Supplement to "Partial Recovery for Top-k Ranking: Optimality of MLE and Sub-Optimality of Spectral Method"

Pinhan Chen[1], Chao Gao[1], and Anderson Y. Zhang[2]

[1] *University of Chicago*
[2] *University of Pennsylvania*

December 27, 2021

This supplement includes all the technical proofs. In Appendix A, we first give proofs for all the results established in Section 4: Theorem 4.1, Theorem 4.2 and Theorem 4.3. After that, we prove Theorem 3.4 and Theorem 6.1 in Appendix B. We then include the proofs of Theorem 7.1 and Theorem 7.2 in Appendix C, the the proof of Lemma 8.5 in Appendix D, and the proofs of all the other technical lemmas in Appendix E. The count method is discussed in Appendix F.

A Analysis of the Spectral Method

We prove results for the spectral method in this section. This includes Theorem 4.1, Theorem 4.2 and Theorem 4.3. The proofs of Theorem 4.1 and Theorem 4.2 are given in Section A.1, and then we prove Theorem 4.3 in Section A.2.

A.1 Proofs of Theorem 4.1 and Theorem 4.2

The proof of Theorem 4.1 relies on a leave-one-out argument introduced by [3]. Without loss of generality, we consider $r_i^* = i$ so that $\theta^*_{r_i^*} = \theta^*_i$. Following [3], we define a transition matrix $P^{(m)}$ for each $m \in [n]$. For any $i \neq j$, $P^{(m)}_{ij} = P_{ij}$ if $i \neq m$ and $j \neq m$ and otherwise $P^{(m)}_{ij} = \frac{p}{d}\psi(\theta^*_i - \theta^*_j)$. For any $i \in [n]$, $P^{(m)}_{ii} = \sum_{j\in[n]\setminus\{i\}} P^{(m)}_{ij}$. Let $\pi^{(m)}$ be the stationary distribution of $P^{(m)}$. The following ℓ_2 norm bound has essentially been proved in [3].

Lemma A.1. *Under the setting of Theorem 4.1, there exists a constant $C > 0$ such that*

$$\max_{m\in[n]} \|\pi^{(m)} - \widehat{\pi}\| \leq C\frac{1}{n}\sqrt{\frac{\log n}{npL}},$$

$$\max_{m\in[n]} \|\pi^{(m)} - \pi^*\|_\infty \leq C\frac{1}{n}\sqrt{\frac{\log n}{npL}},$$

$$\max_{m\in[n]} \|\pi^{(m)} - \pi^*\| \le C\frac{1}{n}\sqrt{\frac{1}{pL}},$$

with probability at least $1 - O(n^{-4})$.

Proof. By Lemma 5.6 and Lemma 5.7 of [3], one can obtain $\|\pi^{(m)} - \widehat{\pi}\| \le C_1\sqrt{\frac{\log n}{npL}}\|\pi^*\|_\infty + \|\widehat{\pi} - \pi^*\|_\infty$ for some constant $C_1 > 0$ with probability at least $1 - O(n^{-5})$. Theorem 2.6 of [3] gives the bound $\|\widehat{\pi} - \pi^*\|_\infty \le C_2\sqrt{\frac{\log n}{npL}}\|\pi^*\|_\infty$ with probability at least $1 - O(n^{-5})$. A union bound argument together with the fact that $\|\pi^*\|_\infty \asymp n^{-1}$ leads to the first conclusion. The second conclusion is a consequence of triangle inequality. By Theorem 5.2 of [3], we have $\|\widehat{\pi} - \pi^*\| \le C_3\frac{1}{n}\sqrt{\frac{1}{pL}}$ with probability at least $1 - O(n^{-1})$. Thus, we obtain the last conclusion by applying triangle inequality again. □

We also need a lemma that relates the asymptotic variance of $\widehat{\pi}_i$ to the function $\overline{V}(\kappa)$.

Lemma A.2. *For any positive* $\kappa_1, \kappa_2 = O(1)$*, we have*

$$\begin{aligned}
&\min_{\substack{x_1,\dots,x_k\in[0,\kappa_1]\\ x_{k+1},\dots,x_n\in[0,\kappa_2]}} \frac{(\sum_{i=1}^k \psi(x_i) + \sum_{i=k+1}^n \psi(-x_i))^2}{\sum_{i=1}^k \psi'(x_i)(1+e^{x_i})^2 + \sum_{i=k+1}^n \psi'(x_i)(1+e^{-x_i})^2} \\
=\ & \frac{(k\psi(\kappa_1) + (n-k)\psi(-\kappa_2))^2}{k\psi'(\kappa_1)(1+e^{\kappa_1})^2 + (n-k)\psi'(\kappa_2)(1+e^{-\kappa_2})^2},
\end{aligned}$$

for n *that is sufficiently large.*

Proof. The problem is equivalent to the solution of the following: the optimum of the problem

$$\min_{\substack{x_1,\dots,x_k\in[1,M_1]\\ x_{k+1},\dots,x_n\in[1,M_2]}} \frac{(\sum_{i=1}^k \frac{2x_i}{1+x_i} + \sum_{i=k+1}^n \frac{2}{1+x_i})^2}{\sum_{i=1}^k x_i + \sum_{i=k+1}^n \frac{1}{x_i}} = \min_{\substack{x_1,\dots,x_k\in[1,M_1]\\ x_{k+1},\dots,x_n\in[1,M_2]}} f(x_1,\cdots,x_n)$$

is obtained at $x_1 = \dots = x_k = M_1, x_{k+1} = \dots = x_n = M_2$. We will show that for any given $x_{k+1}, \dots, x_n \in [1, M_2]$, the function is minimized at $x_1 = \dots = x_k = M_1$. Moreover, for any given $x_1, \dots, x_k$, the function is minimized at $x_{k+1} = \dots = x_n = M_2$. We only need to prove the former claim and the latter one can be proved similarly. Define

$$g(x_1,\cdots,x_k) = \frac{\left(\sum_{i=1}^k \frac{2x_i}{1+x_i} + \alpha\right)^2}{\sum_{i=1}^k x_i + \beta},$$

where $\alpha = \sum_{i=k+1}^n \frac{2}{1+x_i}, \beta = \sum_{i=k+1}^n \frac{1}{x_i}$. We first analyze the behavior of $g(x_1,\cdots,x_k)$ at each coordinate. By direct calculation, we have

$$\begin{aligned}
\frac{\partial \log g(x_1,\cdots,x_k)}{\partial x_1} &= \frac{4}{(1+x_1)^2(\sum_{i=1}^k \frac{2x_i}{1+x_i} + \alpha)} - \frac{1}{\sum_{i=1}^k x_i + \beta} \\
&= \frac{4(\sum_{i=1}^k x_i + \beta) - (1+x_1)^2(\sum_{i=1}^k \frac{2x_i}{1+x_i} + \alpha)}{(1+x_1)^2(\sum_{i=1}^k \frac{2x_i}{1+x_i} + \alpha)(\sum_{i=1}^k x_i + \beta)}.
\end{aligned}$$

The sign of the partial derivative is determined by its numerator

$$\begin{aligned}
& 4(\sum_{i=1}^{k} x_i + \beta) - (1+x_1)^2(\sum_{i=1}^{k} \frac{2x_i}{1+x_i} + \alpha) \\
= \; & -\left(\sum_{i=2}^{k} \frac{2x_i}{1+x_i} + \alpha + 2\right) x_1^2 - \left(\sum_{i=2}^{k} \frac{4x_i}{1+x_i} + 2\alpha - 2\right) x_1 \\
& +4(\sum_{i=2}^{k} x_i + \beta) - \left(\sum_{i=2}^{k} \frac{2x_i}{1+x_i} + \alpha\right),
\end{aligned}$$

which is a quadratic decreasing function of $x_1 \in [1, M_1]$. Therefore, $g(x_1, \cdots, x_k)$ is either monotone of $x_1 \in [1, M_1]$, or it is first increasing then decreasing. This implies that the optimum is achieved either at $x_1 = 1$ or $x_1 = M_1$. Since $g(x_1, \cdots, x_k)$ is symmetric, we therefore know that the optimizer must satisfy $(x_1, \cdots, x_k) \in \{1, M_1\}^k$. Using symmetry again, we can conclude that the value of $\min_{x_1, \cdots, x_k \in [1, M_1]} g(x_1, \cdots, x_k)$ is determined by the number of coordinates that take M_1. For $i \in [k]$, we define g_i to be the value of $g(x_1, \cdots, x_k)$ with $x_1 = \cdots = x_i = M_1$ and $x_{i+1} = \cdots = x_k = 1$. We now need to show g_i is nonincreasing in $i \in [k]$. Note that

$$\begin{aligned}
g_i \geq g_{i+1} \quad & \Longleftrightarrow \quad \frac{(i\frac{2M_1}{M_1+1} + k - i + \alpha)^2}{iM_1 + k - i + \beta} \geq \frac{(\frac{M_1-1}{M_1+1} + i\frac{2M_1}{M_1+1} + k - i + \alpha)^2}{M_1 - 1 + iM_1 + k - i + \beta} \\
& \Longleftrightarrow \quad \frac{M_1 - 1}{iM_1 + k - i + \beta} \geq \frac{(\frac{M_1-1}{M_1+1})^2}{(i\frac{2M_1}{M_1+1} + k - i + \alpha)^2} + \frac{2(\frac{M_1-1}{M_1+1})}{i\frac{2M_1}{M_1+1} + k - i + \alpha} \\
& \Longleftrightarrow \quad \frac{(M_1+1)^2}{iM_1 + k - i + \beta} - \frac{M_1 - 1}{(i\frac{2M_1}{M_1+1} + k - i + \alpha)^2} - \frac{2(M_1+1)}{i\frac{2M_1}{M_1+1} + k - i + \alpha} \geq 0 \\
& \Longleftrightarrow \quad \frac{(i\frac{M_1-1}{M_1+1} + k + \alpha)(M_1+1)^2}{i(M_1-1) + k + \beta} - \frac{M_1 - 1}{i\frac{M_1-1}{M_1+1} + k + \alpha} - 2(M_1+1) \geq 0 \\
& \Longleftrightarrow \quad \frac{i(M_1-1) + (k+\alpha)(M_1+1)}{i(M_1-1) + k + \beta} - \frac{M_1 - 1}{i(M_1-1) + (k+\alpha)(M_1+1)} - 2 \geq 0 \\
& \Longleftarrow \quad \frac{i(M_1-1) + (k+\beta)(M_1+1)}{i(M_1-1) + k + \beta} - \frac{M_1 - 1}{i(M_1-1) + (k+\beta)(M_1+1)} - 2 \geq 0 \qquad \text{(S1)} \\
& \Longleftrightarrow \quad \frac{-i(M_1-1) + (k+\beta)(M_1-1)}{i(M_1-1) + k + \beta} - \frac{M_1 - 1}{i(M_1-1) + (k+\beta)(M_1+1)} \geq 0 \\
& \Longleftarrow \quad \frac{-i + (k+\beta)}{i(M_1-1) + k + \beta} - \frac{1}{i(M_1-1) + (k+\beta)(M_1+1)} \geq 0 \\
& \Longleftrightarrow \quad (k+\beta)^2(M_1+1) \geq i(M_1-1) + i^2(M_1-1) + (2i+1)(k+\beta) \\
& \Longleftarrow \quad (k+\beta)^2(M_1+1) \geq (k-1)^2(M_1-1) + (k-1)(M_1-1) + (2k-1)(k+\beta) \\
& \Longleftrightarrow \quad k^2(M_1+1) + 2\beta(M_1+1)k + \beta^2(M_1+1) \geq k^2(M_1+1) + (-M_1 + 2\beta)k - \beta \\
& \Longleftrightarrow \quad (2\beta+1)M_1 k + \beta^2(M_1+1) + \beta \geq 0
\end{aligned}$$

where the last display is trivially true. We have used $\alpha \geq \beta$ for the step (S1). Therefore, $\min_{x_1, \cdots, x_k \in [1, M_1]} g(x_1, \cdots, x_k) = g_k$, and the proof is complete. □

Now we are ready to prove Theorem 4.1.

Proof of Theorem 4.1. When the error exponent is of constant order, the bound is also a constant, and the result already holds since $\mathsf{H}_k(\widehat{r}, r^*) \leq 1$. Therefore, we only need to consider the case when the error exponent tends to infinity. We first introduce some notation. Define

$$\eta = \frac{1}{2} - \frac{\overline{V}(\kappa)}{(1-\bar{\delta})\Delta^2 npL} \log \frac{n-k}{k}, \tag{S2}$$

where $\bar{\delta} = o(1)$ is chosen so that $\eta > 0$ is satisfied. The specific choice of $\bar{\delta}$ will be determined later in the proof. We will continue to use the notation $\bar{\Delta}_i$ that is defined in (55). Since the diverging exponent implies $\overline{\mathsf{SNR}} \to \infty$, we have $\min_{i\in[n]} \bar{\Delta}_i^2 Lnp \to \infty$ and $\max_{i\in[n]} \bar{\Delta}_i \to 0$.

Since $\widehat{\pi}$ is the stationary distribution of P, we have $\hat{\pi}^T P = \hat{\pi}^T$. This implies that for any $m \in [n]$, we have $\sum_{j=1}^n P_{jm}\widehat{\pi}_j = \widehat{\pi}_m$. We can equivalently write this identity as

$$\widehat{\pi}_m = \frac{\sum_{j\in[n]\setminus\{m\}} P_{jm}\widehat{\pi}_j}{1 - P_{mm}} = \frac{\sum_{j\in[n]\setminus\{m\}} A_{jm}\bar{y}_{mj}\widehat{\pi}_j}{\sum_{j\in[n]\setminus\{m\}} A_{jm}\bar{y}_{jm}}.$$

We approximate $\widehat{\pi}_m$ by

$$\bar{\pi}_m = \frac{\sum_{j\in[n]\setminus\{m\}} A_{jm}\bar{y}_{mj}\pi_j^*}{\sum_{j\in[n]\setminus\{m\}} A_{jm}\bar{y}_{jm}}. \tag{S3}$$

The approximation error can be bounded by

$$|\widehat{\pi}_m - \bar{\pi}_m| \leq \left| \frac{\sum_{j\in[n]\setminus\{m\}} A_{jm}\bar{y}_{mj}(\widehat{\pi}_j - \pi_j^{(m)})}{\sum_{j\in[n]\setminus\{m\}} A_{jm}\bar{y}_{jm}} \right| \tag{S4}$$

$$+ \left| \frac{\sum_{j\in[n]\setminus\{m\}} A_{jm}\bar{y}_{mj}(\pi_j^{(m)} - \pi_j^*)}{\sum_{j\in[n]\setminus\{m\}} A_{jm}\bar{y}_{jm}} \right|. \tag{S5}$$

The two terms (S4) and (S5) share a common denominator, which can be lower bounded by

$$\sum_{j\in[n]\setminus\{m\}} A_{jm}\bar{y}_{jm} \geq \sum_{j\in[n]\setminus\{m\}} A_{jm}\psi(\theta_j^* - \theta_m^*) - \left| \sum_{j\in[n]\setminus\{m\}} A_{jm}(\bar{y}_{jm} - \psi(\theta_j^* - \theta_m^*)) \right|. \tag{S6}$$

By Lemma 8.1 and Lemma 8.4, we have $\sum_{j\in[n]\setminus\{m\}} A_{jm}\bar{y}_{jm} \geq c_1 np$ for some constant $c_1 > 0$ with probability at least $1 - O(n^{-10})$. With this lower bound, we then bound (S4) as

$$\begin{aligned}
\left| \frac{\sum_{j\in[n]\setminus\{m\}} A_{jm}\bar{y}_{mj}(\widehat{\pi}_j - \pi_j^{(m)})}{\sum_{j\in[n]\setminus\{m\}} A_{jm}\bar{y}_{jm}} \right| &\leq \frac{\sqrt{\sum_{j\in[n]\setminus\{m\}} A_{1j}\bar{y}_{mj}^2} \|\widehat{\pi} - \pi^{(m)}\|}{c_1 np} \\
&\leq \frac{\sqrt{\sum_{j\in[n]\setminus\{m\}} A_{1j}} \|\widehat{\pi} - \pi^{(m)}\|}{c_1 np} \\
&\leq C_1 \frac{1}{n}\sqrt{\frac{\log n}{(np)^2 L}},
\end{aligned}$$

with probability at least $1-O(n^{-4})$. In the last inequality, we have used Lemma 8.1 and Lemma A.1. For (S5), we can bound it as

$$\begin{aligned}
&\left|\frac{\sum_{j\in[n]\backslash\{m\}} A_{jm}\bar{y}_{mj}(\pi_j^{(m)}-\pi_j^*)}{\sum_{j\in[n]\backslash\{m\}} A_{jm}\bar{y}_{jm}}\right| \\
\leq\ & \frac{\left|\sum_{j\in[n]\backslash\{m\}} A_{jm}(\bar{y}_{mj}-\psi(\theta_m^*-\theta_j^*))(\pi_j^{(m)}-\pi_j^*)\right|}{c_1 np} + \frac{p\left|\sum_{j\in[n]\backslash\{m\}} \psi(\theta_m^*-\theta_j^*)(\pi_j^{(m)}-\pi_j^*)\right|}{c_1 np} \\
&+\frac{\left|\sum_{j\in[n]\backslash\{m\}} (A_{jm}-p)\psi(\theta_m^*-\theta_j^*)(\pi_j^{(m)}-\pi_j^*)\right|}{c_1 np}.
\end{aligned}$$

We bound the three terms above separately. For the first term, we use Hoeffding's inequality (Lemma E.1), and get

$$\frac{\left|\sum_{j\in[n]\backslash\{m\}} A_{jm}(\bar{y}_{mj}-\psi(\theta_m^*-\theta_j^*))(\pi_j^{(m)}-\pi_j^*)\right|}{c_1 np} \leq C_2\frac{\sqrt{\frac{x}{L}\sum_{j\in[n]\backslash\{m\}} A_{jm}(\pi_j^{(m)}-\pi_j^*)^2}}{np}, \tag{S7}$$

with probability at least $1-e^{-x}$. By Lemma 8.1 and Lemma A.1, we have

$$\sqrt{\sum_{j\in[n]\backslash\{m\}} A_{jm}(\pi_j^{(m)}-\pi_j^*)^2} \leq \|\pi^{(m)}-\pi^*\|_\infty\sqrt{\sum_{j\in[n]\backslash\{m\}} A_{jm}} \leq C_3\frac{1}{n}\sqrt{\frac{\log n}{L}},$$

with probability at least $1-O(n^{-4})$. Taking $x=\bar{\Delta}_m^2 npL\sqrt{\frac{npL}{\log n}}$, we have

$$\frac{\left|\sum_{j\in[n]\backslash\{m\}} A_{jm}(\bar{y}_{mj}-\psi(\theta_m^*-\theta_j^*))(\pi_j^{(m)}-\pi_j^*)\right|}{c_1 np} \leq C_4\frac{1}{n}\bar{\Delta}_m\left(\frac{\log n}{Lnp}\right)^{1/4},$$

with probability at least $1-O(n^{-4})-\exp\left(-\bar{\Delta}_m^2 npL\sqrt{\frac{npL}{\log n}}\right)$. Next, for the second term, we apply Lemma A.1 and get

$$\frac{p\left|\sum_{j\in[n]\backslash\{m\}} \psi(\theta_m^*-\theta_j^*)(\pi_j^{(m)}-\pi_j^*)\right|}{c_1 np} \leq \frac{\|\pi^{(m)}-\pi^*\|}{c_1\sqrt{n}} \leq C_5\frac{1}{n}\sqrt{\frac{1}{npL}},$$

with probability at least $1-O(n^{-4})$. For the third term, we use Bernstein's inequality (Lemma E.2), and get

$$\frac{\left|\sum_{j\in[n]\backslash\{m\}} (A_{jm}-p)\psi(\theta_m^*-\theta_j^*)(\pi_j^{(m)}-\pi_j^*)\right|}{c_1 np} \leq C_6\frac{\sqrt{px}\|\pi^{(m)}-\pi^*\|}{np} + C_6\frac{x\|\pi^{(m)}-\pi^*\|_\infty}{np}, \tag{S8}$$

with probability at least $1-e^{-x}$. We choose $x = \min\left(\bar{\Delta}_m^2 Lnp\frac{np}{\log n}, 4\log n\right)$. Then, with the help of Lemma A.1, we have

$$\frac{\left|\sum_{j\in[n]\setminus\{m\}}(A_{jm}-p)\psi(\theta_m^*-\theta_j^*)(\pi_j^{(m)}-\pi_j^*)\right|}{c_1 np}$$
$$\leq C_7\frac{1}{n}\frac{1}{np\sqrt{L}}\sqrt{\min\left(\bar{\Delta}_m^2 Lnp\frac{np}{\log n}, \log n\right)} + C_7\frac{1}{n}\frac{1}{np}\sqrt{\frac{\log n}{npL}}\min\left(\bar{\Delta}_m^2 Lnp\frac{np}{\log n}, \log n\right), \tag{S9}$$

with probability at least $1-O(n^{-4})-\exp\left(-\bar{\Delta}_m^2 npL\frac{np}{\log n}\right)$.

To summarize, we have proved that

$$\frac{|\widehat{\pi}_m - \bar{\pi}_m|}{\pi_m^*} \leq \delta(1-e^{-\bar{\Delta}_m}), \tag{S10}$$

for some $\delta = o(1)$ with probability at least $1-O(n^{-4})-\exp\left(-\bar{\Delta}_m^2 npL\frac{np}{\log n}\right)-\exp\left(-\bar{\Delta}_m^2 npL\sqrt{\frac{npL}{\log n}}\right)$ under the assumption that $\bar{\Delta}_m = o(1)$, $npL\bar{\Delta}_m^2 \to \infty$ and $\frac{np}{\log n} \to \infty$.

Next, we note that by the definition of $\bar{\pi}_m$, we have

$$\bar{\pi}_m - \pi_m^* = \frac{\sum_{j\in[n]\setminus\{m\}} A_{jm}(\bar{y}_{mj} - \psi(\theta_m^*-\theta_j^*))(\pi_j^*+\pi_m^*)}{\sum_{j\in[n]\setminus\{m\}} A_{jm}\bar{y}_{jm}}. \tag{S11}$$

By Lemma 8.4 and the inequality (S6), the denominator of (S11) satisfies

$$\left|\frac{\sum_{j\in[n]\setminus\{m\}} A_{jm}\bar{y}_{jm}}{\sum_{j\in[n]\setminus\{m\}} A_{jm}\psi(\theta_j^*-\theta_m^*)} - 1\right| \leq \delta, \tag{S12}$$

for some $\delta = o(1)$ with probability at least $1-O(n^{-10})$. Note that we can choose the same δ to accommodate both bounds (S10) and (S12).

We will apply Lemma 3.1 with

$$t = \frac{e^{(1-\eta)\theta_k^*+\eta\theta_{k+1}^*}}{\sum_{j=1}^n e^{\theta_j^*}} \tag{S13}$$

to finish the proof. Recall the definition of η in (S2). For $i \le k$, we have

$$
\begin{aligned}
& \mathbb{P}\left(\widehat{\pi}_i \le \frac{e^{(1-\eta)\theta_k^* + \eta\theta_{k+1}^*}}{\sum_{j=1}^n e^{\theta_j^*}}\right) \\
= \; & \mathbb{P}\left(\frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} \le e^{(1-\eta)\theta_k^* + \eta\theta_{k+1}^* - \theta_i^*} - 1\right) \\
\le \; & \mathbb{P}\left(\frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} \le e^{-\bar{\Delta}_i} - 1\right) \\
\le \; & \mathbb{P}\left(\frac{\bar{\pi}_i - \pi_i^*}{\pi_i^*} \le -(1-\delta)(1-e^{-\bar{\Delta}_i})\right) + \mathbb{P}\left(\frac{|\bar{\pi}_i - \widehat{\pi}_i|}{\pi_i^*} > \delta(1-e^{-\bar{\Delta}_i})\right) \\
\le \; & \mathbb{P}\left(\frac{\sum_{j\in[n]\backslash\{i\}} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))(1+e^{\theta_j^* - \theta_i^*})}{\sum_{j\in[n]\backslash\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)} \le -(1-\delta)^2(1-e^{-\bar{\Delta}_i})\right) \\
& + \mathbb{P}\left(\frac{|\bar{\pi}_i - \widehat{\pi}_i|}{\pi_i^*} > \delta(1-e^{-\bar{\Delta}_i})\right) + \mathbb{P}\left(\left|\frac{\sum_{j\in[n]\backslash\{i\}} A_{ji}\bar{y}_{ji}}{\sum_{j\in[n]\backslash\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)} - 1\right| > \delta\right) \\
\le \; & \mathbb{P}\left(\frac{\sum_{j\in[n]\backslash\{i\}} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))(1+e^{\theta_j^* - \theta_i^*})}{\sum_{j\in[n]\backslash\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)} \le -(1-\delta)^2(1-e^{-\bar{\Delta}_i})\right) \\
& + O(n^{-4}) + \exp\left(-\bar{\Delta}_i^2 npL \frac{np}{\log n}\right) + \exp\left(-\bar{\Delta}_i^2 npL \sqrt{\frac{npL}{\log n}}\right),
\end{aligned}
\tag{S14}
$$

where the last inequality is by (S10) and (S12). Define the event

$$
\mathcal{A}_i = \left\{ A : \left| \frac{\sum_{j\in[n]\backslash\{i\}} A_{ij}\psi'(\theta_i^* - \theta_j^*)\left(1+e^{\theta_j^* - \theta_i^*}\right)^2}{p\sum_{j\in[n]\backslash\{i\}} \psi'(\theta_i^* - \theta_j^*)\left(1+e^{\theta_j^* - \theta_i^*}\right)^2} - 1 \right| \le \delta, \left| \frac{\sum_{j\in[n]\backslash\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)}{p\sum_{j\in[n]\backslash\{i\}} \psi(\theta_j^* - \theta_i^*)} - 1 \right| \le \delta \right\}.
\tag{S15}
$$

Then, by Bernstein's inequality, we have

$$
\begin{aligned}
& \mathbb{P}\left(\frac{\sum_{j\in[n]\backslash\{i\}} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))(1+e^{\theta_j^* - \theta_i^*})}{\sum_{j\in[n]\backslash\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)} \le -(1-\delta)^2(1-e^{-\bar{\Delta}_i})\right) \\
\le \; & \sup_{A\in\mathcal{A}_i} \mathbb{P}\left(\frac{\sum_{j\in[n]\backslash\{i\}} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))(1+e^{\theta_j^* - \theta_i^*})}{\sum_{j\in[n]\backslash\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)} \le -(1-\delta)^2(1-e^{-\bar{\Delta}_i}) \middle| A\right) \\
& + \mathbb{P}(A \in \mathcal{A}_i^c) \\
\le \; & \exp\left(-\frac{(1-o(1))Lp\bar{\Delta}_i^2\left(\sum_{j\in[n]\backslash\{i\}} \psi(\theta_j^* - \theta_i^*)\right)^2}{2\sum_{j\in[n]\backslash\{i\}} \psi'(\theta_i^* - \theta_j^*)\left(1+e^{\theta_j^* - \theta_i^*}\right)^2}\right) + O(n^{-4}) \qquad \text{(S16)} \\
\le \; & \exp\left(-\frac{(1-o(1))Lp(\bar{\Delta} + \theta_i^* - \theta_k^*)^2\left(\sum_{j\in[n]\backslash\{i\}} \psi(\theta_j^* - \theta_i^*)\right)^2}{2\sum_{j\in[n]\backslash\{i\}} \psi'(\theta_i^* - \theta_j^*)\left(1+e^{\theta_j^* - \theta_i^*}\right)^2}\right) + O(n^{-4}). \qquad \text{(S17)}
\end{aligned}
$$

The inequality (S17) is by the same argument that leads to (70) and (71). We use the notation $\bar{\Delta} = \min\left(\eta(\theta_k^* - \theta_{k+1}^*), \left(\frac{\log n}{np}\right)^{1/4}\right)$ in (S17). Define

$$h_i(t) = \frac{(\bar{\Delta} + t)^2 \left(\sum_{j\in[n]\backslash\{i\}} \psi(\theta_j^* - \theta_k^* - t)\right)^2}{\sum_{j\in[n]\backslash\{i\}} \psi'(t + \theta_k^* - \theta_j^*) \left(1 + e^{\theta_j^* - \theta_k^* - t}\right)^2}, \quad \text{for all } t \geq 0.$$

