ℓ_∞-Bounds of the MLE in the BTL Model under General Comparison Graphs (Supplementary Material)

Wanshan Li[1] **Shamindra Shrotriya**[1] **Alessandro Rinaldo**[1]

[1]Department of Statistics & Data Science , Carnegie Mellon University , Pittsburgh, Pennsylvania, USA

A APPENDIX

The Appendix contains following parts:

A.1 Detail on comparison of results.

A.2 Proof in Section 2, i.e., upper bounds in Theorem 1, Corollary 2, Proposition 3 and 4.

A.3 Proof in Section 3, i.e., lower bounds in Theorem 5.

A.4 Additional experiments.

A.5 Other supporting results, including the proof for Lemma 6.

A.6 Special cases of comparison graphs, details related to Section 2.1.

A.7 Upper bound for unregularized/vanilla MLE.

A.1 COMPARISON OF RESULTS

This section is a complement to Section 2.1. We will summarize all existing works on the estimation error of Bradley-Terry model in two tables, and then compare our results with the results in Negahban et al. (2017); Agarwal et al. (2018); Hendrickx et al. (2020) in detail.

For simplicity, in Table 1 and Table 2 we use κ to replace B for results in Hajek et al. (2014); Shah et al. (2016) as $\kappa \asymp B$ when $\mathbf{1}^\top \boldsymbol{\theta}^* = 0$, and in Table 2 we omit the lower bound as they are usually in fairly complex forms.

In Negahban et al. (2017), they establish an ℓ_2 upper bound for $\|\hat{\boldsymbol{\pi}} - \tilde{\boldsymbol{\pi}}\|_2 / \|\tilde{\boldsymbol{\pi}}\|_2$ in the order of $\frac{e^{2.5\kappa}}{\lambda_2(\mathcal{L}_{rw})} \sqrt{\frac{n_{\max} \log n}{L}}$ where $\tilde{\pi}(i) := w_i / \sum_j w_j$ with $w_i = \exp(\theta_i)$, $\hat{\boldsymbol{\pi}}$ is the rank centrality estimamtor of $\tilde{\boldsymbol{\pi}}$, λ_2 refers to the second smallest eigenvalue, $\mathcal{L}_{rw} = D^{-1}A$ (which has the same spectrum as $D^{-1/2}\mathcal{L}_A D^{-1/2}$), and $\mathcal{L}_A = D - A$. Recall that our ℓ_2 upper bound is for $\|\hat{\theta} - \theta\|_2$ and the order is $\frac{e^{\kappa E}}{\lambda_2(\mathcal{L}_A)} \sqrt{\frac{n_{\max} n}{L}}$. We can now see that it's hard to give a general comparison between the two results because 1. for a general graph, there is no precise relationship between $\lambda_2(\mathcal{L}_{rw})$ and $\lambda_2(\mathcal{L}_A)$; 2. more importantly, for a general model parameter θ, there is no tight two-sided bound between $\|\hat{\theta} - \theta\|_2$ and $\|\pi - \tilde{\pi}\|_2 / \|\tilde{\pi}\|_2$. Although, it would be a very interesting future work to give a tight description of these two relevant pairs of quantities and make a meaningful comparison.

Agarwal et al. (2018) establish an ℓ_1-norm upper bound for the score parameter $\tilde{\pi}_i := w_i / \sum_j w_j$ with $w_i = \exp(\theta_i)$. Their bound is of the order $\frac{\eta e^{\kappa} n_{avg}}{\lambda_2(D^{-1}A) n_{\min}} \sqrt{\frac{\log n}{L}}$, where $n_{avg} = \sum_{i \in [n]} \tilde{\pi}_i n_i$, $D = \mathrm{diag}(n_1, \cdots, n_n)$, and $\eta := \log\left(\frac{n_{avg}}{n_{\min} \pi_{\min}}\right)$ with $\pi_{\min} = \min_{i \in [n]} \tilde{\pi}_i$.

