. Zhao (2020). Corrected bayesian information criterion for stochastic block models. *Journal of the American Statistical Association 115*(532), 1771–1783.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory.* New York, NY: Springer New York.

Lei, J. and A. Rinaldo (2015). Consistency of spectral clustering in stochastic block models. *Annals of Statistics 43*(1), 215–237.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics 20*, 53–65.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wang, J., J. Zhang, B. Liu, J. Zhu, and J. Guo (2020). Fast network community detection with profile-pseudo likelihood methods. *Journal of the American Statistical Association 118*, 1359–1372.

Wilson, J. D., J. Palowitch, S. Bhamidi, and A. B. Nobel (2017). Community extraction in multilayer networks with heterogeneous community structure. *Journal of machine learning research 18*, 5458–5506.

Yun, S.-Y. and A. Proutiere (2016). Optimal cluster recovery in the labeled stochastic block model. *Proceedings of the 30th International Conference on Neural Information Processing Systems 29*, 973–981.

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

Zhao, Y., E. Levina, and J. Zhu (2011). Community extraction for social networks. *Proceedings of the National Academy of Sciences 108*(18), 7321–7326.