Though $h_i(t)$ is a complicated function, by the fact that $\bar{\Delta} = o(1)$ and $\max_{j,k} |\theta_j^* - \theta_k^*| \leq \kappa = O(1)$, one can directly analyze the derivative of $h_i(t)$ to conclude that there exists some small constant $c_2 > 0$ such that $h_i(t)$ is increasing on $[0, c_2]$. Moreover, there also exists a small constant $c_3 > 0$ such that $\min_{t\in[c_2,\kappa]} h_i(t) \geq c_3 n$. This implies

$$\begin{aligned}
&\frac{Lp(\bar{\Delta} + \theta_i^* - \theta_k^*)^2 \left(\sum_{j\in[n]\backslash\{i\}} \psi(\theta_j^* - \theta_i^*)\right)^2}{2\sum_{j\in[n]\backslash\{i\}} \psi'(\theta_i^* - \theta_j^*) \left(1 + e^{\theta_j^* - \theta_i^*}\right)^2} \\
\geq\ & \frac{Lp\bar{\Delta}^2 \left(\sum_{j\in[n]\backslash\{i\}} \psi(\theta_j^* - \theta_k^*)\right)^2}{2\sum_{j\in[n]\backslash\{i\}} \psi'(\theta_k^* - \theta_j^*) \left(1 + e^{\theta_j^* - \theta_k^*}\right)^2} \wedge \frac{c_3 npL}{2} \\
=\ & \frac{Lp\bar{\Delta}^2 \left(\sum_{j\in[n]\backslash\{i\}} \psi(\theta_j^* - \theta_k^*)\right)^2}{2\sum_{j\in[n]\backslash\{i\}} \psi'(\theta_k^* - \theta_j^*) \left(1 + e^{\theta_j^* - \theta_k^*}\right)^2},
\end{aligned}$$

where the last inequality is due to the fact that $\bar{\Delta} = o(1)$. We further bound the above exponent by

$$\begin{aligned}
&\frac{Lp\bar{\Delta}^2 \left(\sum_{j\in[n]\backslash\{i\}} \psi(\theta_j^* - \theta_k^*)\right)^2}{2\sum_{j\in[n]\backslash\{i\}} \psi'(\theta_k^* - \theta_j^*) \left(1 + e^{\theta_j^* - \theta_k^*}\right)^2} \\
=\ & (1 - o(1)) \frac{Lp\bar{\Delta}^2 \left(\sum_{j=1}^n \psi(\theta_j^* - \theta_k^*)\right)^2}{2\sum_{j=1}^n \psi'(\theta_k^* - \theta_j^*) \left(1 + e^{\theta_j^* - \theta_k^*}\right)^2} \\
\geq\ & (1 - o(1)) \frac{Lp\bar{\Delta}^2}{2} \min_{\substack{\kappa_1+\kappa_2\leq\kappa \\ \kappa_1,\kappa_2\geq 0}} \min_{\substack{x_1,\cdots,x_k\in[0,\kappa_1] \\ x_{k+1},\cdots,x_n\in[0,\kappa_2]}} \frac{\left(\sum_{j=1}^k \psi(x_j) + \sum_{j=k+1}^n \psi(-x_j)\right)^2}{\sum_{j=1}^k \psi'(x_j)(1 + e^{x_j})^2 + \sum_{j=k+1}^n \psi'(x_j)(1 + e^{-x_j})^2} \\
=\ & (1 - o(1)) \frac{Lp\bar{\Delta}^2}{2} \min_{\substack{\kappa_1+\kappa_2\leq\kappa \\ \kappa_1,\kappa_2\geq 0}} \frac{(k\psi(\kappa_1) + (n-k)\psi(-\kappa_2))^2}{k\psi'(\kappa_1)(1 + e^{\kappa_1})^2 + (n-k)\psi'(\kappa_2)(1 + e^{-\kappa_2})^2} \qquad \text{(S18)} \\
=\ & (1 - o(1)) \frac{Lpn\bar{\Delta}^2}{2\overline{V}(\kappa)}.
\end{aligned}$$

The equality (S18) is due to Lemma A.2. With the above analysis of the error exponent, we can further bound (S17) as

$$
\begin{aligned}
& \exp\left(-\frac{1-o(1)}{2}Lp\min\left(\eta^2\Delta^2, \sqrt{\frac{\log n}{np}}\right)\frac{n}{\overline{V}(\kappa)}\right) + O(n^{-4}) \\
\leq\; & \exp\left(-\frac{(1-o(1))\eta^2\Delta^2 npL}{2\overline{V}(\kappa)}\right) + O(n^{-4}).
\end{aligned}
$$

The last inequality holds because when $\min\left(\eta^2\Delta^2, \sqrt{\frac{\log n}{np}}\right) = \sqrt{\frac{\log n}{np}}$, the first term becomes $\exp\left(-\frac{(1-o(1))L\sqrt{np\log n}}{2\overline{V}(\kappa)}\right)$, which can be absorbed by $O(n^{-4})$. Since $\exp\left(-\bar{\Delta}_i^2 npL\frac{np}{\log n}\right) + \exp\left(-\bar{\Delta}_i^2 npL\sqrt{\frac{npL}{\log n}}\right) \leq \exp\left(-\frac{(1-o(1))\eta^2\Delta^2 npL}{2\overline{V}(\kappa)}\right) + O(n^{-4})$, we have

$$
\mathbb{P}\left(\widehat{\pi}_i \leq \frac{e^{(1-\eta)\theta_k^*+\eta\theta_{k+1}^*}}{\sum_{j=1}^n e^{\theta_j^*}}\right) \leq \exp\left(-\frac{(1-\delta_1)\eta^2\Delta^2 npL}{2\overline{V}(\kappa)}\right) + O(n^{-4}), \tag{S19}
$$

with some $\delta_1 = o(1)$ for all $i \leq k$. With a similar argument, we also have

$$
\mathbb{P}\left(\widehat{\pi}_i \geq \frac{e^{(1-\eta)\theta_k^*+\eta\theta_{k+1}^*}}{\sum_{j=1}^n e^{\theta_j^*}}\right) \leq \exp\left(-\frac{(1-\delta_1)(1-\eta)^2\Delta^2 npL}{2\overline{V}(\kappa)}\right) + O(n^{-4}), \tag{S20}
$$

for all all $i \geq k+1$. It can be checked that the δ_1 above can be set independent of the $\bar{\delta}$ in the definition of η. Now we choose η as in (S2) with $\bar{\delta} = \delta_1$. By Lemma 3.1, we have

$$
\begin{aligned}
\mathbb{E}\mathsf{H}_k(\widehat{r}, r^*) \;&\leq\; \exp\left(-\frac{(1-\bar{\delta})\eta^2\Delta^2 npL}{2\overline{V}(\kappa)}\right) + \frac{n-k}{k}\exp\left(-\frac{(1-\bar{\delta})(1-\eta)^2\Delta^2 npL}{2\overline{V}(\kappa)}\right) + O(n^{-4}) \\
&\leq\; 2\exp\left(-\frac{1}{2}\left(\frac{\sqrt{(1-\bar{\delta})\overline{\mathsf{SNR}}}}{2} - \frac{1}{\sqrt{(1-\bar{\delta})\overline{\mathsf{SNR}}}}\log\frac{n-k}{k}\right)^2\right) + O(n^{-4}).
\end{aligned}
$$

By Markov's inequality, the above bound implies

$$
\mathsf{H}_k(\widehat{r}, r^*) \leq \exp\left(-\frac{1}{2}\left(\frac{\sqrt{(1-\delta')\overline{\mathsf{SNR}}}}{2} - \frac{1}{\sqrt{(1-\delta')\overline{\mathsf{SNR}}}}\log\frac{n-k}{k}\right)^2\right) + O(n^{-3}),
$$

for some $\delta' = o(1)$ with high probability. One can take, for example,

$$
\delta' = \bar{\delta} + \frac{1}{\frac{\sqrt{(1-\bar{\delta})\overline{\mathsf{SNR}}}}{2} - \frac{1}{\sqrt{(1-\bar{\delta})\overline{\mathsf{SNR}}}}\log\frac{n-k}{k}}.
$$

When $O(n^{-3})$ dominates the bound, we have $\mathsf{H}_k(\widehat{r}, r^*) = O(n^{-3})$, which implies $\mathsf{H}_k(\widehat{r}, r^*) = 0$ since $\mathsf{H}_k(\widehat{r}, r^*) \in \{0, (2k)^{-1}, 2(2k)^{-1}, 3(2k)^{-1}, \cdots, 1\}$. Therefore, we always have

$$
\mathsf{H}_k(\widehat{r}, r^*) \leq 2\exp\left(-\frac{1}{2}\left(\frac{\sqrt{(1-\delta')\overline{\mathsf{SNR}}}}{2} - \frac{1}{\sqrt{(1-\delta')\overline{\mathsf{SNR}}}}\log\frac{n-k}{k}\right)^2\right),
$$

with high probability with some $\delta' = o(1)$. The proof is complete. □

Proof of Theorem 4.2. The proof is the same as that of Theorem 3.3. □

A.2 Proof of Theorem 4.3

To prove Theorem 4.3, we need two additional lemmas. The first lemma can be viewed as a reverse version of the inequality in Lemma 3.1.

Lemma A.3. *Suppose $\widehat{r}$ is a rank vector induced by $\widehat{\theta}$, we then have*

$$\mathsf{H}_k(\widehat{r}, r^*) \geq \frac{1}{k} \max_{t\in\mathbb{R}} \min \left(\sum_{i:r_i^*\leq k} \mathbb{I}\left\{\widehat{\theta}_i < t\right\}, \sum_{i:r_i^*>k} \mathbb{I}\left\{\widehat{\theta}_i > t\right\} \right).$$

The inequality holds for any $r^* \in \mathfrak{S}_n$.

Proof. Following the proof of Lemma 3.1, we have

$$\begin{aligned} 2k\mathsf{H}_k(\widehat{r}, r^*) &= 2\max\left(\sum_{i=1}^{k} \mathbb{I}\{\widehat{r}_i > k\}, \sum_{i=k+1}^{n} \mathbb{I}\{\widehat{r}_i \leq k\} \right) \\ &\geq 2\max\left(\sum_{i=1}^{k} \mathbb{I}\left\{\widehat{\theta}_i < \widehat{\theta}_{(k)}\right\}, \sum_{i=k+1}^{n} \mathbb{I}\left\{\widehat{\theta}_i > \widehat{\theta}_{(k+1)}\right\} \right) \\ &\geq 2\min_t \max\left(\sum_{i=1}^{k} \mathbb{I}\left\{\widehat{\theta}_i < t\right\}, \sum_{i=k+1}^{n} \mathbb{I}\left\{\widehat{\theta}_i > t\right\} \right) \qquad \text{(S21)} \\ &= 2\max_t \min\left(\sum_{i=1}^{k} \mathbb{I}\left\{\widehat{\theta}_i < t\right\}, \sum_{i=k+1}^{n} \mathbb{I}\left\{\widehat{\theta}_i > t\right\} \right). \qquad \text{(S22)} \end{aligned}$$

where (S21) and (S22) follow the same argument that leads to (S159) and (S160). □

Proof of Theorem 4.3. We first note that condition (20) necessarily implies $\Delta = o(1)$. Throughout the proof, we assume $\kappa = \Omega(1)$ and there exists some $\delta_1 = o(1)$ such that

$$\frac{\sqrt{(1+\delta_1)\overline{\mathsf{SNR}}}}{2} - \frac{1}{\sqrt{(1+\delta_1)\overline{\mathsf{SNR}}}} \log \frac{n-k}{k} \to \infty. \qquad \text{(S23)}$$

The case with $\kappa = o(1)$ or $\overline{\mathsf{SNR}}$ not satisfying (S23) will be addressed at the end of the proof.

Choose $\kappa_1, \kappa_2 \geq 0$ such that we have both $\kappa_1 + \kappa_2 \leq \kappa$ and

$$\frac{k\psi'(\kappa_1)(1+e^{\kappa_1})^2 + (n-k)\psi'(\kappa_2)(1+e^{-\kappa_2})^2}{(k\psi(\kappa_1) + (n-k)\psi(-\kappa_2))^2/n} = \overline{V}(\kappa).$$

Let $\rho = o(1)$ be a vanishing number that will be specified later. Since $k \to \infty$ and $\kappa = \Omega(1)$, one can easily check that $\kappa_2 = \Omega(1)$. Define $\theta_i^* = \kappa_1$ for all $1 \leq i \leq k - \rho k$, $\theta_i^* = 0$ for $k - \rho k < i \leq k$, $\theta_i^* = -\Delta$ for $k < i \leq k + \rho(n-k)$ and $\theta_i^* = -\kappa_2$ for $k + \rho(n-k) < i \leq n$. For the simplicity of proof, we choose ρ so that both ρk and $\rho(n-k)$ are integers. Define r^* to be $r_i^* = i, \forall i \in [n]$. Then we have

$$\sup_{\substack{r\in\mathfrak{S}_n \\ \theta\in\Theta(k,\Delta,\kappa)}} \mathbb{E}_{(\theta,r)}\mathsf{H}_k(\widehat{r}, r) \geq \mathbb{E}_{(\theta^*,r^*)}\mathsf{H}_k(\widehat{r}, r^*).$$

We will utilize several results established in the proof of Theorem 4.1. Define

$$\eta = \frac{1}{2} - \frac{\overline{V}(\kappa)}{(1+\bar{\delta})\Delta^2 npL} \log \frac{n-k}{k}, \tag{S24}$$

for $\bar{\delta} = o(1)$. The specific choice of $\bar{\delta}$ will be specified later in the proof. Also define $t = \frac{e^{(1-\eta)\theta_k^* + \eta\theta_{k+1}^*}}{\sum_{j=1}^n e^{\theta_j^*}} = \frac{e^{-\eta\Delta}}{\sum_{j=1}^n e^{\theta_j^*}}$. Then, by Lemma A.3, we have

$$\begin{aligned}
\mathsf{H}_k(\widehat{r}, r^*) &\geq \frac{1}{k} \min\left(\sum_{i=1}^{k} \mathbb{I}\{\widehat{\pi}_i < t\}, \sum_{i=k+1}^{n} \mathbb{I}\{\widehat{\pi}_i > t\} \right) \\
&\geq \frac{1}{k} \min\left(\sum_{k-\rho k < i \leq k} \mathbb{I}\{\widehat{\pi}_i < t\}, \sum_{k < i \leq k+\rho(n-k)} \mathbb{I}\{\widehat{\pi}_i > t\} \right).
\end{aligned}$$

For any $\delta > 0$, define the function $\phi(\delta) = \frac{\sqrt{(1+\delta)\overline{\mathsf{SNR}}}}{2} - \frac{1}{\sqrt{(1+\delta)\overline{\mathsf{SNR}}}} \log \frac{n-k}{k}$. It suffices to show there exists some constant $C > 0$ such that

$$\mathbb{P}_{(\theta^*, r^*)}\left(\sum_{k-\rho k < i \leq k} \mathbb{I}\{\widehat{\pi}_i < t\} \geq Ck \exp\left(-\frac{\phi(\bar{\delta})^2}{2} \right) \right) \geq 1 - o(1), \tag{S25}$$

$$\text{and } \mathbb{P}_{(\theta^*, r^*)}\left(\sum_{k < i \leq k+\rho(n-k)} \mathbb{I}\{\widehat{\pi}_i > t\} \geq Ck \exp\left(-\frac{\phi(\bar{\delta})^2}{2} \right) \right) \geq 1 - o(1). \tag{S26}$$

Suppose both inequalities hold, we have

$$\mathbb{P}_{(\theta^*, r^*)}\left(\mathsf{H}_k(\widehat{r}, r^*) > 0\right) \geq 1 - o(1).$$

By Markov's inequality, we also have

$$\mathbb{E}_{(\theta^*, r^*)}\mathsf{H}_k(\widehat{r}, r^*) \geq C\exp\left(-\frac{\phi(\bar{\delta})^2}{2}\right) \mathbb{P}_{(\theta^*, r^*)}\left(\mathsf{H}_k(\widehat{r}, r^*) \geq C\exp\left(-\frac{\phi(\bar{\delta})^2}{2}\right)\right) \geq \frac{C}{2}\exp\left(-\frac{\phi(\bar{\delta})^2}{2}\right).$$

Therefore, we obtain the desired conclusions.

In the rest of the proof, we are going to establish (S25). Recall the definition of $\bar{\pi}$ in (S3). For any $k - \rho k < i \leq k$, define the event $\mathcal{F}$ as

$$\mathcal{F}_i = \left\{ \frac{|\widehat{\pi}_i - \bar{\pi}_i|}{\pi_i^*} \leq \delta_0(1 - e^{-\eta\Delta}) \text{ and } \left| \frac{\sum_{j \in [n]\backslash\{i\}} A_{ji}\bar{y}_{ji}}{\sum_{j \in [n]\backslash\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)} - 1 \right| \leq \delta_0 \right\}.$$

Using a similar argument that leads to (S10) and (S12), we can show that there exists some $\delta_0 = o(1)$ not dependent on $\bar{\delta}$, such that

$$\mathbb{P}_{(\theta^*, r^*)}(\mathcal{F}_i) \geq 1 - \left(O(n^{-4}) + \exp\left(-\eta^2\Delta^2 npL\frac{np}{\log n} \right) + \exp\left(-\eta^2\Delta^2 npL\sqrt{\frac{npL}{\log n}} \right) \right). \tag{S27}$$

Suppose $\mathcal{F}_i$ holds, we then have

$$\begin{aligned}
\mathbb{I}\left\{\widehat{\pi}_i < t\right\} &= \mathbb{I}\left\{\widehat{\pi}_i < \frac{e^{(1-\eta)\theta_k^* + \eta\theta_{k+1}^*}}{\sum_{j=1}^n e^{\theta_j^*}}\right\} \\
&= \mathbb{I}\left\{\frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} \le e^{(1-\eta)\theta_k^* + \eta\theta_{k+1}^* - \theta_i^*} - 1\right\} \\
&= \mathbb{I}\left\{\frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} \le e^{-\eta\Delta} - 1\right\} \\
&\ge \mathbb{I}\left\{\frac{\bar{\pi}_i - \pi_i^*}{\pi_i^*} \le -(1+\delta_0)(1 - e^{-\eta\Delta})\right\} \\
&\ge \mathbb{I}\left\{\frac{\sum_{j\in[n]\setminus\{i\}} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))(1 + e^{\theta_j^* - \theta_i^*})}{\sum_{j\in[n]\setminus\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)} \le -(1+\delta_0)^2(1 - e^{-\eta\Delta})\right\} \\
&\ge \mathbb{I}\left\{\frac{\sum_{j\in[n]\setminus\{i\}} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))(1 + e^{\theta_j^* - \theta_i^*})}{\sum_{j\in[n]\setminus\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)} \le -(1+\delta_0)^2\eta\Delta\right\}. \qquad \text{(S28)}
\end{aligned}$$

We use the notation L_i for the indicator function on the right hand side of (S28). In other words, we have shown that

$$\begin{aligned}
\sum_{k-\rho k < i \le k} \mathbb{I}\left\{\widehat{\pi}_i < t\right\} &\ge \sum_{k-\rho k < i \le k} L_i \mathbb{I}_{\mathcal{F}_i} \\
&\ge \sum_{k-\rho k < i \le k} L_i - \sum_{k-\rho k < i \le k} \mathbb{I}_{\mathcal{F}_i^c}.
\end{aligned}$$

By (S27), we have

$$\mathbb{E}\left(\sum_{k-\rho k < i \le k} \mathbb{I}_{\mathcal{F}_i^c}\right) \le O(n^{-3}) + \rho k \exp\left(-\eta^2\Delta^2 npL\frac{np}{\log n}\right) + \rho k \exp\left(-\eta^2\Delta^2 npL\sqrt{\frac{npL}{\log n}}\right).$$

Since the above bounds is of smaller order than $k\exp\left(-\frac{\eta^2\Delta^2 npL}{2\overline{V}(\kappa)}\left(\frac{np}{\log n}\right)^{1/4}\right)$, we can use Markov's inequality and obtain

$$\mathbb{P}_{(\theta^*, r^*)}\left(\sum_{k-\rho k < i \le k} \mathbb{I}_{\mathcal{F}_i^c} \le k\exp\left(-\frac{\eta^2\Delta^2 npL}{2\overline{V}(\kappa)}\left(\frac{np}{\log n}\right)^{1/4}\right)\right) \ge 1 - o(1). \qquad \text{(S29)}$$

To lower bound $\sum_{k-\rho k<i\leq k} L_i$, we define

$$\mathcal{A} = \left\{ A : \forall k-\rho k < i \leq k, \left| \frac{\sum_{j\in[n]\setminus\{i\}} A_{ij}\psi'(\theta_i^*-\theta_j^*)\left(1+e^{\theta_j^*-\theta_i^*}\right)^2}{p\sum_{j\in[n]\setminus\{i\}} \psi'(\theta_i^*-\theta_j^*)\left(1+e^{\theta_j^*-\theta_i^*}\right)^2} - 1 \right| \leq \delta_0, \right. \tag{S30}$$

$$\left| \frac{\sum_{j\in[n]\setminus\{i\}} A_{ji}\psi(\theta_j^*-\theta_i^*)}{p\sum_{j\in[n]\setminus\{i\}} \psi(\theta_j^*-\theta_i^*)} - 1 \right| \leq \delta_0, \tag{S31}$$

$$\left. \left| \sum_{k-\rho k<j<k} A_{ji}\psi'(\theta_i^*-\theta_j^*)(1+e^{\theta_j^*-\theta_i^*})^2 \right| \leq 2\rho kp + 10\log n \right\}. \tag{S32}$$

By Bernstein's inequality and union bound, we have $\mathbb{P}(A\in\mathcal{A}) \geq 1-O(n^{-3})$. From now on, we use the notation $\mathbb{P}_A$ for the conditional probability $\mathbb{P}_{(\theta^*,r^*)}(\cdot|A)$ given A. For any $s>0$,

$$\mathbb{P}_{(\theta^*,r^*)}\left(\sum_{k-\rho k<i\leq k} L_i \geq s\right) \geq \mathbb{P}(A\in\mathcal{A}) \inf_{A\in\mathcal{A}} \mathbb{P}_A\left(\sum_{k-\rho k<i\leq k} L_i \geq s\right). \tag{S33}$$

To study $\mathbb{P}_A\left(\sum_{k-\rho k<i\leq k} L_i \geq s\right)$, we define the set $S=\{i\in[n]: i\leq k-\rho k \text{ or } i>k\}$. Note that for each $k-\rho k<i\leq k$, we have $L_i \geq L_{i,1}-L_{i,2}-L_{i,3}$, where

$$L_{i,1} = \mathbb{I}\left\{ \frac{\sum_{j\in S} A_{ji}(\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*))(1+e^{\theta_j^*-\theta_i^*})}{\sum_{j\in[n]\setminus\{i\}} A_{ji}\psi(\theta_j^*-\theta_i^*)} \leq -(1+2\delta')(1+\delta_0)^2\eta\Delta \right\}$$

$$L_{i,2} = \mathbb{I}\left\{ \frac{\sum_{k-\rho k<j<i} A_{ji}(\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*))(1+e^{\theta_j^*-\theta_i^*})}{\sum_{j\in[n]\setminus\{i\}} A_{ji}\psi(\theta_j^*-\theta_i^*)} \geq \delta'(1+\delta_0)^2\eta\Delta \right\}$$

$$L_{i,3} = \mathbb{I}\left\{ \frac{\sum_{i<j\leq k} A_{ji}(\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*))(1+e^{\theta_j^*-\theta_i^*})}{\sum_{j\in[n]\setminus\{i\}} A_{ji}\psi(\theta_j^*-\theta_i^*)} \geq \delta'(1+\delta_0)^2\eta\Delta \right\},$$

for some $\delta'=o(1)$ whose value will be determined later. We are going to control each term separately.

(1). Analysis of $L_{i,1}$. Note that conditional on A, $\{L_{i,1}\}_{k-\rho k<i\leq k}$ are all independent Bernoulli random variables. We have $L_{i,1}\sim \text{Bernoulli}(p_i)$, where $p_i=\mathbb{E}_{(\theta^*,r^*)}(L_{i,1}|A)$. By Chebyshev's inequality, we have

$$\mathbb{P}_A\left(\sum_{k-\rho k<i\leq k} L_{i,1} \geq \frac{1}{2}\sum_{k-\rho k<i\leq k} p_i\right) \geq 1-\frac{4}{\sum_{k-\rho k<i\leq k} p_i}.$$

By Lemma A.4 stated and proved at the end of the section, we can lower bound each p_i by

$$p_i = \mathbb{P}_A\left(\frac{\sum_{j\in S} A_{ji}(\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*))(1+e^{\theta_j^*-\theta_i^*})}{\sum_{j\in[n]\setminus\{i\}} A_{ji}\psi(\theta_j^*-\theta_i^*)} \leq -(1+2\delta')(1+\delta_0)^2\eta\Delta \right)$$

$$\geq C_1\exp\left(-\frac{(1+\delta_2)\eta^2\Delta^2 npL}{2\overline{V}(\kappa)} - C_1'\eta\sqrt{\frac{\Delta^2 npL}{\overline{V}(\kappa)}}\right),$$

for some constants $C_1, C_1' > 0$ and some $\delta_2 = o(1)$ that are not dependent on η. By (S23), there exists some $\delta_3 = o(1)$ such that

$$\sum_{k-\rho k < i \leq k} p_i \geq C_1 k \exp\left(-\frac{(1+\delta_3)\eta^2\Delta^2 npL}{2\overline{V}(\kappa)}\right). \tag{S34}$$

To obtain (S138), we need to set ρ that tends to zero sufficiently slow so that it can be absorbed into the exponent. Note that condition (20) is equivalent to $\frac{(1+\epsilon)\overline{\mathsf{SNR}}}{2}\left(\frac{1}{2} - \frac{1}{(1+\epsilon)\overline{\mathsf{SNR}}}\log\frac{n-k}{k}\right)^2 < \log k$. Since ϵ is a constant, it implies

$$\frac{\overline{\mathsf{SNR}}}{2}\left(\frac{1}{2} - \frac{1}{(1+\bar{\delta})\overline{\mathsf{SNR}}}\log\frac{n-k}{k}\right)^2 < (1-\epsilon')^{-1}\log k,$$

for some constant $\epsilon' > 0$. As a result, under the condition that $k \to \infty$, we have

$$\sum_{k-\rho k < i \leq k} p_i \geq \sum_{k-\rho k < i \leq k} C_1 \exp\left(-(1+\delta_3)(1-\epsilon')\log k\right) \geq k^{\frac{\epsilon'}{2}} \to \infty.$$

Hence, we have proved

$$\inf_{A\in\mathcal{A}} \mathbb{P}_A\left(\sum_{k-\rho k < i \leq k} L_{i,1} \geq \frac{1}{2}C_1 k \exp\left(-\frac{(1+\delta_2)\eta^2\Delta^2 npL}{2\overline{V}(\kappa)} - C_1'\eta\sqrt{\frac{\Delta^2 npL}{\overline{V}(\kappa)}}\right)\right) \geq 1 - o(1).$$

(2). Analysis of $L_{i,2}$. By (S133)-(S135) and Bernstein's inequality, we can bound $\mathbb{E}(L_{i,2}|A)$ by

$$\exp\left(-\frac{\left(\delta'(1+\delta_0)^2\eta\Delta L\sum_{j\in[n]\setminus\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)\right)^2}{2\left(L\sum_{k-\rho k<j<i} A_{ji}\psi'(\theta_i^* - \theta_j^*)(1+e^{\theta_j^*-\theta_i^*})^2 + \frac{1}{3}\delta'(1+\delta_0)^2\eta\Delta L\sum_{j\in[n]\setminus\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)\right)}\right)$$

$$\leq \exp\left(-\frac{\left(\delta'(1+\delta_0)^2\eta\Delta L\sum_{j\in[n]\setminus\{i\}} p\psi(\theta_j^* - \theta_i^*)\right)^2}{4\left(2L\rho kp + 10\log n + \frac{1}{3}\delta'(1+\delta_0)^2\eta\Delta L\sum_{j\in[n]\setminus\{i\}} p\psi(\theta_j^* - \theta_i^*)\right)}\right).$$

Now we set $\delta' = \max\{\rho^{\frac{1}{2}}, \Delta^{\frac{4}{3}}, \left(\frac{\log n}{np}\right)^{\frac{1}{2}}\}$. Then, there exists some constant $C_2, C_3 > 0$ such that

$$\mathbb{E}(L_{i,2}|A) \leq \exp\left(-C_2\rho^{-\frac{1}{2}} npL\eta^2\Delta^2\right) \leq \exp\left(-C_3\rho^{-1/2}\frac{\eta^2\Delta^2 npL}{2\overline{V}(\kappa)}\right).$$

Then,

$$\mathbb{E}\left(\sum_{k-\rho k < i \leq k} L_{i,2}\middle| A\right) \leq \rho k \exp\left(-C_3\rho^{-1/2}\frac{\eta^2\Delta^2 npL}{2\overline{V}(\kappa)}\right).$$

By Markov inequality, we have

$$\inf_{A\in\mathcal{A}} \mathbb{P}_A\left(\sum_{k-\rho k < i \leq k} L_{i,2} \geq \rho k \exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\eta^2\Delta^2 npL}{2\overline{V}(\kappa)}\right)\right) \leq \exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\eta^2\Delta^2 npL}{2\overline{V}(\kappa)}\right). \tag{S35}$$

(3). Analysis of $L_{i,3}$. By a similar argument, we also have

$$\inf_{A\in\mathcal{A}} \mathbb{P}_A\left(\sum_{k-\rho k<i\leq k} L_{i,3} \geq \rho k \exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\eta^2\Delta^2 npL}{2\overline{V}(\kappa)}\right)\right) \leq \exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\eta^2\Delta^2 npL}{2\overline{V}(\kappa)}\right). \tag{S36}$$

Now we can combine the above analyses of $L_{i,1}$, $L_{i,2}$ and $L_{i,3}$. Since $\rho = o(1)$, the bounds (S139) and (S140) are of smaller order than (S138). We have

$$\inf_{A\in\mathcal{A}} \mathbb{P}_A\left(\sum_{k-\rho k<i\leq k} L_i \geq C_4 k \exp\left(-\frac{(1+\delta_2)\eta^2\Delta^2 npL}{2\overline{V}(\kappa)} - C_1'\eta\sqrt{\frac{\Delta^2 npL}{\overline{V}(\kappa)}}\right)\right) \geq 1 - o(1), \tag{S37}$$

for some constant $C_4 > 0$. Then (S132) and (S136) lead to

$$\mathbb{P}_{(\theta^*,r^*)}\left(\sum_{k-\rho k<i\leq k} \mathbb{I}\left\{\widehat{\pi}_i < t\right\} \geq C_4 k \exp\left(-\frac{(1+\delta_2)\eta^2\Delta^2 npL}{2\overline{V}(\kappa)} - C_1'\eta\sqrt{\frac{\Delta^2 npL}{\overline{V}(\kappa)}}\right)\right) \geq 1 - o(1). \tag{S38}$$

We are going to show it leads to (S25) by selecting an appropriate $\bar{\delta}$ as follows. We write $\eta = \eta_{\bar{\delta}} = \frac{1}{2} - \frac{\overline{V}(\kappa)}{(1+\bar{\delta})\Delta^2 npL}\log\frac{n-k}{k}$ to make the dependence on $\bar{\delta}$ explicit. Recall that δ_2 and C_1' are independent of the $\bar{\delta}$ in the definition of $\eta_{\bar{\delta}}$. First we can let $\bar{\delta} > \delta_1$, then we have

$$\begin{aligned}
\frac{(1+\delta_2)\eta_{\bar{\delta}}^2\Delta^2 npL}{2\overline{V}(\kappa)} + C_1'\eta_{\bar{\delta}}\sqrt{\frac{\Delta^2 npL}{\overline{V}(\kappa)}} &\leq \left(1+\delta_2+2C_1'\left(\eta_{\bar{\delta}}\frac{\Delta^2 npL}{\overline{V}(\kappa)}\right)^{-\frac{1}{2}}\right)\frac{\eta_{\bar{\delta}}^2\Delta^2 npL}{2\overline{V}(\kappa)} \\
&\leq \left(1+\delta_2+2C_1'\left(\eta_{\delta_1}\frac{\Delta^2 npL}{\overline{V}(\kappa)}\right)^{-\frac{1}{2}}\right)\frac{\eta_{\bar{\delta}}^2\Delta^2 npL}{2\overline{V}(\kappa)} \\
&\leq (1+\delta_4)\,\frac{\eta_{\bar{\delta}}^2\Delta^2 npL}{2\overline{V}(\kappa)},
\end{aligned}$$

for some $\delta_4 = o(1)$ not dependent on $\bar{\delta}$. Here the second inequality is due to the fact that η_δ is in increasing function of δ, and the last inequality is due to (S23). Then we can let $\bar{\delta} \geq \delta_4$ to have the above expression to be upper bounded by $\left(1+\bar{\delta}\right)\frac{\eta_{\bar{\delta}}^2\Delta^2 npL}{2\overline{V}(\kappa)}$. Hence, (S142) leads to

$$\mathbb{P}_{(\theta^*,r^*)}\left(\sum_{k-\rho k<i\leq k} \mathbb{I}\left\{\widehat{\pi}_i < t\right\} \geq C_4 k \exp\left(-\frac{(1+\bar{\delta})\eta_{\bar{\delta}}^2\Delta^2 npL}{2\overline{V}(\kappa)}\right)\right) \geq 1 - o(1), \tag{S39}$$

witch establishes (S25).