In Hendrickx et al. (2020), they propose a novel weighted least square method to estimate vector w, with $w_i = \exp(\theta_i)$, and provide delicate theoretical analysis of their method. Their estimator shows a sharp upper bound for $\mathbb{E}[\sin^2(\hat{w}, w)]$ and equivalently for $\mathbb{E}\|\hat{w}/\|\hat{w}\|_2 - w/\|w\|_2\|_2^2$, in the sense that the upper bound for $\mathbb{E}[\sin^2(\hat{w}, w)]$ matches a instance-wise

Accepted for the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022).

Norm	Reference	Upper Bound
$\|\cdot\|_\infty$	Simons and Yao (1999)	$p=1, \quad \lesssim e^{\kappa}\sqrt{\frac{\log n}{nL}}$
	Yan et al. (2012)	$\lesssim e^{2\kappa}\frac{1}{p}\sqrt{\frac{\log n}{npL}}$
	Han et al. (2020)	$\lesssim e^{2\kappa}\sqrt{\frac{\log n}{np}} \cdot \frac{\log n}{\log(np)}$
	Chen et al. (2019), Chen et al. (2020)	$\lesssim e^{2\kappa}\sqrt{\frac{\log n}{npL}}$
	Our work	$\lesssim e^{2\kappa_E}\sqrt{\frac{\log n}{np^2L}}$
$\|\cdot\|_2^2$	Hajek et al. (2014)	$\lesssim e^{8\kappa}\frac{\log n}{pL}$
	Shah et al. (2016)	$\lesssim e^{8\kappa}\frac{\log n}{pL}, \quad \gtrsim e^{-2\kappa}\frac{1}{pL}$
	Negahban et al. (2017)	$\lesssim e^{4\kappa}\frac{\log n}{pL}, \quad \gtrsim e^{-\kappa}\frac{1}{pL}$
	Chen et al. (2019), Chen et al. (2020)	$\lesssim e^{2\kappa}\frac{1}{pL}$
$\sin^2(\cdot,\cdot)$ ($\hat{\mathbf{w}}$)	Hendrickx et al. (2020)	$\lesssim e^{2\kappa}\frac{1}{pL\|w\|_2^2}$
$\|\cdot\|_1$ ($\hat{\mathbf{w}}$)	Agarwal et al. (2018)	$\lesssim e^{\kappa}\sqrt{\frac{\log n}{L}}$

Table 1: Comparison of results under $ER(n,p)$ in literature.

Norm	Reference	Upper Bound
$\|\cdot\|_\infty$	Yan et al. (2012)	$\frac{e^{\kappa}}{\min_{i,j} n_{ij}}\sqrt{\frac{n_{\max}\log n}{L}}$
	Our work	$\frac{e^{2\kappa_E}}{\lambda_2(\mathcal{L}_{\mathbf{A}})}\frac{n_{\max}}{n_{\min}}\sqrt{\frac{n}{L}} + \frac{e^{\kappa_E}}{\lambda_2(\mathcal{L}_{\mathbf{A}})}\sqrt{\frac{n_{\max}\log n}{L}}$
$\|\cdot\|_2^2$	Hajek et al. (2014)	$\lesssim e^{8\kappa}\frac{\|E\|\log n}{\lambda_2(\mathcal{L}_A)^2 L}$
	Shah et al. (2016)	$\lesssim e^{8\kappa}\frac{n\log n}{\lambda_2(\mathcal{L}_A)L}$
	Our work	$\frac{e^{2\kappa_E}}{\lambda_2(\mathcal{L}_{\mathbf{A}})^2}\frac{n_{\max}n}{L}$
$\sin^2(\cdot,\cdot)$ ($\hat{\mathbf{w}}$)	Hendrickx et al. (2020)	$\lesssim e^{2\kappa}\frac{Tr(\mathcal{L}_A^\dagger)}{L\|w\|_2^2}$
$\|\cdot\|_1$ ($\hat{\mathbf{w}}$)	Agarwal et al. (2018)	$\lesssim \frac{\eta e^{\kappa} n_{avg}}{\lambda_2(D^{-1}A)n_{\min}}\sqrt{\frac{\log n}{L}}$