Similar to (S142), we can establish

$$\mathbb{P}_{(\theta^*,r^*)}\left(\sum_{k<i\le k+\rho(n-k)}\mathbb{I}\left\{\widehat{\pi}_i>t\right\}\ge C_4(n-k)\exp\left(-\frac{(1+\delta_2)(1-\eta_{\bar\delta})^2\Delta^2npL}{2\overline{V}(\kappa)}-C_1'(1-\eta_{\bar\delta})\sqrt{\frac{\Delta^2npL}{\overline{V}(\kappa)}}\right)\right)$$
$$\ge 1-o(1).$$

Due to (S23), we have $(1-\eta_{\bar\delta})\in[0,1]$, then

$$\begin{aligned}\frac{(1+\delta_2)(1-\eta_{\bar\delta})^2\Delta^2npL}{2\overline{V}(\kappa)}+C_1'(1-\eta_{\bar\delta})\sqrt{\frac{\Delta^2npL}{\overline{V}(\kappa)}}&\le\frac{(1+\delta_2)(1-\eta_{\bar\delta})^2\Delta^2npL}{2\overline{V}(\kappa)}+C_1'\sqrt{\frac{\Delta^2npL}{\overline{V}(\kappa)}}\\&\le(1+\delta_5)\,\frac{(1-\eta_{\bar\delta})^2\Delta^2npL}{2\overline{V}(\kappa)},\end{aligned}$$

for some $\delta_5=o(1)$ not dependent on $\bar\delta$. Since $(1-\eta_{\bar\delta})^2\Delta^2npL/(2\overline{V}(\kappa))=\eta_{\bar\delta}^2\Delta^2npL/(2\overline{V}(\kappa))+2\log\frac{n-k}{k}/(1+\bar\delta)$, we have

$$\begin{aligned}&(n-k)\exp\left(-\frac{(1+\delta_2)(1-\eta_{\bar\delta})^2\Delta^2npL}{2\overline{V}(\kappa)}-C_1'\eta_{\bar\delta}\sqrt{\frac{\Delta^2npL}{\overline{V}(\kappa)}}\right)\\&\ge k\exp\left(\log\frac{n-k}{k}-(1+\delta_5)\,\frac{(1-\eta_{\bar\delta})^2\Delta^2npL}{2\overline{V}(\kappa)}\right)\\&=k\exp\left(\frac{\bar\delta-\delta_5}{1+\bar\delta}\log\frac{n-k}{k}-(1+\delta_5)\,\frac{\eta_{\bar\delta}^2\Delta^2npL}{2\overline{V}(\kappa)}\right).\end{aligned}$$

By letting $\bar\delta\ge\delta_5$ and using the same argument as in obtaining (S39), we have

$$\mathbb{P}_{(\theta^*,r^*)}\left(\sum_{k<i\le k+\rho(n-k)}\mathbb{I}\left\{\widehat{\pi}_i>t\right\}\ge C_4k\exp\left(-\frac{(1+\bar\delta)\eta_{\bar\delta}^2\Delta^2npL}{2\overline{V}(\kappa)}\right)\right)\ge 1-o(1), \tag{S40}$$

which establishes (S26). To sum up, we can choose $\bar\delta=\max\{\delta_1,\delta_4,\delta_5\}$ to establish (S25) and (S26).

The above proof assumes that $\kappa=\Omega(1)$ and $\overline{\mathsf{SNR}}$ satisfies (S23). When these two conditions do not hold, we need to slightly modify the argument. When (S23) is not satisfied, there must exist some small constant $\bar\epsilon>0$ such that $\frac{\sqrt{(1+\bar\epsilon)\overline{\mathsf{SNR}}}}{2}-\frac{1}{\sqrt{(1+\bar\epsilon)\overline{\mathsf{SNR}}}}\log\frac{n-k}{k}=O(1)$. We can then take ρ to be a sufficiently small constant, and the proof will go through with some slight modification. When $\kappa=o(1)$, we can simply construct θ^* by $\theta_i^*=0$ for $1\le i\le k$ and $\theta_i^*=-\Delta$ for $k+1\le i\le n$. □

Finally, we state and prove Lemma A.4 to close this section.

Lemma A.4. *Assume $\frac{np}{\log n}\to\infty$, $\kappa=O(1)$, $\rho=o(1)$, $k\to\infty$ and (20) holds for some arbitrarily small constant $\epsilon>0$. Choose $\kappa_1,\kappa_2\ge 0$ such that we have both $\kappa_1+\kappa_2\le\kappa$ and*

$$\frac{k\psi'(\kappa_1)(1+e^{\kappa_1})^2+(n-k)\psi'(\kappa_2)(1+e^{-\kappa_2})^2}{(k\psi(\kappa_1)+(n-k)\psi(-\kappa_2))^2/n}=\overline{V}(\kappa).$$

Define $\theta_i^* = \kappa_1$ *for all* $1 \le i \le k - \rho k$*,* $\theta_i^* = 0$ *for* $k - \rho k < i \le k$*,* $\theta_i^* = -\Delta$ *for* $k+1 \le i \le k + \rho(n-k)$ *and* $\theta_i^* = -\kappa_2$ *for* $k + \rho(n-k) < i \le n$ *and* $S = \{i \in [n] : i \le k - \rho k \text{ or } i > k\}$*. There exists some constants* $C_1 > 0$ *such that for any* $\tilde{\delta} = o(1)$*, there exists* $C_2 > 0$ *and* $\delta_1 = o(1)$ *such that for any* $\eta < 1/2$ *and any* $A \in \mathcal{A}$ *where* $\mathcal{A}$ *is defined in (S133)-(S135), we have*

$$\mathbb{P}\left(\frac{\sum_{j\in S} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))(1 + e^{\theta_j^* - \theta_i^*})}{\sum_{j\in[n]\setminus\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)} \le -(1+\tilde{\delta})\eta\Delta \middle| A \right)$$
$$\ge C_1 \exp\left(-\frac{1+\delta_1}{2}\eta_+^2 \overline{\mathsf{SNR}} - C_2\eta_+\sqrt{\overline{\mathsf{SNR}}} \right). \tag{S41}$$

for any $k - \rho k < i \le k$*.*

Proof. We suggest readers to go through the proof of Lemma B.3 in Section B.2 first. The proof of Lemma A.4 basically follows that of Lemma B.3. We will omit repeated details in the proof of Lemma B.3 and only present key steps and calculations specific to this Lemma A.4.

We denote $q_j = \psi(\theta_i - \theta_j)$. Then $1 + e^{\theta_j^* - \theta_i^*} = 1/q_j$ and $\psi(\theta_j - \theta_i) = 1 - q_j$. Then what we need to lower bound can be written as

$$\mathbb{P}_A\left(\sum_{\ell\in[L]}\sum_{j\in S} A_{ji}\frac{q_j - y_{ij\ell}}{q_j} \ge Lt' \right),$$

where $t' = (1+\delta')\eta\Delta\sum_{j\in[n]\setminus\{i\}} p(1-q_j)$ for some $\delta' = o(1)$ due to (S133)-(S135), and $\mathbb{P}_A$ is the conditional probability given A. Note that δ' can be chosen independent of η. We remark that

$$\overline{\mathsf{SNR}} = (1+\delta'')\frac{L\Delta^2(\sum_{j\in[n]\setminus\{i\}} p(1-q_j))^2}{\sum_{j\in S} p\frac{1-q_j}{q_j}}$$

due to $\rho = o(1)$ for some $\delta'' = o(1)$ independent of η. We still first consider the regime when

$$\eta\sqrt{\overline{\mathsf{SNR}}} \to \infty, \tag{S42}$$

This implies $\eta \in (0, 1/2)$.

The conditional cumulant of $\sum_{j\in S} A_{ji}\frac{q_j - y_{ijl}}{q_j}$ for each $l \in [L]$ is

$$\nu(u) = \sum_{j\in S} A_{ji}\log\left(q_j e^{\frac{u(q_j-1)}{q_j}} + (1-q_j)e^u \right) = \sum_{j\in S} A_{ji}\left[-u\frac{1-q_j}{q_j} + \log((1-q_j)e^{u/q_j} + q_j) \right].$$

The function $\nu(u)$ acts as the same role as $K(u)$ in the proof of Lemma B.3. Define

$$u^* = \arg\min_{u\ge 0}\left(L\nu(u) - uLt' \right).$$

Its first derivative is

$$\nu'(u) = \sum_{j\in S} A_{ji}\left[\frac{\frac{(1-q_j)}{q_j}e^{u/q_j}}{(1-q_j)e^{u/q_j} + q_j} - \frac{1-q_j}{q_j} \right].$$

Following the same argument in the proof of Lemma B.3, we need to pin down a range for u^*. First due to (S42) and $\nu'(0) = 0$, we have $t' > 0$ and thus $\nu'(0) - t' < 0$. Now for $u = o(1)$, we can approximate $\nu'(u)$ by Taylor expansion and obtain

$$1 - \delta_2 \leq \frac{\nu'(u)}{\overline{\nu'}(u)} \leq 1 + \delta_2, \tag{S43}$$

for some $0 < \delta_2 = o(1)$, where $\overline{\nu'}(u) = \sum_{j \in S} p \frac{1-q_i}{q_i} u$. Note that we can replace A_{ji} by p because of the condition $A \in \mathcal{A}$. Then we consider $\tilde{u} = \frac{2t'}{\sum_{j \in S} p \frac{1-q_i}{q_i}}$, which is $o(1)$ since $\Delta = o(1)$ and $\rho = o(1)$. Therefore,

$$\nu'(\tilde{u}) - t' \geq (1 - \delta_2)\overline{\nu'}(\tilde{u}) - t' = (1 - \delta_2)t' > 0.$$

This implies that $u^* \in \left(0, \frac{2t'}{\sum_{j \in S} p \frac{1-q_i}{q_i}}\right)$. Thus $u^* = o(1)$.

When $u = o(1)$, $\nu(u)$ also follows a second order Taylor expansion such that:

$$1 - \delta_3 \leq \frac{\nu(u)}{\bar{\nu}(u)} \leq 1 + \delta_3,$$

where $\bar{\nu}(u) = \frac{1}{2} \sum_{j \in S} p \frac{1-q_j}{q_j} u^2$ and $\delta_3 = o(1)$ due to (S133)-(S135).

Following the change-of-measure argument in the proof of Lemma B.3, the probability of interest can be lower bounded by

$$\exp\left(-u^* T + L\nu(u^*) - Lu^* t'\right) \mathbb{Q}_A \left(0 \leq \sum_{l=1}^{L} \sum_{j \in S} Z_{jl} - Lt' \leq T\right),$$

where $\mathbb{Q}_A$ is a measure under which Z_{jl} are all independent given A and follow

$$\mathbb{Q}_A(Z_{jl} = s) = e^{A_{ji} u^* s - A_{ji} \nu_j(u^*)} \mathbb{P}_A \left(A_{ji} \frac{q_j - y_{ijl}}{q_j} = s\right)$$

and $\nu_j(u) = -u \frac{1-q_j}{q_j} + \log((1 - q_j)e^{u/q_j} + q_j)$. Then for each Z_{jl} such that $A_{ij} = 1$, its second and 4th moment under $\mathbb{Q}_A$ can be analyzed:

$$\mathbb{Q}_A((Z_{jl} - \mathbb{Q}_A(Z_{jl}))^2) = \nu_j''(u^*) = \frac{1 - q_j}{q_j} \frac{e^{u^*/q_j}}{[(1 - q_j)e^{u^*/q_j} + q_j]^2} \in (C_1', C_2'), \tag{S44}$$

$$\mathbb{Q}_A((Z_{jl} - \mathbb{Q}_A(Z_{jl}))^4) = \nu_j''''(u^*) + 3\nu_j''(u^*) \leq (3 + C_4')\nu_j''(u^*) \leq C_3', \tag{S45}$$

where (S45) comes from

$$\nu_j''''(u) = \frac{1 - q_j}{q_j^3} e^{u/q_j} \frac{(1 - q_j)^3 e^{3u/q_j} - 3(1 - q_j)^2 q_j e^{2u/q_j} - 3(1 - q_j) q_j^2 e^{2u/q_j} + q_j^3}{[(1 - q_j)e^{u/q_j} + q_j]^5}$$
$$\leq \max_{j \in S} 1/q_j^2 \nu''(u) \leq C_4' \nu''(u).$$

Now, to lower bound $L\nu(u^*) - Lu^*t'$:

$$
\begin{aligned}
L\nu(u^*) - Lu^*t' &\geq L(1-\delta_3)\frac{1}{2}\sum_{j\in S} p\frac{1-q_j}{q_j}u^{*2} - Lu^*t' \\
&\geq L\min_{u\in(0,1)}\left[(1-\delta_3)\frac{1}{2}\sum_{j\in S} p\frac{1-q_j}{q_j}u^{*2} - u^*t'\right] \qquad \text{(S46)} \\
&\geq -\frac{1}{2}\frac{Lt'^2}{(1-\delta_3)\sum_{j\in S} p\frac{1-q_j}{q_j}} \\
&\geq -\frac{1+\delta_4}{2}\eta^2\mathsf{SNR},
\end{aligned}
$$

where (S46) is achieved at $u = \frac{t'}{(1-\delta_3)\sum_{j\in S} p\frac{1-q_j}{q_j}}$ and $\delta_4 = o(1)$ since $\rho = o(1)$. This gives us the desired exponent. We remark that δ_4 is independent of η.

To choose T, observe that

$$
\mathsf{Var}_{\mathbb{Q}_A}\left(\sum_{l\in[L]}\sum_{j\in S} Z_{jl}\right) \leq \tilde{C}_1 npL,
$$

for some constant $\tilde{C}_1 > 0$ using (S133) - (S135) , (S44) and $\rho = o(1)$. Thus we choose $T = \sqrt{\tilde{C}_1 npL}$, which leads to a term $C_2\eta\sqrt{\mathsf{SNR}}$ in the exponent for some $C_2 > 0$ independent of η.

Finally, to lower bound the $\mathbb{Q}_A$ measure, we only need to verify the vanishing property of the 4th moment approximation bound in Lemma E.3:

$$
\begin{aligned}
&\sqrt{L\sum_{j\in S} A_{ji}\left(\frac{\mathbb{Q}_A((Z_{j1} - \mathbb{Q}_A(Z_{j1})^4)}{(L\sum_{j\in S} A_{ji}\mathbb{Q}_A((Z_{j1} - \mathbb{Q}_A(Z_{j1})^2))^2}\right)^{3/4}} \\
&\leq \tilde{C}_2(npL)^{-1/4} \qquad \text{(S47)}
\end{aligned}
$$

where (S47) is by (S44), (S45) and $\rho = o(1)$. To summarize, we have proved

$$
\mathbb{P}_A\left(\sum_{l\in[L]}\sum_{j\in S} A_{ji}\frac{q_j - y_{ijl}}{q_j} \geq Lt'\right) \geq C_1\exp\left(-\frac{1+\delta_5}{2}\eta^2\mathsf{SNR} - C_2\eta\sqrt{\mathsf{SNR}}\right)
$$

for some constant $C_1, C_2 > 0$ and $\delta_5 = o(1)$, when (S42) holds. This δ_5 can be used as the δ_1 in (S41). We remark that C_1, C_2, δ_5 are all independent of η.

Finally, when

$$
\eta\sqrt{\mathsf{SNR}} \leq C_3
$$

for some constant $C_3 > 0$. This condition, together with (S133)-(S135) and $\rho = o(1)$, implies that

$$
Lt' \leq C_4\sqrt{L\sum_{j\in S} A_{ji}\frac{1-q_i}{q_i}}.
$$

Therefore,

$$\mathbb{P}_A\left(\sum_{l\in[L]}\sum_{j\in S}A_{ji}\frac{q_j-y_{ijl}}{q_j}\geq Lt'\right)\geq\mathbb{P}_A\left(\sum_{l\in[L]}\sum_{j\in S}A_{ji}\frac{q_j-y_{ijl}}{q_j}\geq C_5\sqrt{L\sum_{j\in S}A_{ji}\frac{1-q_i}{q_i}}\right)$$
$$\geq c_1-o(1) \tag{S48}$$

where (S48) comes from Lemma E.3. The 4th moment approximation can be checked to be of order $(npL)^{-1/4}$ similarly as in (S47) using (S133)-(S135) and $\rho=o(1)$ since the second and fourth moment of $\frac{q_j-y_{ijl}}{q_j}$ are at the constant order under measure $\mathbb{P}_A$, which completes the proof.

□

B Proofs of Lower Bounds

This section collects the proofs of lower bound results of the paper. The lower bound for exact recovery is proved in Section B.1, and the partial recovery lower bound is proved in Section B.2.

B.1 Proof of Theorem 3.4

The key mathematical argument in the proof of Theorem 3.4 is to characterize the maximum of dependent binomial random variables. For this purpose, we need a high-dimensional central limit theorem result by [4]. The following lemma is adapted from [4] for our purpose.

Lemma B.1. *Consider independent random vectors $X_1,\cdots,X_n\in\mathbb{R}^d$ with mean zero. Assume there exist constants $c_1,c_2,C_1,C_2>0$ such that $\min_{i,j}\mathbb{E}X_{ij}^2\geq c_1$, $\max_{i,j}\mathbb{E}\exp(|X_{ij}|/C_1)\leq 2$ and $(\log(nd))^7\leq C_2n^{-(1+c_2)}$. Then, there exist independent Gaussian vectors $Z_1,\cdots,Z_n$ satisfying $\mathbb{E}Z_i=0$ and $\mathsf{Cov}(Z_i)=\mathsf{Cov}(X_i)$, such that*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\max_{j\in[d]}\sum_{i=1}^n X_{ij}\leq t\right)-\mathbb{P}\left(\max_{j\in[d]}\sum_{i=1}^n Z_{ij}\leq t\right)\right|\leq Cn^{-c},$$

for some constants $c,C>0$ only depending on c_1,c_2,C_1,C_2.

With the above Gaussian approximation, we only need to analyze the maximum of dependent Gaussian random variables. The following lemma can be found in [6].

Lemma B.2. *Consider $Z=(Z_1,\cdots,Z_n)^T\sim N(0,\Sigma)$. Then, for any $\alpha\in(0,1)$, there exists some constant $C_\alpha>0$ such that for all $n\geq\sqrt{2\pi}e^3\log 1/\alpha$,*

$$\mathbb{P}\left(\max_{i\in[n]}Z_i>\lambda^{1/2}\sqrt{2\log n-\log\log n-C_\alpha}-\Lambda^{1/2}\Phi^{-1}(1-\alpha)\right)\geq 1-2\alpha,$$

where $\lambda=\min_{i\in[n]}\Sigma_{ii}-\frac{\max_{i\in[n]}\sum_{j\in[n]\setminus\{i\}}\Sigma_{ij}^2}{\lambda_{\min}(\Sigma)}$ and $\Lambda=\max_{i\in[n]}\Sigma_{ii}$.

Now we are ready to prove Theorem 3.4.

Proof of Theorem 3.4. We first note that the condition (16) implies that $\Delta = o(1)$. Choose $\kappa_1, \kappa_2 \geq 0$ such that we have both $\kappa_1 + \kappa_2 \leq \kappa$ and

$$\frac{n}{k\psi'(\kappa_1) + (n-k)\psi'(\kappa_2)} = V(\kappa).$$

We first consider the case $k \to \infty$ and $\kappa = \Omega(1)$. In this case, one can easily check that $\kappa_2 = \Omega(1)$. Our least favorable $\theta^* \in \Theta(k, \Delta, \kappa)$ is constructed as follows. Let $\rho = o(1)$ be a vanishing number that will be specified later. Define $\theta_i^* = \kappa_1$ for all $1 \leq i \leq k - \rho k$, $\theta_i^* = 0$ for $k - \rho k < i \leq k$, $\theta_i^* = -\Delta$ for $k < i \leq k + \rho(n-k)$ and $\theta_i^* = -\kappa_2$ for $k + \rho(n-k) < i \leq n$. For the simplicity of proof, we choose ρ so that both ρk and $\rho(n-k)$ are integers. Consider a subset $\mathcal{R}_{k,\rho} \subset \mathfrak{S}_n$ that is defined by

$$\mathcal{R}_{k,\rho} = \{r \in \mathfrak{S}_n : r_i = i \text{ for all } i \leq k - \rho k \text{ or } i > k + \rho(n-k)\}. \tag{S49}$$

We then have the lower bound

$$\inf_{\widehat{r}} \sup_{\substack{r^* \in \mathfrak{S}_n \\ \theta^* \in \Theta(k,\Delta,\kappa)}} \mathbb{P}_{(\theta^*, r^*)} \left(\mathsf{H}_k(\widehat{r}, r^*) > 0\right) \geq \inf_{\widehat{r}} \sup_{r^* \in \mathcal{R}_{k,\rho}} \mathbb{P}_{(\theta^*, r^*)} \left(\mathsf{H}_k(\widehat{r}, r^*) > 0\right).$$

For each $z = \{z_i\}_{k-\rho k < i \leq k+\rho(n-k)} \in \{0,1\}^{\rho n}$, we define $\mathbb{Q}_z$ as a joint probability of the observations $\{A_{ij}\}$ and $\{y_{ijl}\}$. To sample data from $\mathbb{Q}_z$, we first sample $A \sim \mathcal{G}(n,p)$, and then for any (i,j) such that $A_{ij} = 1$, sample $y_{ijl} \sim \text{Bernoulli}(\psi(\mu_i(z) - \mu_j(z)))$ independently for $l \in [L]$. The vector $\mu(z)$ is defined by $\mu_i(z) = \theta_i^*$ for all $i \leq k - \rho k$ or $i > \rho(n-k)$ and $\mu_i(z) = \Delta \mathbb{I}\{z_i = 1\}$ for all $k - \rho k < i \leq k + \rho(n-k)$. Then, we have

$$\begin{aligned}
\inf_{\widehat{r}} \sup_{r^* \in \mathcal{R}_{k,\rho}} \mathbb{P}_{(\theta^*, r^*)} \left(\mathsf{H}_k(\widehat{r}, r^*) > 0\right) &\geq \inf_{\widehat{z}} \sup_{z^* \in \mathcal{Z}_k} \mathbb{Q}_{z^*} \left(\widehat{z} \neq z^*\right) \\
&\geq \inf_{\widehat{z}} \frac{1}{|\mathcal{Z}_k|} \sum_{z^* \in \mathcal{Z}_k} \mathbb{Q}_{z^*}(\widehat{z} \neq z^*),
\end{aligned}$$

where

$$\mathcal{Z}_k = \left\{ z = \{z_i\}_{k-\rho k < i \leq k+\rho(n-k)} \in \{0,1\}^{\rho n} : \sum_i z_i = \rho k \right\}.$$

The Bayes risk $\frac{1}{|\mathcal{Z}_k|} \sum_{z^* \in \mathcal{Z}_k} \mathbb{Q}_{z^*}(\widehat{z} \neq z^*)$ is minimized by

$$\widehat{z} = \operatorname*{argmin}_{z \in \mathcal{Z}_k} \ell_n(\mu(z)), \tag{S50}$$

where

$$\ell_n(\mu(z)) = \sum_{1 \leq i < j \leq n} A_{ij} \left[\bar{y}_{ij} \log \frac{1}{\psi(\mu_i(z) - \mu_j(z))} + (1 - \bar{y}_{ij}) \log \frac{1}{1 - \psi(\mu_i(z) - \mu_j(z))} \right].$$

It suffices to lower bound the probability $\mathbb{Q}_{z^*}(\widehat{z} \neq z^*)$ for the estimator (S50) and for each $z^* \in \mathcal{Z}_k$. By symmetry, the value of $\mathbb{Q}_{z^*}(\widehat{z} \neq z^*)$ is the same for any $z^* \in \mathcal{Z}_k$. We therefore can set $z_i^* = \mathbb{I}\{i \leq k\}$ without loss of generality. Define

$$\mathcal{N}(z^*) = \left\{ z \in \mathcal{Z}_k : \sum_i \mathbb{I}\{z_i \neq z_i^*\} = 2 \right\}.$$

Then, we have

$$\mathbb{Q}_{z^*}(\widehat{z} \neq z^*) \geq \mathbb{Q}_{z^*}\left(\min_{z \in \mathcal{N}(z^*)} \ell_n(\mu(z)) < \ell_n(\mu(z^*)) \right).$$

By direct calculation, we have

$$\begin{aligned}
& \ell_n(\mu(z)) - \ell_n(\mu(z^*)) \\
= & \sum_{1 \leq i < j \leq n} A_{ij}(\bar{y}_{ij} - \psi(\mu_i(z^*) - \mu_j(z^*)))(\mu_i(z^*) - \mu_j(z^*) - \mu_i(z) + \mu_j(z)) \\
& + \sum_{1 \leq i < j \leq n} A_{ij} D\left(\psi(\mu_i(z^*) - \mu_j(z^*)) \| \psi(\mu_i(z) - \mu_j(z))\right).
\end{aligned}$$