Table 2: Comparison of results for a fixed general comparison graph in literature.

lower bound up to constant factors. Such a universal sharp/optimal bound for general comparison graph (although the lower bound is not in the form of minimax rate) is unique in literature. For convenience of comparison, here we assume w_i's are $O(1)$ (otherwise we can put a factor e^{2B} in the bound $\frac{Tr(\mathcal{L}_A^\dagger)}{L}$). Then their upper bound is of the order $\frac{Tr(\mathcal{L}_A^\dagger)}{L\|w\|_2^2}$, where $\mathcal{L}_A^\dagger$ refers to the Moore-Penrose pseudo inverse of the graph Laplacian of the comparison graph. To correct for their different choice of metric, we need to multiply $\|w\|_2^2$ to their bound, and it becomes $\frac{Tr(\mathcal{L}_A^\dagger)}{L}$. On the other hand, the upper bound for expected ℓ_2 loss in Shah et al. (2016) is $\frac{n}{L\lambda_2(\mathcal{L}_A)}$. Since $\lambda_2(\mathcal{L}_A)$ is the smallest positive eigenvalue of $\mathcal{L}_A$, it holds that $Tr(\mathcal{L}_A^\dagger) < n/\lambda_2(\mathcal{L}_A)$, and hence the upper bound for expected error in Hendrickx et al. (2020) is tighter than the one in Shah et al. (2016). Although, it should be noted that the loss function in Hendrickx et al. (2020) is not directly comparable to a plain ℓ_2 loss, and we are just doing an approximate comparison.

In our paper, however, we provide a high probability bound for $\|\hat{\theta}-\theta\|_2^2$ in the order of $\frac{n_{\max}n}{L\lambda_2^2(\mathcal{L}_A)}$. It's usually hard to make a fair comparison between a high probability bound and a bound for expected metrics, but in Hajek et al. (2014); Shah et al. (2016), they also provide a high probability bound in equation (8b) of Theorem 2. As we discussed in Section 2.2 and Section 5, the ℓ_2 bound is not the primary focus of our paper, and although our theoretical analysis is not optimized for ℓ_2 error, our bound is still tighter than the bound in Hajek et al. (2014); Shah et al. (2016) for moderately dense and regular graphs, and is only worse for fairly sparse and irregular graphs.

A.2 PROOF IN SECTION 2

Proof of Theorem 1. We will use a gradient descent sequence defined recursively by $\theta^{(0)} = \theta^*$ and, for $t = 1, 2, \ldots$,

$$\theta^{(t+1)} = \theta^{(t)} - \eta[\nabla \ell_\rho(\theta^{(t)}) + \rho\theta^{(t)}].$$

Our proof builds heavily on the ideas and techniques developed by Chen et al. (2019) and further extended by Chen et al. (2020), and contains two key steps. In the first step, we control $\|\theta^{(T)} - \hat{\theta}_\rho\|$ for T large enough, by leveraging the convergence property of gradient descent for strong convex functions. In the second step, we control $\|\theta^{(T)} - \theta^*\|$ through a leave-one-out argument. The proof can be sketched as follows:

1. Bound $\|\theta^{(T)} - \hat{\theta}_\rho\|_\infty$, for large T using the linear convergence property of gradient descent for strongly-convex and smooth functions.
2. Bound $\|\theta^{(T)} - \theta^*\|_\infty$ for large T using the leave-one-out argument.
3. Finally, $\|\hat{\theta}_\rho - \theta^*\|_\infty$ is controlled by triangle inequality.

Step 1. Bound $\|\theta^{(T)} - \hat{\theta}_\rho\|_\infty$, for large T.