For any $z \in \mathcal{N}(z^*)$, there exists some $k - \rho k < a \leq k$ and some $k < b \leq k + \rho(n-k)$ such that $z_a = 0$, $z_b = 1$ and $z_i = z_i^*$ for all other i's. Then,

$$\begin{aligned}
& \sum_{1 \leq i < j \leq n} A_{ij} D\left(\psi(\mu_i(z^*) - \mu_j(z^*)) \| \psi(\mu_i(z) - \mu_j(z))\right) \\
\leq & \sum_{i=1}^{k-\rho k} A_{ia} D(\psi(\kappa_1) \| \psi(\kappa_1 + \Delta)) + \sum_{i=k+\rho(n-k)+1}^{n} A_{ia} D(\psi(-\kappa_2) \| \psi(-\kappa_2 + \Delta)) \\
& + \sum_{i=1}^{k-\rho k} A_{ib} D(\psi(\kappa_1 + \Delta) \| \psi(\kappa_1)) + \sum_{i=k+\rho(n-k)+1}^{n} A_{ib} D(\psi(-\kappa_2 + \Delta) \| \psi(-\kappa_2)) \\
& + \sum_{i=k-\rho k+1}^{k} A_{ia} D(\psi(0) \| \psi(\Delta)) + \sum_{i=k+1}^{k+\rho(n-k)} A_{ia} D(\psi(-\Delta) \| \psi(0)) \\
& + \sum_{i=k-\rho k+1}^{k} A_{ib} D(\psi(\Delta) \| \psi(0)) + \sum_{i=k+1}^{k+\rho(n-k)} A_{ib} D(\psi(0) \| \psi(-\Delta)) + A_{ab} D(\psi(\Delta) \| \psi(-\Delta)) \\
\leq & (1+\delta)(1-\rho)p\left[kD(\psi(\kappa_1) \| \psi(\kappa_1 + \Delta)) + (n-k)D(\psi(-\kappa_2) \| \psi(-\kappa_2 + \Delta))\right] \qquad \text{(S51)} \\
& + (1+\delta)(1-\rho)p\left[kD(\psi(\kappa_1 + \Delta) \| \psi(\kappa_1)) + (n-k)D(\psi(-\kappa_2 + \Delta) \| \psi(-\kappa_2))\right] \\
& + (1+\delta)\rho p\left[kD(\psi(0) \| \psi(\Delta)) + (n-k)D(\psi(-\Delta) \| \psi(0))\right] \\
& + (1+\delta)\rho p\left[kD(\psi(\Delta) \| \psi(0)) + (n-k)D(\psi(0) \| \psi(-\Delta))\right] + (1+\delta)pD(\psi(\Delta) \| \psi(-\Delta)) \\
\leq & (1+\delta)^2(1-\rho)p\Delta^2\left[k\psi'(\kappa_1) + (n-k)\psi'(\kappa_2)\right] + (1+\delta)^2\rho p\Delta^2\frac{n}{4} \qquad \text{(S52)} \\
\leq & (1+\delta)^3 p\Delta^2 \frac{n}{V(\kappa)}. \qquad \text{(S53)}
\end{aligned}$$

The inequality (S51) holds with probability at least $1 - O(n^{-10})$ by Bernstein's inequality. The inequality (S52) is a Taylor expansion argument with the help of $\Delta = o(1)$. We obtain

(S53) by the choice that $\rho = o(1)$. Note that we can choose some $\delta = o(1)$ to make all of (S51), (S52) and (S53) hold. We also have

$$
\begin{aligned}
& \sum_{1\leq i<j\leq n} A_{ij}(\bar{y}_{ij} - \psi(\mu_i(z^*) - \mu_j(z^*)))(\mu_i(z^*) - \mu_j(z^*) - \mu_i(z) + \mu_j(z)) \\
= \;& -\Delta \sum_{i\in[n]\backslash\{a\}} A_{ia}(\bar{y}_{ia} - \mathbb{E}\bar{y}_{ia}) + \Delta \sum_{i\in[n]\backslash\{b\}} A_{ib}(\bar{y}_{ib} - \mathbb{E}y_{ib}).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& \min_{z\in\mathcal{N}(z^*)} \ell_n(\mu(z)) - \ell_n(\mu(z^*)) \\
\leq \;& -\max_{(1-\rho)k<a\leq k} \Delta \sum_{i\in[n]\backslash\{a\}} A_{ia}(\bar{y}_{ia} - \mathbb{E}\bar{y}_{ia}) + \Delta \min_{k<b\leq k+\rho(n-k)} \sum_{i\in[n]\backslash\{b\}} A_{ib}(\bar{y}_{ib} - \mathbb{E}y_{ib}) \\
& +(1+\delta)^3 p\Delta^2 \frac{n}{V(\kappa)},
\end{aligned}
$$

with probability at least $1 - O(n^{-10})$. This leads to the bound

$$
\begin{aligned}
& \mathbb{Q}_{z^*}\left(\min_{z\in\mathcal{N}(z^*)} \ell_n(\mu(z)) < \ell_n(\mu(z^*))\right) \\
\geq \;& \mathbb{Q}_{z^*}\Bigg(\max_{(1-\rho)k<a\leq k} \sum_{i\in[n]\backslash\{a\}} A_{ia}(\bar{y}_{ia} - \mathbb{E}\bar{y}_{ia}) \\
& - \min_{k<b\leq k+\rho(n-k)} \sum_{i\in[n]\backslash\{b\}} A_{ib}(\bar{y}_{ib} - \mathbb{E}y_{ib}) > (1+\delta)^3 p\Delta \frac{n}{V(\kappa)}\Bigg) - O(n^{-10}) \\
\geq \;& \mathbb{Q}_{z^*}\Bigg(\max_{(1-\rho)k<a\leq k} \sum_{i\in[n]\backslash\{a\}} A_{ia}(\bar{y}_{ia} - \mathbb{E}\bar{y}_{ia}) - \min_{k<b\leq k+\rho(n-k)} \sum_{i\in[n]\backslash\{b\}} A_{ib}(\bar{y}_{ib} - \mathbb{E}y_{ib}) \qquad \text{(S54)} \\
& > \sqrt{2(1-\epsilon/2)}\sqrt{\frac{np}{LV(\kappa)}}\left(\sqrt{\log k} + \sqrt{\log(n-k)}\right)\Bigg) - O(n^{-10}) \\
\geq \;& \mathbb{Q}_{z^*}\left(\max_{(1-\rho)k<a\leq k} \sum_{i\in[n]\backslash\{a\}} A_{ia}(\bar{y}_{ia} - \mathbb{E}\bar{y}_{ia}) > \sqrt{2(1-\epsilon/2)}\sqrt{\frac{np}{LV(\kappa)}}\sqrt{\log k}\right) \qquad \text{(S55)} \\
& +\mathbb{Q}_{z^*}\left(-\min_{k<b\leq k+\rho(n-k)} \sum_{i\in[n]\backslash\{b\}} A_{ib}(\bar{y}_{ib} - \mathbb{E}y_{ib}) > \sqrt{2(1-\epsilon/2)}\sqrt{\frac{np}{LV(\kappa)}}\sqrt{\log(n-k)}\right) \\
& -1 - O(n^{-10}),
\end{aligned}
$$

where we have used the condition of the theorem to derive (S54). The last inequality (S55) is by union bound $\mathbb{P}(A\cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$. To lower bound (S55), we introduce the notation

$$
T_a = \sum_{i\in[n]\backslash\{a\}} A_{ia}(\bar{y}_{ia} - \mathbb{E}\bar{y}_{ia}), \quad (1-\rho)k < a \leq k.
$$

The covariance structure of $\{T_a\}_{(1-\rho)k<a\leq k}$ can be quantified by the matrix $\Sigma \in \mathbb{R}^{(\rho k)\times(\rho k)}$, which is defined by $\Sigma_{ab} = \mathsf{Cov}(T_a, T_b|A)$. We then construct a vector $S = \{S_a\}_{(1-\rho)k<a\leq k}$ that is jointly Gaussian conditioning on A. The conditional covariance of S is also Σ. By Lemma B.1, we have

$$\mathbb{Q}_{z^*}\left(\max_{(1-\rho)k<a\leq k}\sum_{i\in[n]\backslash\{a\}} A_{ia}(\bar{y}_{ia}-\mathbb{E}\bar{y}_{ia}) > \sqrt{2(1-\epsilon/2)}\sqrt{\frac{np}{LV(\kappa)}}\sqrt{\log k}\right) \tag{S56}$$

$$\geq \mathbb{P}\left(\max_{(1-\rho)k<a\leq k} S_a > \sqrt{2(1-\epsilon/2)}\sqrt{\frac{np}{LV(\kappa)}}\sqrt{\log k}\right) - O\left(\frac{1}{(\log n)^c}\right). \tag{S57}$$

To see how Lemma B.1 implies (S57), we can take $X_{la} = \frac{1}{\sqrt{np}}\sum_{i\in[n]\backslash\{a\}} A_{ia}(y_{ial} - \mathbb{E}y_{ial})$. Conditioning on A, we observe that $\{X_{la}\}$ is independent across $l \in [L]$. The conditional variance of X_{la} given A is bounded away from zero with high probability by Lemma 8.1. Moreover, one can find a constant $C > 0$, such that $\mathbb{E}\left[\exp(|X_{la}|/C)\middle|A\right] \leq 2$ by Hoeffding's inequality. Then, we can apply Lemma B.1 for a given A and obtain (S57) under the condition $L > (\log n)^8$. We need Lemma B.2 to lower bound the probability in (S57). For each a,

$$\begin{aligned}
\Sigma_{aa} &= \mathsf{Var}(T_a|A) \\
&= \frac{1}{L}\sum_{i\in[n]\backslash\{a\}} A_{ia}\psi'(\mu_i(z^*)-\mu_a(z^*)) \\
&= \frac{\psi'(\kappa_1)}{L}\sum_{i=1}^{k-\rho k} A_{ia} + \frac{1}{4L}\sum_{i=k-\rho k+1}^{k} A_{ia} + \frac{\psi'(\kappa_2)}{L}\sum_{i=k+1}^{k+\rho(n-k)} A_{ia} + \frac{\psi'(\Delta)}{L}\sum_{i=k+\rho(n-k)+1}^{n} A_{ia}.
\end{aligned}$$

By Lemma 8.1, we have

$$\max_{(1-\rho)k<a\leq k}\Sigma_{aa} \leq \frac{1}{4L}\sum_{i\in[n]\backslash\{a\}} A_{ia} \leq \frac{np}{2L}, \tag{S58}$$

with probability at least $1-O(n^{-10})$. Similar to the proof of Lemma 8.1, we can use Bernstein's inequality and a union bound argument to obtain that

$$\begin{aligned}
\min_{(1-\rho)k<a\leq k}\Sigma_{aa} &\geq \min_{(1-\rho)k<a\leq k}\left[\frac{\psi'(\kappa_1)}{L}\sum_{i=1}^{k-\rho k} A_{ia} + \frac{\psi'(\kappa_2)}{L}\sum_{i=k+1}^{k+\rho(n-k)} A_{ia}\right] \\
&\geq \frac{(1-\delta)(1-\rho)p}{L}\left(k\psi'(\kappa_1)+(n-k)\psi'(\kappa_2)\right) \\
&= \frac{(1-\delta)(1-\rho)pn}{LV(\kappa)},
\end{aligned} \tag{S59}$$

for some $\delta = o(1)$ with probability at least $1-O(n^{-10})$. For each $a \neq b$,

$$\Sigma_{ab} = \mathsf{Cov}(T_a, T_b|A) = A_{ab}\frac{\psi'(\mu_a(z^*)-\mu_b(z^*))}{L}.$$

Then, Bernstein's inequality and a union bound argument, we have

$$\max_a \sum_{b:b\neq a} \Sigma_{ab}^2 \leq \frac{1}{16L^2} \max_{(1-\rho)k<a\leq k} \sum_{b:b\neq a} A_{ab} \leq C_1 \frac{\rho k p + \log n}{L^2}, \tag{S60}$$

with probability at least $1-O(n^{-10})$. We can also obtain a similar bound for $\max_a \sum_{b:b\neq a} \Sigma_{ab}$. This allows us to give a lower bound on $\lambda_{\min}(\Sigma)$:

$$\lambda_{\min}(\Sigma) \geq \min_{(1-\rho)k<a\leq k} \Sigma_{aa} - \max_a \sum_{b:b\neq a} \Sigma_{ab} \geq \frac{(1-\delta)(1-\rho)pn}{LV(\kappa)} - C_2 \frac{\rho k p + \log n}{L} \geq c_1 \frac{pn}{L}. \tag{S61}$$

To apply Lemma B.2, we shall choose ρ that satisfies both $\log(\rho k) = (1+o(1))\log k$ and $\rho = o(1)$. The existence of such ρ is guaranteed by $k \to \infty$. With the bounds (S58)-(S61), we can apply Lemma B.2, and obtain

$$\mathbb{P}\left(\max_{(1-\rho)k<a\leq k} S_a > \sqrt{2(1-\epsilon/2)}\sqrt{\frac{np}{LV(\kappa)}}\sqrt{\log k}\right) \geq 0.98 - O(n^{-1}).$$

We then obtain the desired lower bound for (S56). A similar argument also leads to

$$\begin{aligned} &\mathbb{Q}_{z^*}\left(-\min_{k<b\leq k+\rho(n-k)} \sum_{i\in[n]\setminus\{b\}} A_{ib}(\bar{y}_{ib} - \mathbb{E}y_{ib}) > \sqrt{2(1-\epsilon/2)}\sqrt{\frac{np}{LV(\kappa)}}\sqrt{\log(n-k)}\right) \\ \geq\ & 0.99 - O\left(\frac{1}{(\log n)^c}\right). \end{aligned}$$

Therefore, $\mathbb{Q}_{z^*}(\widehat{z} \neq z^*) \geq 0.95$ and we obtain the desired conclusion.

The above proof assumes that $k \to \infty$ and $\kappa = \Omega(1)$. When these two conditions do not hold, we need to slightly modify the argument. Let us briefly discuss two cases. In the first case, $k = O(1)$ and $\kappa = \Omega(1)$. In this case, we can construct θ^* by $\theta_i^* = 0$ for $1 \leq i \leq k$, $\theta_i^* = -\Delta$ for $k < i \leq k + \rho(n-k)$ and $\theta_i^* = -\kappa$ for $k + \rho(n-k) < i \leq n$. In the second case, $\kappa = o(1)$, and then we can take θ^* with $\theta_i^* = 0$ for $1 \leq i \leq k$ and $\theta_i^* = -\Delta$ for $k < i \leq n$. The remaining part of the proof will go through with similar arguments, and we will omit the details. □

B.2 Proof of Theorem 6.1

We first establish a lemma that lower bounds the error of a critical testing problem.

Lemma B.3. *Assume $\frac{np}{\log n} \to \infty$, $\kappa = O(1)$, $\rho = o(1)$, $k \to \infty$ and (16) holds for some arbitrarily small constant $\epsilon > 0$. Choose $\kappa_1, \kappa_2 \geq 0$ such that we have both $\kappa_1 + \kappa_2 \leq \kappa$ and*

$$\frac{n}{k\psi'(\kappa_1) + (n-k)\psi'(\kappa_2)} = V(\kappa).$$

Define $\theta_i = \kappa_1$ for $1 \leq i \leq k-\rho k$, $\theta_i = 0$ for $k-\rho k < i \leq k$, $\theta_i = -\Delta$ for $k+2 \leq i \leq k+\rho(n-k)$ and $\theta_i = -\kappa_2$ for $k+\rho(n-k) < i \leq n$. Suppose we have independent $A_i \sim \text{Bernoulli}(p)$ and

$z_{il} \sim \text{Bernoulli}(\psi(\theta_i))$ *for all* $i \in [n]\backslash\{k+1\}$ *and* $l \in [L]$. *Then, there exists some* $\delta = o(1)$ *such that*

$$\mathbb{P}\left(\sum_{l=1}^{L}\sum_{i\in[n]\backslash\{k+1\}} A_i\left[z_{il}\log\frac{\psi(\theta_i+\Delta)}{\psi(\theta_i)} + (1-z_{il})\log\frac{1-\psi(\theta_i+\Delta)}{1-\psi(\theta_i)}\right] \geq \log\frac{k}{n-k-1}\right)$$
$$\geq \quad C\exp\left(-\frac{1}{2}\left(\frac{\sqrt{(1+\delta)\mathit{SNR}}}{2} - \frac{1}{\sqrt{(1+\delta)\mathit{SNR}}}\log\frac{n-k}{k}\right)_+^2\right),$$

for some constant $C > 0$.

Proof. We first consider the case

$$\frac{\sqrt{(1+\delta)\mathsf{SNR}}}{2} - \frac{1}{\sqrt{(1+\delta)\mathsf{SNR}}}\log\frac{n-k}{k} \to \infty, \tag{S62}$$

for some $\delta = o(1)$ to be specified later. Throughout the proof, we use $\mathbb{P}_A$ for the conditional distribution $\mathbb{P}(\cdot|A)$. We use the notation

$$Z_l = \sum_{i\in[n]\backslash\{k+1\}} A_i\left[z_{il}\log\frac{\psi(\theta_i+\Delta)}{\psi(\theta_i)} + (1-z_{il})\log\frac{1-\psi(\theta_i+\Delta)}{1-\psi(\theta_i)}\right].$$

Its conditional cumulant generating function is

$$K(u) = \sum_{i\in[n]\backslash\{k+1\}} A_i \log\left(\psi(\theta_i)^{1-u}\psi(\theta_i+\Delta)^u + (1-\psi(\theta_i))^{1-u}(1-\psi(\theta_i+\Delta))^u\right).$$

Define

$$u^* = \operatorname*{argmin}_{u\geq 0}\left(LK(u) - u\log\frac{k}{n-k-1}\right).$$

By direct calculation, we have

$$\begin{aligned} K'(0) &= -\sum_{i\in[n]\backslash\{k+1\}} A_i D(\psi(\theta_i)\|\psi(\theta_i+\Delta)). \\ K'(1) &= \sum_{i\in[n]\backslash\{k+1\}} A_i D(\psi(\theta_i+\Delta)\|\psi(\theta_i)). \end{aligned}$$

By Bernstein's inequality,

$$K'(0) \leq -(1-\delta_1)p\sum_{i\in[n]\backslash\{k+1\}} D(\psi(\theta_i)\|\psi(\theta_i+\Delta)), \tag{S63}$$

$$K'(1) \geq (1-\delta_1)p\sum_{i\in[n]\backslash\{k+1\}} D(\psi(\theta_i+\Delta)\|\psi(\theta_i)), \tag{S64}$$

with some $\delta_1 = o(1)$ for probability at least $1 - O(n^{-1})$. Given that $\Delta = o(1)$, which is implied by (16), and $\rho = o(1)$, we have $\sum_{i\in[n]\backslash\{k+1\}} D(\psi(\theta_i)\|\psi(\theta_i+\Delta)) = (1+o(1))\frac{n\Delta^2}{2V(\kappa)}$

and $\sum_{i\in[n]\backslash\{k+1\}} D(\psi(\theta_i+\Delta)\|\psi(\theta_i)) = (1+o(1))\frac{n\Delta^2}{2V(\kappa)}$. With the condition (S62), we know that $LK'(0)-\log\frac{k}{n-k-1}<0$ and $LK'(1)-\log\frac{k}{n-k-1}>0$. Thus, we must have $u^*\in(0,1)$. In fact, the range of u^* can be further narrowed down. We apply a Taylor expansion of $K'(u)$ as a function of Δ near 0, and we obtain

$$K'(u) = \sum_{i\in[n]\backslash\{k+1\}} A_i\left[-\frac{1}{2}\psi'(\theta_i)\Delta^2+\psi'(\theta_i)u\Delta^2+O(|\Delta|^3)\right].$$

Note that the remainder term $O(|\Delta|^3)$ can be bounded by $|\Delta|^3$ up to some constant uniformly for all $u\in(0,1)$. By Bernstein's inequality, we have

$$K'(u)\geq-(1+\delta_1)\left(\frac{1}{2}-u\right)\frac{np\Delta^2}{V(\kappa)}, \tag{S65}$$

for all $u\in(0,1/2)$ with probability at least $1-O(n^{-1})$. By (S65), there exists $\delta'=o(1)$ such that

$$K'\left(\frac{1}{2}-\frac{1}{(1+\delta')\mathsf{SNR}}\log\frac{n-k}{k}\right)>0,$$

and therefore, we must have

$$u^*\in\left(0,\frac{1}{2}-\frac{1}{(1+\delta')\mathsf{SNR}}\log\frac{n-k}{k}\right). \tag{S66}$$

We also introduce a quadratic approximation for $K(u)$, which is

$$\overline{K}(u)=\frac{np\Delta^2}{2V(\kappa)}(u^2-u).$$

It can be shown that

$$1-\delta_2\leq\frac{K(u)}{\overline{K}(u)}\leq1+\delta_2, \tag{S67}$$

uniformly over all $u\in(0,1)$ for some $\delta_2=o(1)$ with probability at least $1-O(n^{-1})$. The inequality (S67) can be obtained by a Taylor expansion argument followed by Bernstein's inequality, similar to the approximation obtained in (S65).

Define a probability distribution $\mathbb{Q}_A$, under which $Z_1,\cdots,Z_L$ are i.i.d. given A and follow

$$\mathbb{Q}_A(Z_l=s)=\mathbb{P}_A(Z_l=s)e^{u^*s-K(u^*)},$$

for any s. It fact, each Z_l, under the measure $\mathbb{Q}_A$ can be written as the sum of several independent random variables, i.e. $Z_l=\sum_{i\in[n]\backslash\{k+1\}}Z_{il}$ where

$$\mathbb{Q}_A(Z_{il}=s)=e^{A_iu^*s-A_iK_i(u^*)}\mathbb{P}_A\left(A_i\left[z_{il}\log\frac{\psi(\theta_i+\Delta)}{\psi(\theta_i)}+(1-z_{il})\log\frac{1-\psi(\theta_i+\Delta)}{1-\psi(\theta_i)}\right]=s\right),$$

and $K_i(u)=\log\left(\psi(\theta_i)^{1-u}\psi(\theta_i+\Delta)^u+(1-\psi(\theta_i))^{1-u}(1-\psi(\theta_i+\Delta))^u\right)$. Then for each Z_{il} such that $A_i=1$, we can compute its second and 4th moment as

$$\mathbb{Q}_A((Z_{il}-\mathbb{Q}_A(Z_{il}))^2)=K_i''(u^*)=\psi'(\theta_i)\Delta^2\frac{e^{u^*\Delta}}{(1-\psi(\theta_i)+e^{u^*\Delta}\psi(\theta_i))^2}\in(C_1'\Delta^2,C_2'\Delta^2), \tag{S68}$$

$$\mathbb{Q}_A((Z_{il}-\mathbb{Q}_A(Z_{il}))^4) = K_i''''(u^*) + 3K_i''(u^*)^2 \leq \Delta^2 K_i''(u^*) + 3K_i''(u^*)^2, \tag{S69}$$

where $C_1', C_2' > 0$ in (S68) are some constants and we have used

$$K_i''''(u^*) = \psi'(\theta_i)\Delta^4 e^{u^*\Delta} \frac{\psi(\theta_i)^3 e^{3u^*\Delta} - 3\psi(\theta_i)\psi'(\theta_i)e^{2u^*\Delta} - 3\psi'(\theta_i)(1-\psi(\theta^*))e^{u^*\Delta} + (1-\psi(\theta_i))^3}{(1-\psi(\theta_i)+\psi(\theta_i)e^{u^*\Delta})^5}$$
$$\leq \psi'(\theta_i)\Delta^4 e^{u^*\Delta} \frac{1}{(1-\psi(\theta_i)+\psi(\theta_i)e^{u^*\Delta})^2} = \Delta^2 K_i''(u^*)$$

in (S69).

Define $\mathcal{A}$ to be the event of A that (S63), (S64), (S65), (S67) and

$$\frac{1}{2}np \leq \sum_{i\in[n]\setminus\{k+1\}} A_i \leq 2np, \tag{S70}$$

all hold. We know that $\mathbb{P}(A\in\mathcal{A}) \geq 1 - O(n^{-1})$.

With the above preparations, we can lower bound $\mathbb{P}\left(\sum_{l=1}^L Z_l \geq \log\frac{k}{n-k-1}\right)$ by

$$\inf_{A\in\mathcal{A}} \mathbb{P}_A\left(\sum_{l=1}^L Z_l \geq \log\frac{k}{n-k-1}\right)\mathbb{P}(A\in\mathcal{A}) \geq \frac{1}{2}\inf_{A\in\mathcal{A}} \mathbb{P}_A\left(\sum_{l=1}^L Z_l \geq \log\frac{k}{n-k-1}\right).$$

For any $A\in\mathcal{A}$, a change-of-measure argument leads to the lower bound

$$\begin{aligned}
&\mathbb{P}_A\left(\sum_{l=1}^L Z_l \geq \log\frac{k}{n-k-1}\right)\\
= \ & \exp\left(LK(u^*) - u^*\frac{k}{n-k-1}\right)\\
&\times \mathbb{Q}_A\left[\mathbb{I}\left\{\sum_{l=1}^L Z_l - \log\frac{k}{n-k-1} \geq 0\right\}\exp\left(-u^*(\sum_{l=1}^L Z_l - \log\frac{k}{n-k-1})\right)\right]\\
\geq \ & \exp\left(-u^*T + LK(u^*) - u^*\log\frac{k}{n-k-1}\right)\mathbb{Q}_A\left(0 \leq \sum_{l=1}^L Z_l - \log\frac{k}{n-k-1} \leq T\right),
\end{aligned}$$

for any $T > 0$ to be specified. We first lower bound the exponent $LK(u^*) - u^*\log\frac{k}{n-k-1}$ by

$$\begin{aligned}
LK(u^*) - u^*\log\frac{k}{n-k-1} &= \min_{u\in(0,1)}\left(LK(u) - u\log\frac{k}{n-k-1}\right)\\
&\geq \min_{u\in(0,1)}\left(L(1+\delta_2)\overline{K}(u) - u\log\frac{k}{n-k-1}\right)\\
&\geq -\frac{1}{2}\left(\frac{\sqrt{(1+\delta_3)\mathsf{SNR}}}{2} - \frac{1}{\sqrt{(1+\delta_3)\mathsf{SNR}}}\log\frac{n-k}{k}\right)^2,
\end{aligned}$$

for some $\delta_3 = o(1)$. We then need to choose an appropriate T so that the probability $\mathbb{Q}_A\left(0 \leq \sum_{l=1}^L Z_l - \log\frac{k}{n-k-1} \leq T\right)$ can be bounded below by some constant. To achieve this purpose, we note that

$$\mathsf{Var}_{\mathbb{Q}_A}\left(\sum_{l=1}^{L} Z_l\right) = L\sum_{i\in[n]\setminus\{k+1\}} A_i K_i''(u^*) \le C_1\Delta^2 L\sum_{i\in[n]\setminus\{k+1\}} A_i \le 2C_1\Delta^2 Lnp,$$

for some constant $C_1 > 0$ due to (S68), where $\mathsf{Var}_{\mathbb{Q}_A}$ is the variance operator under the measure $\mathbb{Q}_A$. Thus, we set $T = \sqrt{2C_1\Delta^2 Lnp}$. With this choice, and by (S66), we have

$$u^*T \le \sqrt{2C_1\Delta^2 Lnp}\left(\frac{1}{2} - \frac{1}{(1+\delta')\mathsf{SNR}}\log\frac{n-k}{k}\right).$$

Therefore, u^*T is at most the order of the square-root of the desired exponent, and thus it is negligible.

Finally, we need to show $\mathbb{Q}_A\left(0 \le \sum_{l=1}^{L} Z_l - \log\frac{k}{n-k-1} \le T\right)$ is lower bounded by some constant. Note that the definition of u^* implies that $\sum_{l=1}^{L} Z_l - \log\frac{k}{n-k-1}$ has mean zero under $\mathbb{Q}_A$. By the definition of T, we have

$$\begin{aligned}
&\mathbb{Q}_A\left(0 \le \sum_{l=1}^{L} Z_l - \log\frac{k}{n-k-1} \le T\right)\\
\ge\ &\mathbb{Q}_A\left(0 \le \sum_{l=1}^{L} Z_l - \log\frac{k}{n-k-1} \le \sqrt{\mathsf{Var}\left(\sum_{l=1}^{L} Z_l \,\middle|\, A\right)}\right)\\
=\ &\mathbb{Q}_A\left(0 \le \sum_{l=1}^{L}\sum_{i\in[n]\setminus\{k+1\}} Z_{il} - \log\frac{k}{n-k-1} \le \sqrt{\mathsf{Var}\left(\sum_{l=1}^{L}\sum_{i\in[n]\setminus\{k+1\}} Z_{il} \,\middle|\, A\right)}\right).
\end{aligned}$$

We apply the central limit theorem in Lemma E.3 to bound the above probability. The 4th moment approximation bound in Lemma E.3 is

$$\sqrt{L\sum_{i\in[n]\setminus\{k+1\}} A_i\left(\frac{K_i''''(u^*) + 3K_i''(u^*)^2}{(L\sum_{i\in[n]\setminus\{k+1\}} A_i K_i''(u^*))^2}\right)^{3/4}}$$

$$\le \sqrt{L\sum_{i\in[n]\setminus\{k+1\}} A_i\left(\frac{\Delta^2 K_i''(u^*) + 3K_i''(u^*)^2}{(L\sum_{i\in[n]\setminus\{k+1\}} A_i K_i''(u^*))^2}\right)^{3/4}} \tag{S71}$$

$$\le \sqrt{L\sum_{i\in[n]\setminus\{k+1\}} A_i\left(\frac{C_2' + 3C_2'^2}{(L\sum_{i\in[n]\setminus\{k+1\}} A_i C_1')^2}\right)^{3/4}} \tag{S72}$$

$$\le C_2\left(L\sum_{i\in[n]\setminus\{k+1\}} A_i\right)^{-1/4} \tag{S73}$$

which tends to zero by (S70). We have used (S69) in (S71), (S68) in (S72). We thus have

$$\mathbb{Q}_A\left(0 \leq \sum_{l=1}^{L} Z_l - \log\frac{k}{n-k-1} \leq T\right) \geq \mathbb{P}\left(0 \leq N(0,1) \leq 1\right) - o(1),$$

which is bounded below by a constant. To summarize, we have shown that

$$\mathbb{P}\left(\sum_{l=1}^{L} Z_l \geq \log\frac{k}{n-k-1}\right) \geq C_3 \exp\left(-\frac{1}{2}\left(\frac{\sqrt{(1+\delta_4)\mathsf{SNR}}}{2} - \frac{1}{\sqrt{(1+\delta_4)\mathsf{SNR}}}\log\frac{n-k}{k}\right)^2\right),$$

for some $\delta_4 = o(1)$ and some constant $C_3 > 0$ when (S62) holds with $\delta = \delta_4$.

To close the proof, we need a different argument when

$$\frac{\sqrt{(1+\delta_4)\mathsf{SNR}}}{2} - \frac{1}{\sqrt{(1+\delta_4)\mathsf{SNR}}}\log\frac{n-k}{k} \leq C_4,$$

for some constant $C_4 > 0$. This condition, together with Bernstein's inequality, implies that

$$\sum_{l=1}^{L} \mathbb{E}(Z_l|A) - \log\frac{k}{n-k-1} \geq -C_5\sqrt{Lnp\Delta^2}, \tag{S74}$$

with probability at least $1 - O(n^{-1})$. Define $\overline{\mathcal{A}}$ to be an event of A such that both (S70) and (S74) hold. It is clear that $\mathbb{P}(\overline{\mathcal{A}}) \geq 1 - O(n^{-1})$. We then have

$$\begin{aligned}
\mathbb{P}\left(\sum_{l=1}^{L} Z_l \geq \log\frac{k}{n-k-1}\right) &\geq \frac{1}{2}\inf_{A\in\overline{\mathcal{A}}} \mathbb{P}_A\left(\sum_{l=1}^{L} Z_l \geq \log\frac{k}{n-k-1}\right) \\
&\geq \frac{1}{2}\inf_{A\in\overline{\mathcal{A}}} \mathbb{P}_A\left(\sum_{l=1}^{L}(Z_l - \mathbb{E}(Z_l|A)) \geq C_5\sqrt{Lnp\Delta^2}\right) \qquad \text{(S75)} \\
&\geq c_1 - o(1), \qquad \text{(S76)}
\end{aligned}$$

for some constant $c_1 > 0$. The inequality (S75) is by (S74). For (S76), we use the Gaussian approximation in Lemma E.3, and the 4th moment approximation bound is of order $\left(L\sum_{i\in[n]\setminus\{k+1\}} A_i\right)^{-1/4}$ by similar calculation as in (S73) under measure $\mathbb{P}_A$, which tends to zero by (S70). The proof is complete. □

Proof of Theorem 6.1. We first note that the condition (16) implies that $\Delta = o(1)$. Choose $\kappa_1, \kappa_2 \geq 0$ such that we have both $\kappa_1 + \kappa_2 \leq \kappa$ and

$$\frac{n}{k\psi'(\kappa_1) + (n-k)\psi'(\kappa_2)} = V(\kappa).$$

We first consider the case $k \to \infty$ and $\kappa = \Omega(1)$. In this case, one can easily check that $\kappa_2 = \Omega(1)$. Our least favorable $\theta', \theta'' \in \Theta'(k, \Delta, \kappa)$ is constructed as follows. Let $\rho = o(1)$ be a vanishing number that will be specified later. Define $\theta'_i = \kappa_1$ for all $1 \leq i \leq k - \rho k$, $\theta'_i = 0$

for $k-\rho k < i \leq k$, $\theta'_i = -\Delta$ for $k < i \leq k+\rho(n-k)$ and $\theta'_i = -\kappa_2$ for $k+\rho(n-k) < i \leq n$. For the simplicity of proof, we choose ρ so that both ρk and $\rho(n-k)$ are integers. For θ'', we set $\theta''_i = \theta'_i$ for all $i \in [n]\backslash\{k+1\}$ and $\theta''_{k+1} = 0$. Recall the definition of the subset $\mathcal{R}_{k,\rho} \subset \mathfrak{S}_n$ in (S49). We then have

$$\begin{aligned}\inf_{\widehat{r}} \sup_{\substack{r^*\in\mathfrak{S}_n \\ \theta^*\in\Theta'(k,\Delta,\kappa)}} \mathbb{E}_{(\theta^*,r^*)}\mathsf{H}_k(\widehat{r},r^*) &\geq \inf_{\widehat{r}} \sup_{\substack{r^*\in\mathcal{R}_{k,\rho} \\ \theta^*\in\{\theta',\theta''\}}} \mathbb{E}_{(\theta^*,r^*)}\mathsf{H}_k(\widehat{r},r^*) \\ &\geq \inf_{\widehat{r}} \frac{1}{2} \sum_{\theta^*\in\{\theta',\theta''\}} \frac{1}{|\mathcal{R}_{k,\rho}|} \sum_{r^*\in\mathcal{R}_{k,\rho}} \mathbb{E}_{(\theta^*,r^*)}\mathsf{H}_k(\widehat{r},r^*).\end{aligned}$$

That is, we first lower bound the minimax risk by the Bayes risk. Since

$$\mathsf{H}_k(\widehat{r},r^*) \geq \frac{1}{2k} \sum_{k-\rho k<i\leq k+\rho(n-k)} \left(\mathbb{I}\{\widehat{r}_i > k, r^*_i \leq k\} + \mathbb{I}\{\widehat{r}_i \leq k, r^*_i > k\}\right),$$

we have

$$\begin{aligned}&\inf_{\widehat{r}} \sup_{\substack{r^*\in\mathfrak{S}_n \\ \theta^*\in\Theta'(k,\Delta,\kappa)}} \mathbb{E}_{(\theta^*,r^*)}\mathsf{H}_k(\widehat{r},r^*) \\ &\geq \inf_{\widehat{r}} \frac{1}{2} \sum_{\theta^*\in\{\theta',\theta''\}} \frac{1}{|\mathcal{R}_{k,\rho}|} \sum_{r^*\in\mathcal{R}_{k,\rho}} \mathbb{E}_{(\theta^*,r^*)} \frac{1}{2k} \sum_{k-\rho k<i\leq k+\rho(n-k)} \left(\mathbb{I}\{\widehat{r}_i > k, r^*_i \leq k\} + \mathbb{I}\{\widehat{r}_i \leq k, r^*_i > k\}\right) \\ &\geq \frac{1}{4k|\mathcal{R}_{k,\rho}|} \sum_{k-\rho k<i\leq k+\rho(n-k)} \inf_{\widehat{r}} \sum_{\theta^*\in\{\theta',\theta''\}} \left(\sum_{\substack{r^*\in\mathcal{R}_{k,\rho} \\ r^*_i\leq k}} \mathbb{P}_{(\theta^*,r^*)}(\widehat{r}_i > k) + \sum_{\substack{r^*\in\mathcal{R}_{k,\rho} \\ r^*_i\geq k+2}} \mathbb{P}_{(\theta^*,r^*)}(\widehat{r}_i \leq k) \right) \\ &\geq \frac{1}{4k|\mathcal{R}_{k,\rho}|} \sum_{k-\rho k<i\leq k+\rho(n-k)} \inf_{\widehat{r}} \left(\sum_{\substack{r^*\in\mathcal{R}_{k,\rho} \\ r^*_i\leq k}} \mathbb{P}_{(\theta'',r^*)}(\widehat{r}_i > k) + \sum_{\substack{r^*\in\mathcal{R}_{k,\rho} \\ r^*_i\geq k+2}} \mathbb{P}_{(\theta',r^*)}(\widehat{r}_i \leq k) \right).\end{aligned}$$

At this point, we need to introduce some extra notation. For any $r, r' \in \mathfrak{S}_n$, we define the Hamming distance without normalization as $\mathcal{H}(r,r') = \sum_{i=1}^n \mathbb{I}\{r_i \neq r'_i\}$. For each $k-\rho k < i \leq k+\rho(n-k)$, we can partition the set $\mathcal{R}_{k,\rho}$ into three disjoint subsets. Define

$$\begin{aligned}\mathcal{R}^{(1)}_{k,\rho} &= \{r\in\mathcal{R}_{k,\rho} : r_i \leq k\}, \\ \mathcal{R}^{(2)}_{k,\rho} &= \{r\in\mathcal{R}_{k,\rho} : r_i = k+1\}, \\ \mathcal{R}^{(3)}_{k,\rho} &= \{r\in\mathcal{R}_{k,\rho} : r_i \geq k+2\}.\end{aligned}$$

It is easy to see that $\mathcal{R}_{k,\rho} = \cup_{j=1}^3 \mathcal{R}^{(j)}_{k,\rho}$. We note that the three subsets all depend on the index i, but we shall suppress this dependence to avoid notational clutter. For any $r \in \mathcal{R}^{(2)}_{k,\rho}$, define

$$\begin{aligned}\mathcal{N}_{2\to1}(r) &= \left\{r''\in\mathcal{R}^{(1)}_{k,\rho} : \mathcal{H}(r,r'') = 2\right\}, \\ \mathcal{N}_{2\to3}(r) &= \left\{r'\in\mathcal{R}^{(3)}_{k,\rho} : \mathcal{H}(r,r') = 2\right\}.\end{aligned}$$

Since for any different permutations, the smallest Hamming distance between them is 2, $\mathcal{N}_{2\to1}(r)$ and $\mathcal{N}_{2\to3}(r)$ can be understood as neighborhoods r within $\mathcal{R}_{k,\rho}^{(1)}$ and $\mathcal{R}_{k,\rho}^{(3)}$, respectively. It is easy to check that $\{\mathcal{N}_{2\to1}(r)\}_{r\in\mathcal{R}_{k,\rho}^{(2)}}$ are disjoint subsets, and they form a partition of $\mathcal{R}_{k,\rho}^{(1)}$. Similarly, $\{\mathcal{N}_{2\to3}(r)\}_{r\in\mathcal{R}_{k,\rho}^{(2)}}$ are disjoint subsets, and form a partition of $\mathcal{R}_{k,\rho}^{(3)}$. With these notation, we have

$$
\begin{aligned}
&\inf_{\widehat{r}} \sup_{\substack{r^*\in\mathfrak{S}_n \\ \theta^*\in\Theta'(k,\Delta,\kappa)}} \mathbb{E}_{(\theta^*,r^*)}\mathsf{H}_k(\widehat{r},r^*) \\
&\geq \frac{1}{4k\,|\mathcal{R}_{k,\rho}|}\sum_{k-\rho k<i\leq k+\rho(n-k)} \inf_{\widehat{r}} \sum_{r\in\mathcal{R}_{k,\rho}^{(2)}} \left(\sum_{r''\in\mathcal{N}_{2\to1}(r)} \mathbb{P}_{(\theta'',r'')}(\widehat{r}_i>k) + \sum_{r'\in\mathcal{N}_{2\to3}(r)} \mathbb{P}_{(\theta',r')}(\widehat{r}_i\leq k)\right) \\
&= \frac{1}{4k\,|\mathcal{R}_{k,\rho}|}\sum_{k-\rho k<i\leq k+\rho(n-k)} \inf_{\widehat{r}} \sum_{r\in\mathcal{R}_{k,\rho}^{(2)}} \sum_{\substack{r''\in\mathcal{N}_{2\to1}(r)\\ r'\in\mathcal{N}_{2\to3}(r)}} \left(\frac{1}{n-k-1}\mathbb{P}_{(\theta'',r'')}(\widehat{r}_i>k) + \frac{1}{k}\mathbb{P}_{(\theta',r')}(\widehat{r}_i\leq k)\right) \\
&\geq \frac{1}{4k(n-k-1)\,|\mathcal{R}_{k,\rho}|}\sum_{k-\rho k<i\leq k+\rho(n-k)} \sum_{r\in\mathcal{R}_{k,\rho}^{(2)}} \sum_{\substack{r''\in\mathcal{N}_{2\to1}(r)\\ r'\in\mathcal{N}_{2\to3}(r)}} \inf_{0\leq\phi\leq1}\left[\mathbb{E}_{(\theta'',r'')}\phi + \frac{n-k-1}{k}\mathbb{E}_{(\theta',r')}(1-\phi)\right],
\end{aligned}
$$

where we have used the fact $|\mathcal{N}_{2\to1}(r)| = k$ and $|\mathcal{N}_{2\to3}(r)| = n-k-1$ to obtain the equality in the above display. To this end, it suffices to give a lower bound for the testing problem

$$
\inf_{0\leq\phi\leq1}\left[\mathbb{E}_{(\theta'',r'')}\phi + \frac{n-k-1}{k}\mathbb{E}_{(\theta',r')}(1-\phi)\right], \tag{S77}
$$

for any $r''\in\mathcal{N}_{2\to1}(r)$ and any $r'\in\mathcal{N}_{2\to3}(r)$ with any $r\in\mathcal{R}_{k,\rho}^{(2)}$ and any $k-\rho k<i\leq k+\rho(n-k)$.

For the two probability distributions in (S77), the probability $\mathbb{P}_{(\theta'',r'')}$ is the BTL model with parameter $\{\theta''_{r''_i}\}_{i\in[n]}$ and the probability $\mathbb{P}_{(\theta',r')}$ is the BTL model with parameter $\{\theta'_{r'_i}\}_{i\in[n]}$. It turns out the two vectors $\{\theta''_{r''_i}\}_{i\in[n]}$ and $\{\theta'_{r'_i}\}_{i\in[n]}$ only differ by one entry. To see this, let i and j' be the two coordinates that r and r' differ and let i and j'' be the two coordinates that r and r'' differ. Then, r' and r'' differ at the ith, the j'th and the j''th coordinates. This immediately implies $\theta'_{r'_l} = \theta''_{r''_l}$ for all $l\in[n]\backslash\{i,j',j''\}$. By the definitions of $\mathcal{N}_{2\to1}$ and $\mathcal{N}_{2\to3}$, we have $r'_i = r_{j'}, r'_{j'} = k+1, r'_{j''} = r_{j''}$ and $r''_i = r_{j''}, r''_{j'} = r_{j'}, r''_{j''} = k+1$. Moreover, we also have $r_{j'}\geq k+2$ and $r_{j''}\leq k$. We remind the readers that all the three coordinates are in the interval $[k-\rho k+1, k+\rho(n-k)]$. According to the definitions of θ' and θ'', we then have $\theta'_{r'_{j''}} = \theta''_{r''_{j''}} = 0$ and $\theta'_{r'_{j'}} = \theta''_{r''_{j'}} = -\Delta$. For the only different coordinate, we have $\theta'_{r'_i} = -\Delta$ and $\theta''_{r''_i} = 0$.

Since $\{\theta''_{r''_i}\}_{i\in[n]}$ and $\{\theta'_{r'_i}\}_{i\in[n]}$ only differ by a single coordinate, the testing problem (S77) is equivalent to

$$
\inf_{0\leq\phi\leq1}\left[\mathbb{E}_{(\theta'',\bar{r})}\phi + \frac{n-k-1}{k}\mathbb{E}_{(\theta',\bar{r})}(1-\phi)\right], \tag{S78}
$$

where $\bar{r}_i = i$ for all $i\in[n]$. The equivalence between (S77) and (S78) can be obtained by the existence of a simultaneous permutation that maps the two vectors $\{\theta''_{r''_i}\}_{i\in[n]}$ and $\{\theta'_{r'_i}\}_{i\in[n]}$

to θ'' and θ'. By Neyman-Pearson lemma, we can lower bound (S78) by

$$\mathbb{P}_{(\theta'',\bar{r})}\left(\frac{d\mathbb{P}_{(\theta',\bar{r})}}{d\mathbb{P}_{(\theta'',\bar{r})}} \geq \frac{k}{n-k-1}\right). \tag{S79}$$

This probability can be lower bounded by

$$C\exp\left(-\frac{1}{2}\left(\frac{\sqrt{(1+\delta)\mathsf{SNR}}}{2} - \frac{1}{\sqrt{(1+\delta)\mathsf{SNR}}}\log\frac{n-k}{k}\right)_+^2\right),$$

with some constant $C > 0$ and some $\delta = o(1)$ according to Lemma B.3. Since $|\mathcal{R}_{k,\rho}^{(2)}|/|\mathcal{R}_{(k,\rho)}| = (1-\rho)n$, we have

$$\begin{aligned}
&\inf_{\widehat{r}} \sup_{\substack{r^*\in\mathfrak{S}_n \\ \theta^*\in\Theta'(k,\Delta,\kappa)}} \mathbb{E}_{(\theta^*,r^*)}\mathsf{H}_k(\widehat{r},r^*) \\
\geq\ & C_1\rho\exp\left(-\frac{1}{2}\left(\frac{\sqrt{(1+\delta)\mathsf{SNR}}}{2} - \frac{1}{\sqrt{(1+\delta)\mathsf{SNR}}}\log\frac{n-k}{k}\right)_+^2\right),
\end{aligned}$$

for some constant $C_1 > 0$. When the exponent diverges, we can choose ρ that tends to zero sufficiently slow so that it can be absorbed into the exponent. Otherwise, we can simply set ρ to be a sufficiently small constant, and the above proof will still go through. One can use a similar argument as Lemma B.3 to show (S79) is bounded below by some constant. In this case, we have $\inf_{\widehat{r}}\sup_{\substack{r^*\in\mathfrak{S}_n \\ \theta^*\in\Theta'(k,\Delta,\kappa)}} \mathbb{E}_{(\theta^*,r^*)}\mathsf{H}_k(\widehat{r},r^*)$ bounded below by some constant as desired.

Finally, we briefly discuss how to modify the proof when either $k\to\infty$ or $\kappa=\Omega(1)$ does not hold. When $k\to\infty$ and $\kappa = o(1)$, we can take $\theta'_i = 0$ for $1\leq i\leq k$ and $\theta'_i = -\Delta$ for $k < i \leq n$. The vector θ'' is still defined according to $\theta''_i = \theta'_i$ for all $i\in[n]\backslash\{k+1\}$ and $\theta''_{k+1} = 0$. The proof will go through with some slight modifcation. When $k = O(1)$, the condition (16) is equivalent to $\mathsf{SNR} < (1-\epsilon)2\log n$ for some constant $\epsilon > 0$, and we only need to prove a constant minimax lower bound. This is obviously true becasue

$$\begin{aligned}
\inf_{\widehat{r}} \sup_{\substack{r^*\in\mathfrak{S}_n \\ \theta^*\in\Theta'(k,\Delta,\kappa)}} \mathbb{E}_{(\theta^*,r^*)}\mathsf{H}_k(\widehat{r},r^*) &\geq \inf_{\widehat{r}} \sup_{\substack{r^*\in\mathfrak{S}_n \\ \theta^*\in\Theta(k,\Delta,\kappa)}} \mathbb{E}_{(\theta^*,r^*)}\mathsf{H}_k(\widehat{r},r^*) \\
&\geq \inf_{\widehat{r}} \sup_{\substack{r^*\in\mathfrak{S}_n \\ \theta^*\in\Theta(k,\Delta,\kappa)}} \frac{1}{2k}\mathbb{P}_{(\theta^*,r^*)}\left(\mathsf{H}_k(\widehat{r},r^*) > 0\right),
\end{aligned}$$

which is lower bounded by a constant by Theorem 3.4 and the condition that $k = O(1)$. □

C Proofs of Local Error Rates

In this section, we prove Theorem 7.1 and Theorem 7.2.

C.1 Proof of Theorem 7.1

We first give Lemma C.1 to characterize entrywise tail behaviors of the MLE (6) which is crucial to the upper bound in Theorem 7.1.

Lemma C.1. *Assume $\frac{np}{\log n} \to \infty$ and $\kappa = O(1)$. Then, for the rank vector $\widehat{r}$ that is induced by the MLE (6), for any small constant $0.1 > \delta > 0$, there exists some constant $C > 0$, such that for any $t \in \mathbb{R}$, any $\theta^* \in \Theta(k, 0, \kappa)$, $r^* \in \mathfrak{S}_n$, we have*

$$\mathbb{P}_{(\theta^*, r^*)}\left(\widehat{\theta}_i \leq t\right) \leq C \exp\left(-\frac{(1-\delta)(\theta^*_{r_i^*} - t)_+^2 npL}{2V_{r_i^*}(\theta^*)}\right) + Cn^{-7}, r_i^* \leq k; \tag{S80}$$

$$\mathbb{P}_{(\theta^*, r^*)}\left(\widehat{\theta}_i \geq t\right) \leq C \exp\left(-\frac{(1-\delta)(t - \theta^*_{r_i^*})_+^2 npL}{2V_{r_i^*}(\theta^*)}\right) + Cn^{-7}, r_i^* \geq k+1 \tag{S81}$$

Proof. The proof follows the proof of Theorem 3.2 with slight modifications. Without loss of generality, we can assume $r_i^* = i$ for all $\in [n]$. Let

$$\bar{\Delta}_i = \begin{cases} \min\left((\theta_i^* - t)_+, \left(\frac{\log n}{np}\right)^{1/4}\right), & 1 \leq i \leq k, \\ \min\left((t - \theta_i^*)_+, \left(\frac{\log n}{np}\right)^{1/4}\right), & k+1 \leq i \leq n. \end{cases} \tag{S82}$$

We only need to prove (S80) since (S81) can be proved similarly.

Consider any $m \in [k]$. When $(\theta_m^* - t)_+^2 npL \leq c'$ for some large enough constant to be specified later, we can directly bound the probability using the trivial bound 1. Thus, we only need to consider the regime when $(\theta_m^* - t)_+^2 npL > c'$.

Following the proof of Theorem 3.2, we have (57)-(63) and (65) hold. Note that we now have $\bar{\Delta}_m^2 Lnp > c'$ instead of $\bar{\Delta}_m^2 Lnp \to \infty$ which is needed in the proof of Theorem 3.2. As a consequence, we now have (64) and (66) hold with $\delta = 4C_4 e^{\kappa}/\sqrt{c'}$ instead of some $o(1)$ as in the proof of Theorem 3.2. To sum up, with this δ, we have

$$|\widehat{\theta}_m - \bar{\theta}_m| \leq \delta \bar{\Delta}_m, \tag{S83}$$

$$\frac{|f^{(m)}(\theta_m^* | \widehat{\theta}_{-m}) - f^{(m)}(\theta_m^* | \theta_{-m}^*)|}{g^{(m)}(\theta_m^* | \theta_{-m}^*)} \leq \delta \bar{\Delta}_m, \tag{S84}$$

$$\frac{\left|g^{(m)}(\theta_m^* | \widehat{\theta}_{-m}) - g^{(m)}(\theta_m^* | \theta_{-m}^*)\right|}{g^{(m)}(\theta_m^* | \theta_{-m}^*)} \leq \delta, \tag{S85}$$

hold with probability at least $1 - O(n^{-7}) - \exp(-\bar{\Delta}_m^{3/2} Lnp) - \exp\left(-\bar{\Delta}_m^2 npL \frac{np}{\log n}\right)$. We can make δ to be an arbitrarily small constant by setting c' large as $\kappa = O(1)$.

Then for any $i \le k$, by the same argument as in the proof of Theorem 3.2, we have

$$\begin{aligned}
&\mathbb{P}\left(\widehat{\theta}_i \le t\right) \\
\le\ & \mathbb{P}\left(\widehat{\theta}_i - \theta_i^* \le -(\theta_i^* - t)\right) \\
\le\ & \mathbb{P}\left(\bar{\theta}_i - \theta_i^* \le -(1-\delta)\bar{\Delta}_i\right) + \mathbb{P}\left(|\bar{\theta}_i - \widehat{\theta}_i| > \delta\bar{\Delta}_i\right) \\
\le\ & \mathbb{P}\left(-\frac{f^{(i)}(\theta_i^*|\theta_{-i}^*)}{g^{(i)}(\theta_i^*|\theta_{-i}^*)} \le -(1-3\delta)\bar{\Delta}_i\right) + O(n^{-7}) \qquad \text{(S86)} \\
& + \exp(-\bar{\Delta}_i^{3/2}Lnp) + \exp\left(-\bar{\Delta}_i^2 npL\frac{np}{\log n}\right),
\end{aligned}$$

which has the same upper bound as in (67). We then have the same (68) and the event $\mathcal{A}_i$ as in the proof of Theorem 3.2. As a result,

$$\begin{aligned}
& \mathbb{P}\left(-\frac{f^{(i)}(\theta_i^*|\theta_{-i}^*)}{g^{(i)}(\theta_i^*|\theta_{-i}^*)} \le -(1-3\delta)\bar{\Delta}_i\right) \\
\le\ & \sup_{A\in\mathcal{A}_i} \exp\left(-\frac{\frac{1}{2}(1-3\delta)^2\bar{\Delta}_i^2\left(L\sum_{j\in[n]\setminus\{i\}}A_{ij}\psi'(\theta_i^*-\theta_j^*)\right)^2}{L\sum_{j\in[n]\setminus\{i\}}A_{ij}\psi'(\theta_i^*-\theta_j^*) + \frac{1-3\delta}{3}\bar{\Delta}_i L\sum_{j\in[n]\setminus\{i\}}A_{ij}\psi'(\theta_i^*-\theta_j^*)}\right) \\
& + O(n^{-7}) \\
=\ & \exp\left(-\frac{1-\delta'}{2}\bar{\Delta}_i^2 Lp\sum_{j\in[n]\setminus\{i\}}\psi'(\theta_i^*-\theta_j^*)\right) + O(n^{-7}) \qquad \text{(S87)} \\
\le\ & \exp\left(-\frac{1-\delta'}{2}(\theta_i^*-t)^2 Lp\sum_{j\in[n]\setminus\{i\}}\psi'(\theta_i^*-\theta_j^*)\right) + O(n^{-7}) \qquad \text{(S88)} \\
=\ & \exp\left(-\frac{1-\delta''}{2V_i(\theta^*)}(\theta_i^*-t)^2 npL\right) + O(n^{-7}) \qquad \text{(S89)}
\end{aligned}$$

where δ', δ'' are able to be any small constant (by adjusting c'). We use the definition of $\mathcal{A}_i$ to obtain the expression (S87). To see why (S88) is true, note that when $\bar{\Delta}_i^2 = \sqrt{\frac{\log n}{np}}$, the first term of (S87) can be absorbed into $O(n^{-7})$. (S89) comes from $\frac{\sum_{j\in[n]\setminus\{i\}}\psi'(\theta_i^*-\theta_j^*)}{\sum_{j\in[n]}\psi'(\theta_i^*-\theta_j^*)} = 1 + o(1)$.

Since $\exp(-\bar{\Delta}_i^{3/2}Lnp) + \exp\left(-\bar{\Delta}_i^2 npL\frac{np}{\log n}\right) \le \exp\left(-\frac{1+o(1)}{2V_i(\theta^*)}(\theta_i^*-t)^2 npL\right) + O(n^{-7})$, we have for any small constant $\delta > 0$, there exists some constant $C > 0$, such that

$$\mathbb{P}\left(\widehat{\theta}_i \le t\right) \le C\exp\left(-\frac{1-\delta}{2V_i(\theta^*)}(\theta_i^*-t)^2 npL\right) + Cn^{-7}, \qquad \text{(S90)}$$

for all $i \le k$ which completes the proof. □

Proof of (29) of Theorem 7.1. The upper bound (29) is a straightforward consequence of

Lemma 3.1 and Lemma C.1. We have

$$\begin{aligned}
&\mathbb{E}_{(\theta^*,r^*)}\mathsf{H}_k(\widehat{r},r^*) \\
&\leq C\frac{1}{k}\left[\sum_{i=1}^{k}\exp\left(-\frac{(1-\delta)(\theta_i^*-t)_+^2 npL}{2V_i(\theta^*)}\right)+\sum_{i=k+1}^{n}\exp\left(-\frac{(1-\delta)(t-\theta_i^*)_+^2 npL}{2V_i(\theta^*)}\right)\right]+Cn^{-6}.
\end{aligned}$$

□

The rest of the section focuses on the lower bound (30). The proof follows the proof of Theorem 3.4 with some modification. We include it below for completeness.

Proof of (30) of Theorem 7.1. We are going to prove

$$\mathbb{E}_{(\theta^*,r^*)}\mathsf{H}_k(\widehat{r},r^*) \gtrsim \frac{R_1([k],\theta^*,t^*,-\delta)+R_2([n]\backslash[k],\theta^*,t^*,-\delta)}{k} \tag{S91}$$

where t^* is the unique solution such that $R_1([k],\theta^*,t^*,-\delta) = R_2([n]\backslash[k],\theta^*,t^*,-\delta)$. We first show the existence and uniqueness of t^*. Note that $R_1([k],\theta^*,t,-\delta)$ increases with t while $R_2([n]\backslash[k],\theta^*,t,-\delta)$ decreases with t. Moreover, since $\lim_{t\to-\infty}R_1([k],\theta^*,t,-\delta) = \lim_{t\to+\infty}R_2([n]\backslash[k],\theta^*,t,-\delta) = 0$, such t^* must exist due to continuity. The uniqueness comes from $R_1([k],\theta^*,t,-\delta)$, as a function of t, is strictly increasing on $(-\infty,\theta_1^*]$ and $R_2([n]\backslash[k],\theta^*,t,-\delta)$, as a function of t, is strictly decreasing on $[\theta_n^*,+\infty)$ and $\theta_1^*\geq\theta_n^*$.

Define

$$S_1(t) = \left\{i\in[n]: i\leq k, (\theta_i^*-t)_+ \leq (\log n/np)^{1/4}\right\}, \tag{S92}$$

$$S_2(t) = \left\{i\in[n]: i\geq k+1, (t-\theta_i^*)_+ \leq (\log n/np)^{1/4}\right\}.$$

Since we assume $\inf_t(R_1([k],\theta^*,t,-\delta)+R_2([n]\backslash[k],\theta^*,t,-\delta))\to\infty$, we must have

$$R_1([k],\theta^*,t^*,-\delta)\to\infty \tag{S93}$$

and hence,

$$\frac{R_1(S_1(t^*),\theta^*,t^*,-\delta)}{R_1([k],\theta^*,t^*,-\delta)}\geq\frac{1}{2}, \frac{R_1(S_2(t^*),\theta^*,t^*,-\delta)}{R_1([n]\backslash[k],\theta^*,t^*,-\delta)}\geq\frac{1}{2}. \tag{S94}$$

This is because $R_1([k],\theta^*,t^*,-\delta) - R_1(S_1(t^*),\theta^*,t^*,-\delta) \leq n^{-6}$ and $R_2([n]\backslash[k],\theta^*,t^*,-\delta) - R_2(S_2(t^*),\theta^*,t^*,-\delta) \leq n^{-6}$ by the definition of $S_1(t^*)$, $S_2(t^*)$ and $np/\log n\to\infty$.

Now by Lemma A.3, we have

$$\begin{aligned}
\mathsf{H}_k(\widehat{r},r^*) &\geq \frac{1}{k}\min\left(\sum_{i=1}^{k}\mathbb{I}\left\{\widehat{\theta}_i<t^*\right\}, \sum_{i=k+1}^{n}\mathbb{I}\left\{\widehat{\theta}_i>t^*\right\}\right) \\
&\geq \frac{1}{k}\min\left(\sum_{i\in S_1(t^*)}\mathbb{I}\left\{\widehat{\theta}_i<t^*\right\}, \sum_{i\in S_2(t^*)}\mathbb{I}\left\{\widehat{\theta}_i>t^*\right\}\right).
\end{aligned} \tag{S95}$$

It suffices to show there exists some constant $C > 0$ such that

$$\mathbb{P}_{(\theta^*,r^*)}\left(\frac{1}{k}\sum_{i\in S_1(t^*)}\mathbb{I}\left\{\widehat{\theta}_i < t^*\right\} \geq \frac{4C}{k}R_1(S_1(t^*),\theta^*,t^*,-\delta)\right) \geq 3/4 \tag{S96}$$

$$\text{and } \mathbb{P}_{(\theta^*,r^*)}\left(\frac{1}{k}\sum_{i\in S_2(t^*)}\mathbb{I}\left\{\widehat{\theta}_i > t\right\} \geq \frac{4C}{k}R_2(S_2(t^*),\theta^*,t^*,-\delta)\right) \geq 3/4. \tag{S97}$$

This is because

$$\begin{aligned}
&\mathbb{E}_{(\theta^*,r^*)}\mathsf{H}_k(\widehat{r},r^*) \\
&\geq C\frac{R_1([k],\theta^*,t^*,-\delta)+R_2([n]\backslash[k],\theta^*,t^*,-\delta)}{k} \\
&\quad\times\mathbb{P}_{(\theta^*,r^*)}\left(\mathsf{H}_k(\widehat{r},r^*)\geq C\frac{R_1([k],\theta^*,t^*,-\delta)+R_2([n]\backslash[k],\theta^*,t^*,-\delta)}{k}\right) &\text{(S98)}\\
&\geq C\frac{R_1([k],\theta^*,t^*,-\delta)+R_2([n]\backslash[k],\theta^*,t^*,-\delta)}{k}\mathbb{P}_{(\theta^*,r^*)}\left(\begin{array}{c}\frac{\sum_{i\in S_1(t^*)}\mathbb{I}\{\widehat{\theta}_i<t^*\}}{k}\geq\frac{2C}{k}R_1([k],\theta^*,t^*,-\delta)\text{ and}\\ \frac{\sum_{i\in S_2(t^*)}\mathbb{I}\{\widehat{\theta}_i>t\}}{k}\geq\frac{2C}{k}R_2([n]\backslash[k],\theta^*,t^*,-\delta)\end{array}\right) &\text{(S99)}\\
&\geq C\frac{R_1([k],\theta^*,t^*,-\delta)+R_2([n]\backslash[k],\theta^*,t^*,-\delta)}{k}\mathbb{P}_{(\theta^*,r^*)}\left(\begin{array}{c}\frac{\sum_{i\in S_1(t^*)}\mathbb{I}\{\widehat{\theta}_i<t^*\}}{k}\geq\frac{4C}{k}R_1(S_1(t^*),\theta^*,t^*,-\delta)\text{ and}\\ \frac{\sum_{i\in S_2(t^*)}\mathbb{I}\{\widehat{\theta}_i>t\}}{k}\geq\frac{4C}{k}R_2(S_2(t^*),\theta^*,t^*,-\delta)\end{array}\right) &\text{(S100)}\\
&\geq\frac{C}{2}\frac{R_1([k],\theta^*,t^*,-\delta)+R_2([n]\backslash[k],\theta^*,t^*,-\delta)}{k}. &\text{(S101)}
\end{aligned}$$

Therefore, we obtain the desired conclusion. (S98) is a consequence of Markov inequality; (S99) comes from (S95) and the choice of t^*; (S100) is due to (S94); (S96) and (S97) lead to (S101).

In the rest of the proof, we are going to establish (S96) and then (S97) can be proved similarly. Define

$$S_1'(\rho,t^*) = \left\{i\in S_1(t^*): \rho\,|S_1(t^*)| \text{ indices in } S_1(t^*) \text{ with the smallest } \frac{(\theta_i^*-t^*)_+^2}{V_i(\theta^*)}\right\} \tag{S102}$$

for some small enough constant $\rho > 0$ to be specified later. That is, $S_1'(\rho,t^*)$ is a subset of $S_1(t^*)$ of size $\rho\,|S_1(t^*)|$ with the smallest $\frac{(\theta_i^*-t^*)_+^2}{V_i(\theta^*)}$ values. We remark that condition (S93) and (S94) necessarily imply $|S_1'(\rho,t^*)|\to\infty$ when ρ is a constant. We shall also assume $\rho\,|S_1'(\rho,t^*)|$ is an integer. Furthermore, note that the definition of $S_1'(\rho,t^*)$ implies:

$$R_1(S_1(t^*),\theta^*,t^*,-\delta)\geq R_1(S_1'(\rho,t^*),\theta^*,t^*,-\delta)\geq \rho R_1(S_1(t^*),\theta^*,t^*,-\delta). \tag{S103}$$

Therefore, to establish (S96), we only need to show

$$\mathbb{P}_{(\theta^*,r^*)}\left(\sum_{i\in S_1'(\rho,t^*)}\mathbb{I}\left\{\widehat{\theta}_i<t^*\right\}\geq C'R_1(S_1'(\rho,t^*),\theta^*,t^*,-\delta)\right)\geq 3/4. \tag{S104}$$

for some constant $C' > 0$. The remaining proof is then devoted to proving (S104).

Recall the definition of $\bar{\theta}$ in (62). Define $\tilde{\Delta}_i = (\theta_i^* - t^*)_+ \vee \alpha\sqrt{\frac{1}{npL}}$ where α is some large enough constant to be determined later. Define the event $\mathcal{F}_i$ as

$$\mathcal{F}_i = \left\{ |\widehat{\theta}_i - \bar{\theta}_i| \le \frac{\delta_0}{3}\tilde{\Delta}_i, \frac{|f^{(i)}(\theta_i^*|\widehat{\theta}_{-i}) - f^{(i)}(\theta_i^*|\theta_{-i}^*)|}{g^{(i)}(\theta_i^*|\theta_{-i}^*)} \le \frac{\delta_0}{3}\tilde{\Delta}_i, \frac{\left|g^{(i)}(\theta_i^*|\widehat{\theta}_{-i}) - g^{(i)}(\theta_i^*|\theta_{-i}^*)\right|}{g^{(i)}(\theta_i^*|\theta_{-i}^*)} \le \frac{\delta_0}{3} \right\}.$$

When $(\theta_i^* - t^*)_+^2 npL > \alpha$, using a similar argument that leads to (S83)-(S85), we can show that there exists some constant $\delta_0 > 0$, such that

$$\mathbb{P}_{(\theta^*, r^*)}(\mathcal{F}_i) \ge 1 - \left(O(n^{-7}) + \exp\left(-\tilde{\Delta}_i^2 npL \frac{np}{\log n} \right) + \exp\left(-\tilde{\Delta}_i^{3/2} npL \right) \right). \tag{S105}$$

When $(\theta_i^* - t^*)_+^2 npL \le \alpha$, we can show

$$\mathbb{P}_{(\theta^*, r^*)}(\mathcal{F}_i) \ge 1 - \left(O(n^{-7}) + e^{-(npL)^{1/4}} + e^{-\sqrt{\log n}} \right). \tag{S106}$$

instead. To establish it, we can choose $x = (npL)^{1/4}$ in (56) and $x = \sqrt{\log n}$ in (61) and then follow the same proof of (63), (64), and (66) as in the proof of Theorem 3.2. In both cases, this δ_0 can be made arbitrarily small by setting α large.

Assuming $\mathcal{F}_i$ is true, we can use arguments similar to the establishment of (67) to have

$$\mathbb{I}\left\{ \widehat{\theta}_i < t^* \right\} \ge \mathbb{I}\left\{ \frac{\sum_{j \in [n]\setminus\{i\}} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))}{\sum_{j \in [n]\setminus\{i\}} A_{ji}\psi'(\theta_j^* - \theta_i^*)} \le -(1 + \delta_0)(\theta_i^* - t^*)_+ \right\}.$$

Define the RHS of the above display as L_i. Then we have shown that

$$\sum_{i \in S_1'(\rho, t^*)} \mathbb{I}\left\{ \widehat{\theta}_i < t^* \right\} \ge \sum_{i \in S_1'(\rho, t^*)} L_i \mathbb{I}_{\mathcal{F}_i} \ge \sum_{i \in S_1'(\rho, t^*)} L_i - \sum_{i \in S_1'(\rho, t^*)} \mathbb{I}_{\mathcal{F}_i^c}. \tag{S107}$$

By (S105) and (S106), we have

$$\begin{aligned}
&\mathbb{E}\left(\sum_{i \in S_1'(\rho, t^*)} \mathbb{I}_{\mathcal{F}_i^c} \right) \\
&\le O(n^{-6}) + \sum_{i: i \in S_1'(\rho, t^*), (\theta_i^* - t^*)_+^2 npL > \alpha} \left(\exp\left(-\tilde{\Delta}^2 npL \frac{np}{\log n} \right) + \exp\left(-\tilde{\Delta}^{3/2} npL \right) \right) \\
&\quad + \sum_{i: i \in S_1'(\rho, t^*), (\theta_i^* - t^*)_+^2 npL \le \alpha} \left(\exp\left(-(npL)^{1/4} \right) + \exp\left(-\sqrt{\log n} \right) \right).
\end{aligned}$$

Using $\theta_i^* - t^* \le (\log n / np)^{1/4}$ for $i \in S_1(t^*)$ and $np/\log n \to \infty$, we see that the above bound is of smaller order than

$$n^{-5.9} + \sum_{i \in S_1'(\rho, t^*)} \exp\left[-\frac{\tilde{\Delta}_i^2 npL}{2\overline{V}_i(\theta^*)} \left(\left(\frac{np}{\log n} \right)^{1/9} \wedge (\log n)^{1/5} \right) \right],$$

and we can use Markov's inequality and obtain

$$\mathbb{P}_{(\theta^*,r^*)}\left(\sum_{i\in S_i'(t^*)}\mathbb{I}_{\mathcal{F}_i^c}\leq n^{-5.9}+\sum_{i\in S_1'(\rho,t^*)}\exp\left[-\frac{\tilde{\Delta}_i^2 npL}{2V_i(\theta^*)}\left(\left(\frac{np}{\log n}\right)^{1/9}\wedge(\log n)^{1/5}\right)\right]\right)\geq 1-o(1). \tag{S108}$$

Now to lower bound $\sum_{i\in S_1'(\rho,t^*)}L_i$, we define

$$\mathcal{A}=\Bigg\{A:\forall i\in S_1(t^*),\left|\frac{\sum_{j\in[n]\backslash\{i\}}A_{ij}\psi'(\theta_i^*-\theta_j^*)}{p\sum_{j\in[n]\backslash\{i\}}\psi'(\theta_i^*-\theta_j^*)}-1\right|\leq\delta_0, \tag{S109}$$

$$\left|\sum_{j\in S_1'(\rho,t^*)}A_{ji}\psi'(\theta_i^*-\theta_j^*)\right|\leq 2\rho kp+10\log n\Bigg\}. \tag{S110}$$

By Bernstein's inequality and union bound, we have $\mathbb{P}(A\in\mathcal{A})\geq 1-O(n^{-10})$. From now on, we use the notation $\mathbb{P}_A$ for the conditional probability $\mathbb{P}_{(\theta^*,r^*)}(\cdot|A)$ given A. For any $s>0$,

$$\mathbb{P}_{(\theta^*,r^*)}\left(\sum_{i\in S_1'(\rho,t^*)}L_i\geq s\right)\geq\mathbb{P}(A\in\mathcal{A})\inf_{A\in\mathcal{A}}\mathbb{P}_A\left(\sum_{i\in S_1'(\rho,t^*)}L_i\geq s\right). \tag{S111}$$

Now we study $\mathbb{P}_A\left(\sum_{i\in S_1'(\rho,t^*)}L_i\geq s\right)$. Define $S=[n]\backslash S_1'(\rho,t^*)$. Note that for each $i\in S_1'(\rho,t^*)$, we have $L_i\geq L_{i,1}-L_{i,2}-L_{i,3}$, where

$$L_{i,1}=\mathbb{I}\left\{\frac{\sum_{j\in S}A_{ji}(\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*))}{\sum_{j\in[n]\backslash\{i\}}A_{ji}\psi'(\theta_j^*-\theta_i^*)}\leq-(1+2\delta')(1+\delta_0)\tilde{\Delta}_i\right\},$$
$$L_{i,2}=\mathbb{I}\left\{\frac{\sum_{j\in S_1'(\rho,t^*):j<i}A_{ji}(\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*))}{\sum_{j\in[n]\backslash\{i\}}A_{ji}\psi'(\theta_j^*-\theta_i^*)}\geq\delta'(1+\delta_0)\tilde{\Delta}_i\right\},$$
$$L_{i,3}=\mathbb{I}\left\{\frac{\sum_{j\in S_1'(\rho,t^*):i<j}A_{ji}(\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*))}{\sum_{j\in[n]\backslash\{i\}}A_{ji}\psi'(\theta_j^*-\theta_i^*)}\geq\delta'(1+\delta_0)\tilde{\Delta}_i\right\}$$

for some small constant $\delta'>0$ whose value will be determined later. We are going to control each term separately.

(1). Analysis of $L_{i,1}$. Note that conditional on A, $\{L_{i,1}\}_{i\in S_1'(\rho,t^*)}$ are all independent Bernoulli random variables. We have $L_{i,1}\sim\text{Bernoulli}(p_i)$, where $p_i=\mathbb{E}_{(\theta^*,r^*)}(L_{i,1}|A)$. By Chebyshev's inequality, we have

$$\mathbb{P}_A\left(\sum_{i\in S_1'(\rho,t^*)}L_{i,1}\geq\frac{1}{2}\sum_{i\in S_1'(\rho,t^*)}p_i\right)\geq 1-\frac{4}{\sum_{i\in S_1'(\rho,t^*)}p_i}.$$

By Lemma C.2, we can lower bound each p_i by

$$p_i = \mathbb{P}_A \left(\frac{\sum_{j\in S} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))(1 + e^{\theta_j^* - \theta_i^*})}{\sum_{j\in[n]\setminus\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)} \le -(1+2\delta')(1+\delta_0)^2\tilde{\Delta}_i \right)$$
$$\ge C_1 \exp\left(-\frac{1+\delta_2}{2}\frac{\tilde{\Delta}_i^2 npL}{V_i(\theta^*)} - C_1'\sqrt{\frac{\tilde{\Delta}_i^2 npL}{V_i(\theta^*)}} \right),$$

for some constants $C_1, C_1' > 0$ and some small constant $\delta_2 > 0$. Note that δ_2 can be an arbitrarily small constant by making δ' and ρ small as well as making α large. Thus we can choose δ', ρ small enough and α large enough to let $\delta_2 < \delta/2$. Then we have

$$\sum_{i\in S_1'(\rho,t^*)} p_i \ge C_1 \sum_{i\in S_1'(\rho,t^*)} \exp\left(-\frac{1+\delta_2}{2}\frac{\tilde{\Delta}_i^2 npL}{V_i(\theta^*)} - C_1'\sqrt{\frac{\tilde{\Delta}_i^2 npL}{V_i(\theta^*)}} \right)$$
$$\ge C_1 R_1(S_1'(\rho,t^*), \theta^*, t^*, -\delta) \tag{S112}$$
$$\ge C_1 \rho R_1(S_1(t^*), \theta^*, t^*, -\delta). \tag{S113}$$

where (S112) can be achieved by setting α large and (S113) comes from (S103). As a result, under the condition (S93), we have $\sum_{i\in S_1'(\rho,t^*)} p_i \to \infty$.

Hence, we have proved

$$\inf_{A\in\mathcal{A}} \mathbb{P}_A \left(\sum_{i\in S_1'(\rho,t^*)} L_{i,1} \ge \frac{1}{2} C_1 \sum_{i\in S_1'(\rho,t^*)} \exp\left(-\frac{1+\delta_2}{2}\frac{\tilde{\Delta}_i^2 npL}{V_i(\theta^*)} - C_1'\sqrt{\frac{\tilde{\Delta}_i^2 npL}{V_i(\theta^*)}} \right) \right) \ge 1 - o(1).$$

(2). Analysis of $L_{i,2}$. By (S109)-(S110) and Bernstein's inequality, we can bound $\mathbb{E}(L_{i,2}|A)$ by

$$\exp\left(-\frac{\left(\delta'(1+\delta_0)^2\tilde{\Delta}_i L \sum_{j\in[n]\setminus\{i\}} A_{ji}\psi'(\theta_j^* - \theta_i^*)\right)^2}{2\left(L\sum_{j\in S_1'(\rho,t^*):j<i} A_{ji}\psi'(\theta_i^* - \theta_j^*) + \frac{1}{3}\delta'(1+\delta_0)^2\tilde{\Delta}_i L \sum_{j\in[n]\setminus\{i\}} A_{ji}\psi'(\theta_j^* - \theta_i^*)\right)} \right)$$
$$\le \exp\left(-\frac{\left(\delta'(1+\delta_0)^2\tilde{\Delta}_i L \sum_{j\in[n]\setminus\{i\}} p\psi'(\theta_j^* - \theta_i^*)\right)^2}{4\left(2L\rho kp + 10\log n + \frac{1}{3}\delta'(1+\delta_0)^2\tilde{\Delta}_i L \sum_{j\in[n]\setminus\{i\}} p\psi'(\theta_j^* - \theta_i^*)\right)} \right).$$

Now we set $\delta' = \rho^{1/8}$, and make ρ small enough to ensure (S113). Then, there exists some constants $C_2, C_3 > 0$ such that

$$\mathbb{E}(L_{i,2}|A) \le \exp\left(-C_2\rho^{-\frac{1}{2}} npL\tilde{\Delta}_i^2\right) \le \exp\left(-C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2V_i(\theta^*)} \right).$$

due to $\tilde{\Delta}_i = o(1)$ and $np/\log n \to \infty$. Then,

$$\mathbb{E}\left(\sum_{i\in S_1'(\rho,t^*)} L_{i,2} \middle| A \right) \le \sum_{i\in S_1'(\rho,t^*)} \exp\left(-C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2V_i(\theta^*)} \right).$$

By Markov inequality, we have

$$\inf_{A\in\mathcal{A}} \mathbb{P}_A\left(\sum_{i\in S_1'(\rho,t^*)} L_{i,2} \geq \sum_{i\in S_1'(\rho,t^*)} \exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2V_i(\theta^*)}\right)\right) \leq \frac{\sum_{i\in S_1'(\rho,t^*)}\exp\left(-C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2V_i(\theta^*)}\right)}{\sum_{i\in S_1'(\rho,t^*)}\exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2V_i(\theta^*)}\right)}. \tag{S114}$$

(3). Analysis of $L_{i,3}$. By a similar argument, we also have

$$\inf_{A\in\mathcal{A}} \mathbb{P}_A\left(\sum_{i\in S_1'(\rho,t^*)} L_{i,3} \geq \sum_{i\in S_1'(\rho,t^*)} \exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2V_i(\theta^*)}\right)\right) \leq \frac{\sum_{i\in S_1'(\rho,t^*)}\exp\left(-C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2V_i(\theta^*)}\right)}{\sum_{i\in S_1'(\rho,t^*)}\exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2V_i(\theta^*)}\right)}. \tag{S115}$$

Now we can combine the above analyses of $L_{i,1}$, $L_{i,2}$ and $L_{i,3}$. Since we are allowed to choose ρ to be an arbitrarily small constant, we shall make

$$\sum_{i\in S_1'(\rho,t^*)} \exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2V_i(\theta^*)}\right) \leq \frac{1}{8}C_1 \sum_{i\in S_1'(\rho,t^*)} \exp\left(-\frac{1+\delta_2}{2}\frac{\tilde{\Delta}_i^2 npL}{V_i(\theta^*)} - C_1'\sqrt{\frac{\tilde{\Delta}_i^2 npL}{V_i(\theta^*)}}\right)$$

and

$$\frac{\sum_{i\in S_1'(\rho,t^*)}\exp\left(-C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2V_i(\theta^*)}\right)}{\sum_{i\in S_1'(\rho,t^*)}\exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2V_i(\theta^*)}\right)} \leq \frac{1}{16}.$$

Thus, we have

$$\inf_{A\in\mathcal{A}} \mathbb{P}_A\left(\sum_{i\in S_1'(\rho,t^*)} L_i \geq C_4 \sum_{i\in S_1'(\rho,t^*)} \exp\left(-\frac{1+\delta_2}{2}\frac{\tilde{\Delta}_i^2 npL}{V_i(\theta^*)} - C_1'\sqrt{\frac{\tilde{\Delta}_i^2 npL}{V_i(\theta^*)}}\right)\right) \geq \frac{7}{8} - o(1), \tag{S116}$$

for some constant $C_4 > 0$. Then (S107), (S108), (S111) together with (S93) lead to

$$\mathbb{P}_{(\theta^*,r^*)}\left(\sum_{i\in S_1'(\rho,t^*)} \mathbb{I}\left\{\widehat{\theta}_i < t^*\right\} \geq \frac{C_4}{2} \sum_{i\in S_1'(\rho,t^*)} \exp\left(-\frac{1+\delta_2}{2}\frac{\tilde{\Delta}_i^2 npL}{V_i(\theta^*)} - C_1'\sqrt{\frac{\tilde{\Delta}_i^2 npL}{V_i(\theta^*)}}\right)\right) \geq \frac{7}{8} - o(1). \tag{S117}$$

Finally, (S104) follows from (S113) which completes the proof. □

We state Lemma C.2 to close this section. Its proof is essentially the same as the proof of Lemma A.4 and hence is omitted here.

Lemma C.2. *Assume $\frac{np}{\log n} \to \infty$, $\kappa = O(1)$. Recall the definition of $S_1'(\rho, t^*)$ in (S102), $S = [n]\backslash S_1'(\rho, t^*)$ and $\tilde{\Delta}_i = (\theta_i^* - t^*)_+ \vee \alpha\sqrt{\frac{1}{npL}}$. There exists some constants $C_1, C_2 > 0$ such that for any small constant $0.1 > \tilde{\delta} > 0$, there exists constant $\delta_1 > 0$ such that for* *any constant* $\alpha > 0$*,* $i \in S_1'(\rho, t^*)$*, any* $A \in \mathcal{A}$ *where* $\mathcal{A}$ *is defined in (S109)-(S110), any* $\theta^* \in \Theta(k, 0, \kappa)$ *and any* $r^* \in \mathfrak{S}_n$*, we have*

$$\mathbb{P}_{(\theta^*,r^*)}\left(\frac{\sum_{j\in S} A_{ji}(\bar{y}_{ij} - \psi(\theta^*_{r_i^*} - \theta^*_{r_j^*}))}{\sum_{j\in[n]\backslash\{i\}} A_{ji}\psi'(\theta^*_{r_j^*} - \theta^*_{r_i^*})} \leq -(1+\tilde{\delta})\tilde{\Delta}_i \middle| A\right)$$
$$\geq C_1 \exp\left(-\frac{1+\delta_1}{2}\frac{\tilde{\Delta}_i^2 npL}{V_{r_i^*}(\theta^*)} - C_2\sqrt{\frac{\tilde{\Delta}_i^2 npL}{V_{r_i^*}(\theta^*)}}\right). \tag{S118}$$

Moreover, δ_1 *is able to be arbitrarily small if* $\tilde{\delta}$ *and* ρ *are small enough.*

C.2 Proof of Theorem 7.2

We first give Lemma C.3 to characterize entrywise tail behaviors of the spectral method (7) which is crucial to the upper bound in Theorem 7.2.

Lemma C.3. *Assume* $\frac{np}{\log n} \to \infty$ *and* $\kappa = O(1)$*. Then, for the rank vector* $\widehat{r}$ *that is induced by the stationary distribution of the Markov chain (7), for any small constant* $0.1 > \delta > 0$*, there exists some constant* $C > 0$*, such that for any* $t \in \mathbb{R}$*, any* $\theta^* \in \Theta(k, 0, \kappa)$*,* $r^* \in \mathfrak{S}_n$*, we have*

$$\mathbb{P}_{(\theta^*,r^*)}\left(\widehat{\pi}_i \leq \frac{e^t}{\sum_{j\in[n]} e^{\theta^*_j}}\right) \leq C\exp\left(-\frac{(1-\delta)(\theta^*_{r_i^*} - t)_+^2 npL}{2\overline{V}_{r_i^*}(\theta^*)}\right) + Cn^{-4}, r_i^* \leq k; \tag{S119}$$

$$\mathbb{P}_{(\theta^*,r^*)}\left(\widehat{\pi}_i \geq \frac{e^t}{\sum_{j\in[n]} e^{\theta^*_j}}\right) \leq C\exp\left(-\frac{(1-\delta)(t - \theta^*_{r_i^*})_+^2 npL}{2\overline{V}_{r_i^*}(\theta^*)}\right) + Cn^{-4}, r_i^* \geq k+1 \tag{S120}$$

Proof. The proof follows the proof of Theorem 4.1 with slight modifications. Without loss of generality, we can assume $r_i^* = i$ for all $\in [n]$. Define $\bar{\Delta}_i$ as in (S82). We only need to prove (S119) since (S120) can be proved similarly.

Consider any $m \in [k]$. When $(\theta^*_m - t)_+^2 npL \leq c'$ for some large enough constant to be specified later, we can directly bound the probability using the trivial bound 1. Thus, we only need to consider the regime when $(\theta^*_m - t)_+^2 npL > c'$.

Following the proof of Theorem 4.1, we have (S3)-(S14) and (S12) hold. Note that we now have $\bar{\Delta}_m^2 Lnp > c'$ instead of $\bar{\Delta}_m^2 Lnp \to \infty$ which is needed in the proof of Theorem 4.1. As a consequence, we now have (S10) hold with $\delta = 4C_4 e^\kappa/\sqrt{c'}$ instead of some $o(1)$ as in the proof of Theorem 4.1. To sum up, with this δ, we have

$$\frac{|\widehat{\pi}_m - \bar{\pi}_m|}{\pi^*_m} \leq \delta(1 - e^{-\bar{\Delta}_m}), \tag{S121}$$

$$\left|\frac{\sum_{j\in[n]\backslash\{m\}} A_{jm}\bar{y}_{jm}}{\sum_{j\in[n]\backslash\{m\}} A_{jm}\psi(\theta^*_j - \theta^*_m)} - 1\right| \leq \delta, \tag{S122}$$

hold with probability at least $1 - O(n^{-4}) - \exp\left(-\bar{\Delta}_m^2 npL\frac{np}{\log n}\right) - \exp\left(-\bar{\Delta}_m^2 npL\sqrt{\frac{npL}{\log n}}\right)$. We can make δ to be an arbitrarily small constant by setting c' large as $\kappa = O(1)$.

Then for any $i \leq k$, by the same argument as in the proof of Theorem 4.1, we have

$$\begin{aligned}
&\mathbb{P}\left(\widehat{\pi}_i \leq \frac{e^t}{\sum_{j=1}^n e^{\theta_j^*}}\right) \\
= \ &\mathbb{P}\left(\frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} \leq e^{-(\theta_i^* - t)} - 1\right) \\
\leq \ &\mathbb{P}\left(\frac{\widehat{\pi}_i - \pi_i^*}{\pi_i^*} \leq e^{-\bar{\Delta}_i} - 1\right) \\
\leq \ &\mathbb{P}\left(\frac{\sum_{j\in[n]\backslash\{i\}} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))(1 + e^{\theta_j^* - \theta_i^*})}{\sum_{j\in[n]\backslash\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)} \leq -(1-\delta)^2(1 - e^{-\bar{\Delta}_i})\right) \\
&+O(n^{-4}) + \exp\left(-\bar{\Delta}_i^2 npL\frac{np}{\log n}\right) + \exp\left(-\bar{\Delta}_i^2 npL\sqrt{\frac{npL}{\log n}}\right),
\end{aligned}$$

which has the same upper bound as in (S14). We then have the same (S16) as in the proof of Theorem 4.1 which leads to

$$\begin{aligned}
&\mathbb{P}\left(\frac{\sum_{j\in[n]\backslash\{i\}} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))(1 + e^{\theta_j^* - \theta_i^*})}{\sum_{j\in[n]\backslash\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)} \leq -(1-\delta)^2(1 - e^{-\bar{\Delta}_i})\right) \\
\leq \ &\exp\left(-\frac{(1-o(1))Lp\bar{\Delta}_i^2\left(\sum_{j\in[n]\backslash\{i\}}\psi(\theta_j^* - \theta_i^*)\right)^2}{2\sum_{j\in[n]\backslash\{i\}}\psi'(\theta_i^* - \theta_j^*)\left(1 + e^{\theta_j^* - \theta_i^*}\right)^2}\right) + O(n^{-4}) \\
= \ &\exp\left(-\frac{(1-\delta_2)npL\bar{\Delta}_i^2}{2\overline{V}_i(\theta^*)}\right) + O(n^{-4}) \\
\leq \ &\exp\left(-\frac{(1-\delta_2)npL(\theta_i^* - t)_+^2}{2\overline{V}_i(\theta^*)}\right) + O(n^{-4})
\end{aligned}$$

with $\delta_1, \delta_2 > 0$ being some constant that can be arbitrarily small. The last inequality holds because when $\min\left((\theta_i^* - t)_+^2, \sqrt{\frac{\log n}{np}}\right) = \sqrt{\frac{\log n}{np}}$, the first term becomes $\exp\left(-\frac{(1-\delta_2)L\sqrt{np\log n}}{2\overline{V}_i(\theta^*)}\right)$, which can be absorbed by $O(n^{-4})$. Since $\exp\left(-\bar{\Delta}_i^2 npL\frac{np}{\log n}\right) + \exp\left(-\bar{\Delta}_i^2 npL\sqrt{\frac{npL}{\log n}}\right) \leq \exp\left(-\frac{(1-\delta_2)(\theta_i^* - t)_+^2 npL}{2\overline{V}_i(\theta^*)}\right) + O(n^{-4})$, we have

$$\mathbb{P}\left(\widehat{\pi}_i \leq \frac{e^t}{\sum_{j=1}^n e^{\theta_j^*}}\right) \leq 2\exp\left(-\frac{(1-\delta_2)(\theta_i^* - t)_+^2 npL}{2\overline{V}_i(\theta^*)}\right) + O(n^{-4}), \tag{S123}$$

for all $i \leq k$. The proof is complete. □

Proof of (32) of Theorem 7.2. The upper bound (32) is a straightforward consequence of Lemma C.3 in the same way as the proof of (29) of Theorem 7.1, and hence is omitted here. □

The rest of the section focuses on the lower bound (30). The proof follows the proof of Theorem 3.4 with some modification and is also very similar to the proof of (30) of Theorem 7.1. We include it below for completeness.

Proof of (33) of Theorem 7.2. To prove the lower bound (33), we are going to show

$$\mathbb{E}_{(\theta^*,r^*)}\mathsf{H}_k(\widehat{r},r^*) \gtrsim \frac{\overline{R}_1([k],\theta^*,t^*,-\delta)+\overline{R}_2([n]\backslash[k],\theta^*,t^*,-\delta)}{k} \tag{S124}$$

where t^* is the unique solution such that $\overline{R}_1([k],\theta^*,t^*,-\delta) = \overline{R}_2([n]\backslash[k],\theta^*,t^*,-\delta)$. The existence and uniqueness of t^* follow the same argument as in the proof of (30) of Theorem 7.1. Recall the definition of $S_1(t)$ in (S92). Since we assume $\inf_t(\overline{R}_1([k],\theta^*,t,-\delta) + \overline{R}_2([n]\backslash[k],\theta^*,t,-\delta)) \to \infty$, we have

$$\overline{R}_1([k],\theta^*,t^*,-\delta) \to \infty. \tag{S125}$$

The proof of (S124) follows the proof of Theorem 4.3. We will omit repeated details and only present the differences. Define

$$\overline{S}_1'(\rho,t^*) = \left\{ i\in S_1(t^*) : \rho\,|S_1(t^*)| \text{ indices in } S_1(t^*) \text{ with the smallest } \frac{(\theta_i^*-t^*)_+^2}{\overline{V}_i(\theta^*)}\right\} \tag{S126}$$

for some small enough constant $\rho > 0$ to be specified later. Following the same argument as in the proof of (30) of Theorem 7.1, we only need to show

$$\mathbb{P}_{(\theta^*,r^*)}\left(\sum_{i\in\overline{S}_1'(\rho,t^*)}\mathbb{I}\left\{\widehat{\pi}_i < t\right\} \geq C'\overline{R}_1(\overline{S}_1'(\rho,t^*),\theta^*,t^*,-\delta)\right) \geq 3/4. \tag{S127}$$

for some constant $C' > 0$. The remaining proof is then devoted to proving (S127).

Recall the definition of $\bar{\pi}$ in (S3). Define $\tilde{\Delta}_i = (\theta_i^*-t^*)_+ \vee \alpha\sqrt{\frac{1}{npL}}$ where α is some large enough constant to be determined later. Define the event $\overline{\mathcal{F}}_i$ as

$$\overline{\mathcal{F}}_i = \left\{\frac{|\widehat{\pi}_i-\bar{\pi}_i|}{\pi_i^*} \leq \delta_0(1-e^{-\tilde{\Delta}_i}) \text{ and } \left|\frac{\sum_{j\in[n]\backslash\{i\}}A_{ji}\bar{y}_{ji}}{\sum_{j\in[n]\backslash\{i\}}A_{ji}\psi(\theta_j^*-\theta_i^*)}-1\right| \leq \delta_0\right\}.$$

When $(\theta_i^*-t^*)_+^2 npL > \alpha$, using a similar argument that leads to (S121)-(S122), we can show that there exists some constant $\delta_0 > 0$, such that

$$\mathbb{P}_{(\theta^*,r^*)}(\overline{\mathcal{F}}_i) \geq 1-\left(O(n^{-4}) + \exp\left(-\tilde{\Delta}_i^2 npL\frac{np}{\log n}\right) + \exp\left(-\tilde{\Delta}_i^2 npL\sqrt{\frac{npL}{\log n}}\right)\right). \tag{S128}$$

When $(\theta_i^*-t^*)_+^2 npL \leq \alpha$, we can show

$$\mathbb{P}_{(\theta^*,r^*)}(\overline{\mathcal{F}}_i) \geq 1-\left(O(n^{-4}) + e^{-(np/\log n)^{1/2}} + e^{-\sqrt{\log n}}\right). \tag{S129}$$

instead. To establish it, we can choose $x = (np/\log n)^{1/2}$ in (S7) and $x = \sqrt{\log n}$ in (S8) and then follow the same proof of (S10) and (S12) as in the proof of Theorem 3.2. In both cases, this δ_0 can be made arbitrarily small by setting α large.

Assuming $\overline{\mathcal{F}}_i$ is true, we can use arguments similar to the establishment of (S28) to have

$$\mathbb{I}\{\widehat{\pi}_i < t\} \geq \mathbb{I}\left\{\frac{\sum_{j\in[n]\setminus\{i\}} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))(1+e^{\theta_j^*-\theta_i^*})}{\sum_{j\in[n]\setminus\{i\}} A_{ji}\psi(\theta_j^* - \theta_i^*)} \leq -(1+\delta_0)^2\alpha\sqrt{\frac{1}{npL}}\right\}. \tag{S130}$$

Define the RHS of the above display as $\overline{L}_i$.

$$\sum_{i\in\overline{S}_1'(\rho,t^*)} \mathbb{I}\{\widehat{\pi}_i < t\} \quad \geq \quad \sum_{i\in\overline{S}_1'(\rho,t^*)} \overline{L}_i \mathbb{I}_{\overline{\mathcal{F}}_i} \geq \sum_{i\in\overline{S}_1'(\rho,t^*)} \overline{L}_i - \sum_{i\in\overline{S}_1'(\rho,t^*)} \mathbb{I}_{\overline{\mathcal{F}}_i^c}. \tag{S131}$$

By (S128) and (S129), we have

$$\begin{aligned}
&\mathbb{E}\left(\sum_{i\in\overline{S}_1'(\rho,t^*)} \mathbb{I}_{\overline{\mathcal{F}}_i^c}\right) \\
&\leq O(n^{-3}) + \sum_{i:i\in\overline{S}_1'(\rho,t^*),(\theta_i^*-t^*)_+^2 npL>\alpha} \exp\left(-\tilde{\Delta}_i^2 npL\frac{np}{\log n}\right) + \exp\left(-\tilde{\Delta}_i^2 npL\sqrt{\frac{npL}{\log n}}\right) \\
&\quad + \sum_{i:i\in\overline{S}_1'(\rho,t^*),(\theta_i^*-t^*)_+^2 npL\leq\alpha} \exp\left(-(np/\log n)^{1/2}\right) + \exp\left(-\sqrt{\log n}\right).
\end{aligned}$$

Since the above bound is of smaller order than

$$n^{-2.9} + \sum_{i\in\overline{S}_1'(\rho,t^*)} \exp\left[-\frac{\tilde{\Delta}_i^2 npL}{2\overline{V}_i(\theta^*)}\left(\left(\frac{np}{\log n}\right)^{1/4} \wedge (\log n)^{1/4}\right)\right],$$

we can use Markov's inequality and obtain

$$\mathbb{P}_{(\theta^*,r^*)}\left(\sum_{i\in S_i'(t^*)} \mathbb{I}_{\overline{\mathcal{F}}_i^c} \leq n^{-2.9} + \sum_{i\in\overline{S}_1'(\rho,t^*)} \exp\left[-\frac{\tilde{\Delta}_i^2 npL}{2\overline{V}_i(\theta^*)}\left(\left(\frac{np}{\log n}\right)^{1/4} \wedge (\log n)^{1/4}\right)\right]\right) \geq 1-o(1). \tag{S132}$$

Now to lower bound $\sum_{i\in\overline{S}_1'(\rho,t^*)} \overline{L}_i$, we define

$$\overline{\mathcal{A}} = \left\{A : \forall i \in S_1(t^*), \left|\frac{\sum_{j\in[n]\setminus\{i\}} A_{ij}\psi'(\theta_i^*-\theta_j^*)\left(1+e^{\theta_j^*-\theta_i^*}\right)^2}{p\sum_{j\in[n]\setminus\{i\}}\psi'(\theta_i^*-\theta_j^*)\left(1+e^{\theta_j^*-\theta_i^*}\right)^2} - 1\right| \leq \delta_0,\right. \tag{S133}$$

$$\left|\frac{\sum_{j\in[n]\setminus\{i\}} A_{ji}\psi(\theta_j^*-\theta_i^*)}{p\sum_{j\in[n]\setminus\{i\}}\psi(\theta_j^*-\theta_i^*)} - 1\right| \leq \delta_0, \tag{S134}$$

$$\left.\left|\sum_{j\in\overline{S}_1'(\rho,t^*)} A_{ji}\psi'(\theta_i^*-\theta_j^*)(1+e^{\theta_j^*-\theta_i^*})^2\right| \leq 2\rho kp + 10\log n\right\}. \tag{S135}$$

By Bernstein's inequality and union bound, we have $\mathbb{P}(A \in \overline{\mathcal{A}}) \geq 1 - O(n^{-3})$. From now on, we use the notation $\mathbb{P}_A$ for the conditional probability $\mathbb{P}_{(\theta^*, r^*)}(\cdot|A)$ given A. For any $s > 0$,

$$\mathbb{P}_{(\theta^*, r^*)}\left(\sum_{i \in \overline{S}_1'(\rho, t^*)} \overline{L}_i \geq s \right) \geq \mathbb{P}(A \in \overline{\mathcal{A}}) \inf_{A \in \overline{\mathcal{A}}} \mathbb{P}_A \left(\sum_{i \in \overline{S}_1'(\rho, t^*)} \overline{L}_i \geq s \right). \tag{S136}$$

Now we study $\mathbb{P}_A \left(\sum_{i \in \overline{S}_1'(\rho, t^*)} L_i \geq s \right)$. Define $S = [n] \backslash \overline{S}_1'(\rho, t^*)$. Note that for each $i \in \overline{S}_1'(\rho, t^*)$, we have $L_i \geq L_{i,1} - L_{i,2} - L_{i,3}$, where

$$\overline{L}_{i,1} = \mathbb{I}\left\{ \frac{\sum_{j \in S} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))(1 + e^{\theta_j^* - \theta_i^*})}{\sum_{j \in [n] \backslash \{i\}} A_{ji} \psi(\theta_j^* - \theta_i^*)} \leq -(1 + 2\delta')(1 + \delta_0)^2 \tilde{\Delta}_i \right\},$$

$$\overline{L}_{i,2} = \mathbb{I}\left\{ \frac{\sum_{j \in \overline{S}_1'(\rho, t^*): j < i} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))(1 + e^{\theta_j^* - \theta_i^*})}{\sum_{j \in [n] \backslash \{i\}} A_{ji} \psi(\theta_j^* - \theta_i^*)} \geq \delta'(1 + \delta_0)^2 \tilde{\Delta}_i \right\},$$

$$\overline{L}_{i,3} = \mathbb{I}\left\{ \frac{\sum_{j \in \overline{S}_1'(\rho, t^*): i < j} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))(1 + e^{\theta_j^* - \theta_i^*})}{\sum_{j \in [n] \backslash \{i\}} A_{ji} \psi(\theta_j^* - \theta_i^*)} \geq \delta'(1 + \delta_0)^2 \tilde{\Delta}_i \right\}$$

for some small constant $\delta' > 0$ whose value will be determined later. We are going to control each term separately.

(1). Analysis of $\overline{L}_{i,1}$. Note that conditional on A, $\{\overline{L}_{i,1}\}_{i \in \overline{S}_1'(\rho, t^*)}$ are all independent Bernoulli random variables. We have $\overline{L}_{i,1} \sim \text{Bernoulli}(p_i)$, where $p_i = \mathbb{E}_{(\theta^*, r^*)}(\overline{L}_{i,1}|A)$. By Chebyshev's inequality, we have

$$\mathbb{P}_A \left(\sum_{i \in \overline{S}_1'(\rho, t^*)} \overline{L}_{i,1} \geq \frac{1}{2} \sum_{i \in \overline{S}_1'(\rho, t^*)} p_i \right) \geq 1 - \frac{4}{\sum_{i \in \overline{S}_1'(\rho, t^*)} p_i}.$$

By Lemma C.4, we can lower bound each p_i by

$$\begin{aligned} p_i &= \mathbb{P}_A \left(\frac{\sum_{j \in S} A_{ji}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))(1 + e^{\theta_j^* - \theta_i^*})}{\sum_{j \in [n] \backslash \{i\}} A_{ji} \psi(\theta_j^* - \theta_i^*)} \leq -(1 + 2\delta')(1 + \delta_0)^2 \tilde{\Delta}_i \right) \\ &\geq C_1 \exp \left(-\frac{1 + \delta_2}{2} \frac{\tilde{\Delta}_i^2 npL}{\overline{V}_i(\theta^*)} - C_1' \sqrt{\frac{\tilde{\Delta}_i^2 npL}{\overline{V}_i(\theta^*)}} \right), \end{aligned}$$

for some constants $C_1, C_1' > 0$ and some small constant $\delta_2 > 0$. Note that δ_2 can be an arbitrarily small constant by making δ' and ρ small as well as making α large. Thus we can choose δ', ρ small enough and α large enough to let $\delta_2 < \delta/2$. Then we have

$$\sum_{i \in \overline{S}_1'(\rho, t^*)} p_i \geq C_1 \sum_{i \in \overline{S}_1'(\rho, t^*)} \exp \left(-\frac{1 + \delta_2}{2} \frac{\tilde{\Delta}_i^2 npL}{\overline{V}_i(\theta^*)} - C_1' \sqrt{\frac{\tilde{\Delta}_i^2 npL}{\overline{V}_i(\theta^*)}} \right)$$

$$\geq C_1 \overline{R}_1(\overline{S}_1'(\rho, t^*), \theta^*, t^*, -\delta) \tag{S137}$$

$$\geq C_1 \rho \overline{R}_1(S_1(t^*), \theta^*, t^*, -\delta). \tag{S138}$$

by the same argument as in the proof of (30) of Theorem 7.1. As a result, under the condition (S125), we have $\sum_{i\in\overline{S}_1'(\rho,t^*)} p_i \to \infty$.

Hence, we have proved

$$\inf_{A\in\overline{\mathcal{A}}} \mathbb{P}_A\left(\sum_{i\in\overline{S}_1'(\rho,t^*)} \overline{L}_{i,1} \geq \frac{1}{2}C_1 \sum_{i\in\overline{S}_1'(\rho,t^*)} \exp\left(-\frac{1+\delta_2}{2}\frac{\tilde{\Delta}_i^2 npL}{\overline{V}_i(\theta^*)} - C_1'\sqrt{\frac{\tilde{\Delta}_i^2 npL}{\overline{V}_i(\theta^*)}}\right)\right) \geq 1-o(1).$$

(2). Analysis of $\overline{L}_{i,2}$. By (S133)-(S135) and Bernstein's inequality, we can bound $\mathbb{E}(\overline{L}_{i,2}|A)$ by

$$\exp\left(-\frac{\left(\delta'(1+\delta_0)^2\tilde{\Delta}_i L\sum_{j\in[n]\backslash\{i\}} A_{ji}\psi(\theta_j^*-\theta_i^*)\right)^2}{2\left(L\sum_{j\in\overline{S}_1'(\rho,t^*):j<i} A_{ji}\psi'(\theta_i^*-\theta_j^*)(1+e^{\theta_j^*-\theta_i^*})^2 + \frac{1}{3}\delta'(1+\delta_0)^2\tilde{\Delta}_i L\sum_{j\in[n]\backslash\{i\}} A_{ji}\psi(\theta_j^*-\theta_i^*)\right)}\right)$$

$$\leq \exp\left(-\frac{\left(\delta'(1+\delta_0)^2\tilde{\Delta}_i L\sum_{j\in[n]\backslash\{i\}} p\psi(\theta_j^*-\theta_i^*)\right)^2}{4\left(2L\rho kp + 10\log n + \frac{1}{3}\delta'(1+\delta_0)^2\tilde{\Delta}_i L\sum_{j\in[n]\backslash\{i\}} p\psi(\theta_j^*-\theta_i^*)\right)}\right).$$

Now we set $\delta' = \rho^{1/8}$, and make ρ small enough to ensure (S138). Then, there exists some constants $C_2, C_3 > 0$ such that

$$\mathbb{E}(\overline{L}_{i,2}|A) \leq \exp\left(-C_2\rho^{-\frac{1}{2}}npL\tilde{\Delta}_i^2\right) \leq \exp\left(-C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2\overline{V}_i(\theta^*)}\right).$$

Then,

$$\mathbb{E}\left(\sum_{i\in\overline{S}_1'(\rho,t^*)} \overline{L}_{i,2}\middle| A\right) \leq \sum_{i\in\overline{S}_1'(\rho,t^*)} \exp\left(-C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2\overline{V}_i(\theta^*)}\right).$$

By Markov inequality, we have

$$\inf_{A\in\overline{\mathcal{A}}} \mathbb{P}_A\left(\sum_{i\in\overline{S}_1'(\rho,t^*)} \overline{L}_{i,2} \geq \sum_{i\in\overline{S}_1'(\rho,t^*)} \exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2\overline{V}_i(\theta^*)}\right)\right) \leq \frac{\sum_{i\in\overline{S}_1'(\rho,t^*)} \exp\left(-C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2\overline{V}_i(\theta^*)}\right)}{\sum_{i\in\overline{S}_1'(\rho,t^*)} \exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2\overline{V}_i(\theta^*)}\right)}. \tag{S139}$$

(3). Analysis of $\overline{L}_{i,3}$. By a similar argument, we also have

$$\inf_{A\in\overline{\mathcal{A}}} \mathbb{P}_A\left(\sum_{i\in\overline{S}_1'(\rho,t^*)} \overline{L}_{i,3} \geq \sum_{i\in\overline{S}_1'(\rho,t^*)} \exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2\overline{V}_i(\theta^*)}\right)\right) \leq \frac{\sum_{i\in\overline{S}_1'(\rho,t^*)} \exp\left(-C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2\overline{V}_i(\theta^*)}\right)}{\sum_{i\in\overline{S}_1'(\rho,t^*)} \exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2\overline{V}_i(\theta^*)}\right)}. \tag{S140}$$

Now we can combine the above analyses of $\overline{L}_{i,1}$, $\overline{L}_{i,2}$ and $\overline{L}_{i,3}$. Since we are allowed to choose ρ to be an arbitrarily small constant, we shall make

$$\sum_{i\in\overline{S}_1'(\rho,t^*)} \exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2\overline{V}_i(\theta^*)}\right) \leq \frac{1}{8}C_1 \sum_{i\in\overline{S}_1'(\rho,t^*)} \exp\left(-\frac{1+\delta_2}{2}\frac{\tilde{\Delta}_i^2 npL}{\overline{V}_i(\theta^*)} - C_1'\sqrt{\frac{\tilde{\Delta}_i^2 npL}{\overline{V}_i(\theta^*)}}\right)$$

and

$$\frac{\sum_{i\in\overline{S}_1'(\rho,t^*)}\exp\left(-C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2\overline{V}_i(\theta^*)}\right)}{\sum_{i\in\overline{S}_1'(\rho,t^*)}\exp\left(-\frac{1}{2}C_3\rho^{-1/2}\frac{\tilde{\Delta}_i^2 npL}{2\overline{V}_i(\theta^*)}\right)}\leq\frac{1}{16}.$$

Thus, we have

$$\inf_{A\in\overline{\mathcal{A}}}\mathbb{P}_A\left(\sum_{i\in\overline{S}_1'(\rho,t^*)}\overline{L}_i\geq C_4\sum_{i\in\overline{S}_1'(\rho,t^*)}\exp\left(-\frac{1+\delta_2}{2}\frac{\tilde{\Delta}_i^2 npL}{\overline{V}_i(\theta^*)}-C_1'\sqrt{\frac{\tilde{\Delta}_i^2 npL}{\overline{V}_i(\theta^*)}}\right)\right)\geq\frac{7}{8}-o(1), \tag{S141}$$

for some constant $C_4>0$. Then (S131), (S132), (S136) together with (S125) lead to

$$\mathbb{P}_{(\theta^*,r^*)}\left(\sum_{i\in\overline{S}_1'(\rho,t^*)}\mathbb{I}\left\{\widehat{\pi}_i<t\right\}\geq\frac{C_4}{2}\sum_{i\in\overline{S}_1'(\rho,t^*)}\exp\left(-\frac{1+\delta_2}{2}\frac{\tilde{\Delta}_i^2 npL}{\overline{V}_i(\theta^*)}-C_1'\sqrt{\frac{\tilde{\Delta}_i^2 npL}{\overline{V}_i(\theta^*)}}\right)\right)\geq\frac{7}{8}-o(1). \tag{S142}$$

Finally, (S127) follows from (S138) which completes the proof. □

We state Lemma C.4 to close this section. Its proof is essentially the same as the proof of Lemma A.4 and hence is omitted here.

Lemma C.4. *Assume $\frac{np}{\log n}\to\infty$, $\kappa=O(1)$. Recall the definition of $\overline{S}_1'(\rho,t^*)$ in (S126), $S=[n]\backslash\overline{S}_1'(\rho,t^*)$ and $\tilde{\Delta}_i=(\theta_i^*-t^*)_+\vee\alpha\sqrt{\frac{1}{npL}}$. There exists some constants $C_1,C_2>0$ such that for any small constant $0.1>\tilde{\delta}>0$, there exists constant $\delta_1>0$ such that for any constant $\alpha>0$, $i\in S_1'(t^*)$, any $A\in\overline{\mathcal{A}}$ where $\overline{\mathcal{A}}$ is defined in (S133)-(S135), any $\theta^*\in\Theta(k,0,\kappa)$ and any $r^*\in\mathfrak{S}_n$, we have*

$$\begin{aligned}&\mathbb{P}_{(\theta^*,r^*)}\left(\frac{\sum_{j\in S}A_{ji}(\bar{y}_{ij}-\psi(\theta^*_{r_i^*}-\theta^*_{r_j^*}))(1+e^{\theta^*_{r_j^*}-\theta^*_{r_i^*}})}{\sum_{j\in[n]\backslash\{i\}}A_{ji}\psi(\theta^*_{r_j^*}-\theta^*_{r_i^*})}\leq-(1+\tilde{\delta})\tilde{\Delta}_i\middle|A\right)\\&\geq C_1\exp\left(-\frac{1+\delta_1}{2}\frac{\tilde{\Delta}_i^2 npL}{\overline{V}_{r_i^*}(\theta^*)}-C_2\sqrt{\frac{\tilde{\Delta}_i^2 npL}{\overline{V}_{r_i^*}(\theta^*)}}\right).\end{aligned} \tag{S143}$$

Moreover, δ_1 is able to be arbitrarily small if $\tilde{\delta}$ and ρ are small enough.

D Proof of Lemma 8.5

Define a gradient descent sequence

$$\theta^{(t+1)}=\theta^{(t)}-\eta\left(\nabla\ell_n(\theta^{(t)})+\lambda\theta^{(t)}\right). \tag{S144}$$

We also need to introduce a leave-one-out gradient descent sequence. Define

$$\begin{aligned}\ell_n^{(m)}(\theta) &= \sum_{1\leq i<j\leq n: i,j\neq m} A_{ij}\left[\bar{y}_{ij}\log\frac{1}{\psi(\theta_i-\theta_j)} + (1-\bar{y}_{ij})\log\frac{1}{1-\psi(\theta_i-\theta_j)}\right] \\ &\quad + \sum_{i\in[n]\backslash\{m\}} p\left[\psi(\theta_i^*-\theta_m^*)\log\frac{1}{\psi(\theta_i-\theta_m)} + \psi(\theta_m^*-\theta_i^*)\log\frac{1}{\psi(\theta_m-\theta_i)}\right].\end{aligned}$$

With the objective $\ell_n^{(m)}(\theta)$, we define

$$\theta^{(t+1,m)} = \theta^{(t,m)} - \eta\left(\nabla\ell_n^{(m)}(\theta^{(t,m)}) + \lambda\theta^{(t,m)}\right). \tag{S145}$$

We initialize both (S144) and (S145) by $\theta^{(0)} = \theta^{(0,m)} = \theta^*$ and use the same step size $\eta = \frac{1}{\lambda+np}$. Note that $\mathbb{1}_n^T\theta^* = 0$ implies $\mathbb{1}_n^T\theta^{(t)} = \mathbb{1}_n^T\theta^{(t,m)} = 0$ for all t. See Section 4.3 of [2]. We will establish the following bounds,

$$\max_{m\in[n]}\|\theta^{(t,m)} - \theta^{(t)}\| \leq 1, \tag{S146}$$

$$\|\theta^{(t)} - \theta^*\| \leq \sqrt{\frac{n}{\log n}}, \tag{S147}$$

$$\max_{m\in[n]}|\theta_m^{(t,m)} - \theta_m^*| \leq 1. \tag{S148}$$

It is obvious that (S146), (S147) and (S148) hold for $t=0$. We use a mathematical induction argument to show (S146), (S147) and (S148) for a general t. Let us suppose (S146), (S147) and (S148) are true, and we need to show the same conclusions continue to hold for $t+1$.

First, we have

$$\begin{aligned}\theta^{(t+1)} - \theta^{(t+1,m)} &= (1-\eta\lambda)(\theta^{(t)} - \theta^{(t,m)}) - \eta(\nabla\ell_n(\theta^{(t)}) - \nabla\ell_n^{(m)}(\theta^{(t,m)})) \\ &= ((1-\eta\lambda)I_n - \eta H(\xi))(\theta^{(t)} - \theta^{(t,m)}) - \eta\left(\nabla\ell_n(\theta^{(t,m)}) - \nabla\ell_n^{(m)}(\theta^{(t,m)})\right),\end{aligned}$$

where ξ is a convex combination of $\theta^{(t)}$ and $\theta^{(t,m)}$. By (S146) and (S148), we have

$$\|\theta^{(t)} - \theta^*\|_\infty \leq \max_{m\in[n]}\|\theta^{(t,m)} - \theta^{(t)}\| + \max_{m\in[n]}|\theta_m^{(t,m)} - \theta_m^*| \leq 2, \tag{S149}$$

and

$$\|\theta^{(t,m)} - \theta^*\|_\infty \leq \|\theta^{(t)} - \theta^*\|_\infty + \|\theta^{(t,m)} - \theta^{(t)}\| \leq 3. \tag{S150}$$

We thus have $\|\xi - \theta^*\|_\infty \leq 3$, and we can apply Lemma 8.3 to obtain the bound

$$\|((1-\eta\lambda)I_n - \eta H(\xi))(\theta^{(t)} - \theta^{(t,m)})\| \leq (1-\eta\lambda - c_1\eta np)\|\theta^{(t)} - \theta^{(t,m)}\|, \tag{S151}$$

for some constant $c_1 > 0$. We also note that

$$\begin{aligned}
&\|\nabla \ell_n(\theta^{(t,m)}) - \nabla \ell_n^{(m)}(\theta^{(t,m)})\|^2 \\
= &\left(\sum_{j\in[n]\setminus\{m\}} A_{jm}(\bar{y}_{jm} - \psi(\theta_j^* - \theta_m^*)) - \sum_{j\in[n]\setminus\{m\}} (A_{jm} - p)(\psi(\theta_j^{(t,m)} - \theta_m^{(t,m)}) - \psi(\theta_j^* - \theta_m^*)) \right)^2 \\
&+ \sum_{j\in[n]\setminus\{m\}} \Big(A_{jm}(\bar{y}_{jm} - \psi(\theta_j^* - \theta_m^*)) - (A_{jm} - p)(\psi(\theta_j^{(t,m)} - \theta_m^{(t,m)}) - \psi(\theta_j^* - \theta_m^*)) \Big)^2 \\
\le &\; C_1 \frac{np\log n}{L} + C_1 np \log n \|\theta^{(t,m)} - \theta^*\|_\infty^2, \qquad \text{(S152)}
\end{aligned}$$

for some constant $C_1 > 0$ by Lemma 8.2 and Lemma 8.4. We combine the two bounds (S151) and (S152), and obtain

$$\begin{aligned}
\|\theta^{(t+1)} - \theta^{(t+1,m)}\| &\le (1 - \eta\lambda - c_1\eta np)\|\theta^{(t)} - \theta^{(t,m)}\| + \eta\sqrt{C_1 np \log n \left(L^{-1} + \|\theta^{(t,m)} - \theta^*\|_\infty^2\right)} \\
&\le (1 - c_1\eta np) + \eta\sqrt{C_1 np\log n \,(L^{-1} + 9)} \qquad \text{(S153)} \\
&\le 1 \qquad \text{(S154)}
\end{aligned}$$

where the inequality (S153) is by (S146) and (S150). The inequality (S154) requires that $\sqrt{C_1 np \log n\,(L^{-1}+9)} \le c_1 np$, which is implied by the condition that $p \ge \frac{c_0 \log n}{n}$ for some sufficiently large $c_0 > 0$. We thus have proved (S146) for $t+1$.

Next, we have

$$\begin{aligned}
\theta^{(t+1)} - \theta^* &= \theta^{(t)} - \theta^* - \eta\left(\nabla \ell_n(\theta^{(t)}) + \lambda\theta^{(t)}\right) \\
&= (1-\eta\lambda)(\theta^{(t)} - \theta^*) - \eta\left(\nabla\ell_n(\theta^{(t)}) - \nabla\ell_n(\theta^*)\right) - \eta\lambda\theta^* - \eta\nabla\ell_n(\theta^*) \\
&= \left((1-\eta\lambda)I_n - \eta H(\xi)\right)(\theta^{(t)} - \theta^*) - \eta\lambda\theta^* - \eta\nabla\ell_n(\theta^*),
\end{aligned}$$

where ξ is abused for a vector that is a convex combination of $\theta^{(t)}$ and θ^*. Since by (S149) we get $\|\xi - \theta^*\|_\infty \le \|\theta^{(t)} - \theta^*\|_\infty \le 2$, we can use Lemma 8.3 to obtain the bound

$$\left((1-\eta\lambda)I_n - \eta H(\xi)\right)(\theta^{(t)} - \theta^*) \le (1 - \eta\lambda - c_2\eta np)\|\theta^{(t)} - \theta^*\|, \qquad \text{(S155)}$$

for some constant $c_2 > 0$. We also note that

$$\|\nabla\ell_n(\theta^*)\|^2 = \sum_{i=1}^n \left(\sum_{j\in[n]\setminus\{i\}} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)) \right)^2 \le C_2 \frac{n^2 p}{L}, \qquad \text{(S156)}$$

for some constant $C_2 > 0$ with high probability by Lemma 8.4. Combine the bounds (S155) and (S156), and we obtain

$$\begin{aligned}
\|\theta^{(t+1)} - \theta^*\| &\le (1 - \eta\lambda - c_2\eta np)\|\theta^{(t)} - \theta^*\| + \eta\sqrt{C_2\frac{n^2p}{L}} + \eta\lambda\|\theta^*\| \\
&\le (1 - c_2\eta np)\sqrt{\frac{n}{\log n}} + \eta\sqrt{C_2\frac{n^2p}{L}} + \eta\lambda\|\theta^*\| \\
&\le \sqrt{\frac{n}{\log n}},
\end{aligned}$$

where the last inequality is due to $\eta\sqrt{C_2\frac{n^2p}{L}}+\eta\lambda\|\theta^*\| \lesssim \frac{1}{\sqrt{Lp}}+\frac{1}{n^{3/2}p} = o\left(\eta np\sqrt{\frac{n}{\log n}}\right)$ by the choice of η and λ. Hence, (S147) holds for $t+1$.

Finally, we have

$$\begin{aligned}
\theta_m^{(t+1,m)}-\theta_m^* &= \theta_m^{(t,m)}-\theta_m^*+\eta p\sum_{j\in[n]\backslash\{m\}}\left(\psi(\theta_m^*-\theta_j^*)-\psi(\theta_m^{(t,m)}-\theta_j^{(t,m)})\right)-\lambda\eta\theta_m^{(t,m)} \\
&= \theta_m^{(t,m)}-\theta_m^*+\eta p\sum_{j\in[n]\backslash\{m\}}\psi'(\xi_j)(\theta_m^*-\theta_j^*-\theta_m^{(t,m)}+\theta_j^{(t,m)})-\lambda\eta\theta_m^{(t,m)} \\
&= \left(1-\eta\lambda-\eta p\sum_{j\in[n]\backslash\{m\}}\psi'(\xi_j)\right)(\theta_m^{(t,m)}-\theta_m^*)-\lambda\eta\theta_m^* \\
&\quad +\eta p\sum_{j\in[n]\backslash\{m\}}\psi'(\xi_j)(\theta_j^{(t,m)}-\theta_j^*),
\end{aligned}$$

where ξ_j is a scalar between $\theta_m^*-\theta_j^*$ and $\theta_m^{(t,m)}-\theta_j^{(t,m)}$. By (S150), we have $|\xi_j-\theta_m^*+\theta_j^*|\leq|\theta_m^*-\theta_j^*-\theta_m^{(t,m)}+\theta_j^{(t,m)}|\leq 6$, which implies $\|\xi\|_\infty$ is bounded. We then have $\sum_{j\in[n]\backslash\{m\}}\psi'(\xi_j)\geq c_3 n$ for some constant $c_3>0$, and thus

$$\left|\left(1-\eta\lambda-\eta p\sum_{j\in[n]\backslash\{m\}}\psi'(\xi_j)\right)(\theta_m^{(t,m)}-\theta_m^*)\right|\leq(1-\eta\lambda-c_3\eta np)|\theta_m^{(t,m)}-\theta_m^*|. \tag{S157}$$

We also have

$$\left|\sum_{j\in[n]\backslash\{m\}}\psi'(\xi_j)(\theta_j^{(t,m)}-\theta_j^*)\right|\leq\|\theta^{(t,m)}-\theta^*\|_1\leq\sqrt{n}\|\theta^{(t,m)}-\theta^*\|\leq\sqrt{n}\left(1+\sqrt{\frac{n}{\log n}}\right), \tag{S158}$$

where the last inequality is by (S146) and (S147). Combine the bounds (S157) and (S158), and we get

$$\begin{aligned}
|\theta_m^{(t+1,m)}-\theta_m^*| &\leq (1-\eta\lambda-c_3\eta np)|\theta_m^{(t,m)}-\theta_m^*|+\eta p\sqrt{n}\left(1+\sqrt{\frac{n}{\log n}}\right)+\lambda\eta|\theta_m^*| \\
&\leq (1-c_3\eta np)+\eta p\sqrt{n}+\eta p\frac{n}{\sqrt{\log n}}+\lambda\eta|\theta_m^*| \\
&\leq 1,
\end{aligned}$$

where the last inequality is because of $\eta p\sqrt{n}+\eta p\frac{n}{\sqrt{\log n}}+\lambda\eta|\theta_m^*| = o(\eta np)$ by the choice of η and λ. Hence, (S148) holds for $t+1$.

To summarize, we have shown that (S146), (S147) and (S148) hold for all $t\leq t^*$ with probability at least $1-O(t^*n^{-10})$. The reason why we have the probability $1-O(t^*n^{-10})$ is because we need to apply Lemma 8.2 with a different weight at each iteration to show (S152). Note that the bound (S149) holds for all $t\leq t^*$ as well and we thus have $\|\theta^{(t^*)}-\theta^*\|_\infty\leq 2$. With a standard optimization result for a strongly convex objective function, we have

$$\|\theta^{(t^*)}-\widehat{\theta}_\lambda\|\leq\left(1-\frac{\lambda}{\lambda+np}\right)^{t^*}\|\widehat{\theta}_\lambda-\theta^*\|.$$

See Lemma 6.7 of [3]. By triangle inequality, we have

$$\|\widehat{\theta}_\lambda - \theta^*\|_\infty \leq \|\theta^{(t^*)} - \widehat{\theta}_\lambda\| + \|\theta^{(t^*)} - \theta^*\|_\infty \leq \left(1 - \frac{\lambda}{\lambda + np}\right)^{t^*} \sqrt{n}\|\widehat{\theta}_\lambda - \theta^*\|_\infty + 2.$$

Since $\left(1 - \frac{\lambda}{\lambda+np}\right) \leq 1 - \frac{1}{1+n^2}$, we can take $t^* = n^3$ in order that $\left(1 - \frac{\lambda}{\lambda+np}\right)^{t^*} \sqrt{n} \leq \frac{1}{2}$. This implies $\|\widehat{\theta}_\lambda - \theta^*\|_\infty \leq 4$ with probability at least $1 - O(n^{-7})$ as desired.

E Proofs of Technical Lemmas

In this section, we prove Lemma 3.1, Lemma 8.1, Lemma 8.2, Lemma 8.3 and Lemma 8.4. We first list some additional technical results that will be needed in the proofs.

Lemma E.1 (Hoeffding's inequality)**.** *For independent random variables $X_1, \cdots, X_n$ that satisfy $a_i \leq X_i \leq b_i$, we have*

$$\mathbb{P}\left(\sum_{i=1}^n (X_i - \mathbb{E}X_i) \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right),$$

for any $t > 0$.

Lemma E.2 (Bernstein's inequality)**.** *For independent random variables $X_1, \cdots, X_n$ that satisfy $|X_i| \leq M$ and $\mathbb{E}X_i = 0$, we have*

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{\frac{1}{2}t^2}{\sum_{i=1}^n \mathbb{E}X_i^2 + \frac{1}{3}Mt}\right),$$

for any $t > 0$.

Lemma E.3 (Central limit theorem, Theorem 2.20 of [7])**.** *If $Z \sim N(0,1)$ and $W = \sum_{i=1}^n X_i$ where X_i are independent mean 0 and $\mathsf{Var}(W) = 1$, then*

$$\sup_t |\mathbb{P}(W \leq t) - \mathbb{P}(Z \leq t)| \leq 2\sqrt{3\sum_{i=1}^n \left(\mathbb{E}X_i^4\right)^{3/4}}.$$

Proof of Lemma 3.1. Without loss of generality, we consider $r_i^* = i$ so that $\theta_1^* \geq \cdots \geq \theta_n^*$. Then, we can write the loss as $2k\mathsf{H}_k(\widehat{r}, r^*) = \sum_{i=1}^k \mathbb{I}\{\widehat{r}_i > k\} + \sum_{i=k+1}^n \mathbb{I}\{\widehat{r}_i \leq k\}$. Since

$\widehat{r} \in \mathfrak{S}_n$, we must have $\sum_{i=1}^k \mathbb{I}\{\widehat{r}_i > k\} = \sum_{i=k+1}^n \mathbb{I}\{\widehat{r}_i \leq k\}$. This implies

$$\begin{aligned}
2k\mathsf{H}_k(\widehat{r}, r^*) &= 2\min\left(\sum_{i=1}^k \mathbb{I}\{\widehat{r}_i > k\}, \sum_{i=k+1}^n \mathbb{I}\{\widehat{r}_i \leq k\}\right) \\
&\leq 2\min\left(\sum_{i=1}^k \mathbb{I}\left\{\widehat{\theta}_i \leq \widehat{\theta}_{(k+1)}\right\}, \sum_{i=k+1}^n \mathbb{I}\left\{\widehat{\theta}_i \geq \widehat{\theta}_{(k)}\right\}\right) \\
&\leq 2\max_t \min\left(\sum_{i=1}^k \mathbb{I}\left\{\widehat{\theta}_i \leq t\right\}, \sum_{i=k+1}^n \mathbb{I}\left\{\widehat{\theta}_i \geq t\right\}\right) \qquad \text{(S159)} \\
&= 2\min_t \max\left(\sum_{i=1}^k \mathbb{I}\left\{\widehat{\theta}_i \leq t\right\}, \sum_{i=k+1}^n \mathbb{I}\left\{\widehat{\theta}_i \geq t\right\}\right) \qquad \text{(S160)} \\
&\leq 2\min_t \left(\sum_{i=1}^k \mathbb{I}\left\{\widehat{\theta}_i \leq t\right\} + \sum_{i=k+1}^n \mathbb{I}\left\{\widehat{\theta}_i \geq t\right\}\right).
\end{aligned}$$

The inequality (S159) uses the fact that $\widehat{\theta}_{(k)} \geq \widehat{\theta}_{(k+1)}$ where $\{\theta_{(i)}\}_{i=1}^n$ are the order statistics with $\widehat{\theta}_{(1)}$ being the largest and $\widehat{\theta}_{(n)}$ being the smallest. The equality (S160) holds since $\sum_{i=1}^k \mathbb{I}\left\{\widehat{\theta}_i \leq t\right\}$ is a nondecreasing function of t and $\sum_{i=k+1}^n \mathbb{I}\left\{\widehat{\theta}_i \geq t\right\}$ is a nonincreasing function of t. □

Proof of Lemma 8.1. The first conclusion is a direct consequence of Bernstein's inequality and a union bound argument. The second and third conclusion is a standard property of random graph Laplacian [8]. □

Proof of Lemma 8.2. To see the first conclusion, we note that $\mathbb{E}(A_{ij} - p)^2 \leq p$ and $\mathsf{Var}((A_{ij} - p)^2) \lesssim p$, and thus we can apply Bernstein's inequality followed by a union bound argument to obtain the desired result. The second conclusion is a direct consequence of Bernstein's inequality and a union bound argument. □

Proof of Lemma 8.3. For any $u \in \mathbb{R}^n$ such that $\mathbb{1}_n^T u = 0$,

$$u^T H(\theta) u = \sum_{1 \leq i < j \leq n} A_{ij} \psi(\theta_i - \theta_j)\psi(\theta_j - \theta_i)(u_i - u_j)^2.$$

Since $\psi(\theta_i - \theta_j)\psi(\theta_j - \theta_i) \geq \frac{1}{4}e^{-M}$, we have $\lambda_{\min,\perp}(H(\theta)) \geq \frac{1}{4}e^{-M}\lambda_{\min,\perp}(\mathcal{L}_A)$. By Lemma 8.1, we obtain the desired result. □

Proof of Lemma 8.4. Let $\mathcal{U} = \left\{u \in \mathbb{R}^n : \sum_{i \in [n]} u_i^2 \leq 1\right\}$ be the unit ball in $\mathbb{R}^n$. Then there exists a subset of $\mathcal{V} \subset \mathcal{U}$ such that for any $u \in \mathcal{U}$, there is a $v \in \mathcal{V}$ satisfying $\|u - v\| \leq 1/2$. Moreover, we also have $\log|\mathcal{V}| \leq C'n$ for some constant C'. See Lemma 5.2 of [9]. Then for any $u \in \mathcal{U}$, with the corresponding $v \in \mathcal{V}$, we have

$$\begin{aligned}
&\sum_{i=1}^n u_i \left(\sum_{j \in [n]\backslash\{i\}} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)) \right) \\
&= \sum_{i=1}^n v_i \left(\sum_{j \in [n]\backslash\{i\}} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)) \right) + \sum_{i=1}^n (u_i - v_i) \left(\sum_{j \in [n]\backslash\{i\}} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)) \right) \\
&\leq \sum_{i=1}^n v_i \left(\sum_{j \in [n]\backslash\{i\}} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)) \right) + \frac{1}{2}\sqrt{\sum_{i=1}^n \left(\sum_{j \in [n]\backslash\{i\}} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)) \right)^2}.
\end{aligned}$$

Maximize u and v on both sides of the inequality, after rearrangement, we have

$$\begin{aligned}
&\sqrt{\sum_{i=1}^n \left(\sum_{j \in [n]\backslash\{i\}} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)) \right)^2} \\
&\leq 2 \max_{v \in \mathcal{V}} \sum_{i=1}^n v_i \left(\sum_{j \in [n]\backslash\{i\}} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)) \right) \\
&= 2 \max_{v \in \mathcal{V}} \sum_{i<j} A_{ij}(v_i - v_j)(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)).
\end{aligned}$$

Conditional on A, applying Hoeffding's inequality and union bound on the last line, we have

$$\begin{aligned}
\sum_{i=1}^n \left(\sum_{j \in [n]\backslash\{i\}} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)) \right)^2 &\leq C'' \frac{(\log n + n) \max_{v \in \mathcal{V}} \sum_{i<j} A_{ij}(v_i - v_j)^2}{L} \\
&\leq C'' \frac{(\log n + n)\lambda_{\max}(\mathcal{L}_A)}{L}
\end{aligned}$$

with probability at least $1 - O(n^{-10})$. By Lemma 8.1, we obtain the desired bound for the first conclusion.

The second conclusion is a direct application of Hoeffding's inequality and a union bound argument.

The proof of the third conclusion is similar to that of the first one. Define $\mathcal{U}_i = \left\{ u \in \mathbb{R}^{n-1} : \sum_{j \in [n]\backslash\{i\}} A_{ij} u_j^2 \leq 1 \right\}$. Conditioning on A, one can think of $\mathcal{U}_i$ as a unit ball with dimension $\sum_{j \in [n]\backslash\{i\}} A_{ij} - 1$. Then, there exists a subset $\mathcal{V}_i \subset \mathcal{U}_i$ such that for any $u \in \mathcal{U}_i$, there is a $v \in \mathcal{V}_i$ that satisfies $\|u - v\| \leq \frac{1}{2}$. Moreover, we also have $\log |\mathcal{V}_i| \leq 2 \sum_{j \in [n]\backslash\{i\}} A_{ij}$ by Lemma 5.2 of [9]. For any $u \in \mathcal{U}_i$, with the corresponding $v \in \mathcal{V}_i$, following a similar argument of the proof of the first conclusion, we have

$$\sqrt{\sum_{j \in [n]\backslash\{i\}} A_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*))^2} \leq 2 \max_{v \in \mathcal{V}_i} \sum_{j \in [n]\backslash\{i\}} A_{ij} v_{ij}(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)),$$

which implies

$$\sqrt{\max_{i\in[n]}\sum_{j\in[n]\setminus\{i\}} A_{ij}(\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*))^2} \leq 2\max_{i\in[n]}\max_{v\in\mathcal{V}_i}\sum_{j\in[n]\setminus\{i\}} A_{ij}v_{ij}(\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)).$$

Applying Hoeffding's inequality and union bound, we have

$$\max_{i\in[n]}\sum_{j\in[n]\setminus\{i\}} A_{ij}(\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*))^2 \leq C_1\frac{\log n+\max_{i\in[n]}\sum_{j\in[n]\setminus\{i\}}A_{ij}}{L},$$

with probability at least $1-O(n^{-10})$. Finally, applying Lemma 8.1, we obtain the desired bound for the third conclusion, which concludes the proof. □

F Some Discussion on the Count Method

One of the simplest ranking methods is a count-based algorithm, often referred to as the Borda count [1] or the Copeland count [5] method. In this method, the players are ranked according to the number of games won. We will argue that this method is in general not optimal under the BTL model. Define

$$S_i=\sum_{j\in[n]\setminus\{i\}}A_{ij}\bar{y}_{ij},$$

and then LS_i is the number of games won by the ith player. The top-k set is determined by the players with the largest values of S_i's. To understand the condition of exact recovery, let us assume $r_i^*=i$ without loss of generality. We first compute the signal gap

$$\mathbb{E}S_k-\mathbb{E}S_{k+1}=p\sum_{j\in[n]\setminus\{k,k+1\}}\left(\psi(\theta_k^*-\theta_j^*)-\psi(\theta_{k+1}^*-\theta_j^*)\right)+p\psi(\theta_k^*-\theta_{k+1}^*)-p\psi(\theta_{k+1}^*-\theta_k^*).$$

Under the condition that $\theta_1^*-\theta_n^*\leq\kappa=O(1)$ and $\theta_k^*-\theta_{k+1}^*=\Delta$, we have

$$C_1\Delta\leq\psi(\theta_k^*-\theta_j^*)-\psi(\theta_{k+1}^*-\theta_j^*)\leq C_2\Delta,$$

for some constants $C_1,C_2>0$ for all j. Therefore,

$$\mathbb{E}S_k-\mathbb{E}S_{k+1}\asymp np\Delta,$$

which is the order of the signal gap. Next, we compute the variance,

$$\begin{aligned}
\mathsf{Var}(S_i) &= \sum_{j\in[n]\setminus\{i\}}\mathsf{Var}(A_{ij}\bar{y}_{ij})\\
&= \sum_{j\in[n]\setminus\{i\}}\left[\mathbb{E}\mathsf{Var}(A_{ij}\bar{y}_{ij}|A_{ij})+\mathsf{Var}(\mathbb{E}(A_{ij}\bar{y}_{ij}|A_{ij}))\right]\\
&= \sum_{j\in[n]\setminus\{i\}}\left(\frac{p\psi'(\theta_i^*-\theta_j^*)}{L}+\psi(\theta_i^*-\theta_j^*)p(1-p)\right)\\
&\asymp np(1-p).
\end{aligned}$$

In order that exact recovery is possible, it is necessary that the signal gap exceeds the standard deviation, which leads to the condition $np\Delta \gtrsim \sqrt{np(1-p)}$ (we have ignored the possible logarithmic factor due to a potential union bound argument). Suppose p is bounded away from 1, this condition becomes $\Delta \gtrsim \frac{1}{\sqrt{np}}$. In comparison, both the MLE and the spectral method achieve exact recovery under the condition $\Delta \gtrsim \frac{1}{\sqrt{npL}}$ (again, we ignore the logarithmic factor here). It is very clear that compared with the MLE or the spectral method, top-k ranking based on sums of wins $\{S_i\}$ does not even achieve the optimal rate. The condition $\Delta \gtrsim \frac{1}{\sqrt{np}}$ does not depend on L, which means increasing the number of games does not improve the accuracy of ranking.

To better illustrate the role of L, let us consider the extreme case $L = \infty$, which implies that $\bar{y}_{ij} = \psi(\theta_i^* - \theta_j^*)$ almost surely. In this situation, it is clear that both the MLE and the spectral method exactly recovers the top-k set without any error (the exact recovery condition becomes $\Delta > 0$). However, since

$$S_i = \sum_{j \in [n] \setminus \{i\}} A_{ij} \psi(\theta_i^* - \theta_j^*),$$

it is still possible that $S_k < S_{k+1}$ due to the randomness of $\{A_{ij}\}$, which leads to error. In fact, by Lindeberg's central limit theorem, the probability of $S_k < S_{k+1}$ is of constant order as long as $\Delta \lesssim \frac{1}{\sqrt{np}}$.

In addition to the number of games won, ranking based on the statistic of win ratio $\frac{\sum_{j \in [n] \setminus \{i\}} A_{ij} \bar{y}_{ij}}{\sum_{j \in [n] \setminus \{i\}} A_{ij}}$ suffers from a similar issue. Even when $L = \infty$, the average statistic becomes $\frac{\sum_{j \in [n] \setminus \{i\}} A_{ij} \psi(\theta_i^* - \theta_j^*)}{\sum_{j \in [n] \setminus \{i\}} A_{ij}}$ almost surely, which is still noisy. One can similarly show via the central limit theorem that

$$\frac{\sum_{j \in [n] \setminus \{k\}} A_{kj} \psi(\theta_k^* - \theta_j^*)}{\sum_{j \in [n] \setminus \{k\}} A_{kj}} < \frac{\sum_{j \in [n] \setminus \{k+1\}} A_{k+1j} \psi(\theta_{k+1}^* - \theta_j^*)}{\sum_{j \in [n] \setminus \{k+1\}} A_{k+1j}},$$

with a constant probability as long as $\Delta \lesssim \frac{1}{\sqrt{np}}$.

References

[1] Borda, J. d. [1784]. Mémoire sur les élections au scrutin, *Histoire de l'Academie Royale des Sciences pour 1781 (Paris, 1784)* .

[2] Chen, Y., Fan, J., Ma, C. and Wang, K. [2019a]. Additional proofs for the paper "spectral method and regularized mle are both optimal for top-k ranking", *The Annals of Statistics* .

[3] Chen, Y., Fan, J., Ma, C. and Wang, K. [2019b]. Spectral method and regularized mle are both optimal for top-k ranking, *The Annals of Statistics* **47**(4): 2204–2235.

[4] Chernozhukov, V., Chetverikov, D. and Kato, K. [2013]. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors, *The Annals of Statistics* **41**(6): 2786–2819.

[5] Copeland, A. [1951]. A ‘reasonable’social welfare function, seminar on mathematics in social sciences, university of michigan, *Cited indirectly from its mention by Luce and Raiffa (1957)* p. 358.

[6] Hartigan, J. [2014]. Bounding the maximum of dependent random variables, *Electronic Journal of Statistics* **8**(2): 3126–3140.

[7] Ross, S. M. and Peköz, E. A. [2007]. *A second course in probability*, www. Probability-Bookstore. com.

[8] Tropp, J. A. [2015]. An introduction to matrix concentration inequalities, *Foundations and Trends® in Machine Learning* **8**(1-2): 1–230.

[9] Vershynin, R. [2010]. Introduction to the non-asymptotic analysis of random matrices, *arXiv preprint arXiv:1011.3027* .