1. **Linear convergence, orthogonality to $\mathbf{1}_n$.** We say that a function ℓ is α-strongly convex if $\nabla^2 \ell(x) \succeq \alpha I_n$ and β-smooth if $\|\nabla \ell(x) - \nabla \ell(y)\|_2 \leq \beta \|x - y\|_2$ for all $x, y \in \mathrm{dom}(\ell)$. By Lemma 2, we know that $\ell_\rho(\cdot)$ is ρ-strongly convex and $(\rho + n_{\max})$-smooth. By Theorem 3.10 in Bubeck (2015), we have

   $$\|\theta^{(t)} - \hat{\theta}_\rho\|_2 \leq (1 - \frac{\rho}{\rho + n_{\max}})^t \|\theta^{(0)} - \hat{\theta}_\rho\|_2. \tag{1}$$

   Besides, as we start with θ^* that satisfies $1_n^\top \theta^* = 0$, it holds that $1_n^\top \theta^{(t)} = 0$ for all $t \geq 0$. To see this, just notice that

   $$\nabla \ell_\rho(\theta) = \rho\theta + \sum_{(i,j) \in E} [-\bar{y}_{ij} + \psi(\theta_i - \theta_j)](\boldsymbol{e}_i - \boldsymbol{e}_j)$$

   and $1_n^\top(\boldsymbol{e}_i - \boldsymbol{e}_j)$ for any i, j. Then by $1_n^\top \theta^{(t)} = 0, \forall t$ and (1), we have $1_n^\top \hat{\theta}_\rho = 0$.
2. **Control $\|\theta^* - \hat{\theta}_\rho\|_2$.** By a Taylor expansion, we have that

   $$\begin{aligned}\ell_\rho(\hat{\theta}_\rho; y) = \ell_\rho(\theta^*; y) &+ (\hat{\theta}_\rho - \theta^*)^\top \nabla \ell_\rho(\theta^*; y) \\ &+ \frac{1}{2}(\hat{\theta}_\rho - \theta^*)^\top \nabla^2 \ell_\rho(\xi; y)(\hat{\theta}_\rho - \theta^*),\end{aligned}$$

   where ξ is a convex combination of θ^* and $\hat{\theta}_\rho$. By Cauchy-Schwartz inequality,

   $$|(\hat{\theta}_\rho - \theta^*)^\top \nabla \ell_\rho(\theta^*; y)| \leq \|\nabla \ell_\rho(\theta^*; y)\|_2 \|\theta^* - \hat{\theta}_\rho\|_2.$$

   The two inequalities above and the fact that $\ell_\rho(\theta^*; y) \geq \ell_\rho(\hat{\theta}_\rho; y)$ yield that

   $$\|\theta^* - \hat{\theta}_\rho\|_2 \leq \frac{2\|\nabla \ell_\rho(\theta^*; y)\|_2}{\rho_{\min}(\nabla^2 \ell_\rho(\xi; y))}.$$

   By Lemma 1, $\|\nabla \ell_\rho(\theta^*; y)\|_2 \lesssim \sqrt{\frac{n_{\max}(n+r)}{L}}$. This fact, together with $\rho_{\min}(\nabla^2 \ell_\rho(\xi; y)) \geq \rho$ and $\rho \asymp \frac{1}{\kappa}\sqrt{\frac{n_{\max}}{L}}$, gives that $\|\theta^* - \hat{\theta}_\rho\|_2 \leq c\kappa\sqrt{n + r}$, for some $c > 0$.
3. **Bound $\|\theta^{(T)} - \hat{\theta}_\rho\|_2$.** Take $T = \lfloor \kappa^2 e^{3\kappa_E} n^6 \rfloor$ and remember that $L \leq \kappa^2 e^{4\kappa_E} n^8$. The previous two steps imply that

   $$\|\theta^{(T)} - \hat{\theta}_\rho\|_2 \leq c(1 - \frac{\rho}{\rho + n_{\max}})^T \kappa\sqrt{n + r} \leq c \exp\left(-\frac{T\rho}{\rho + n_{\max}}\right) \kappa\sqrt{n + r}.$$

   Let $\tilde{f}_d = \frac{e^{2\kappa_E}}{\lambda_2(\mathcal{L}_A)} \frac{n_{\max}}{n_{\min}} \sqrt{\frac{n+r}{L}} + \frac{e^{\kappa_E}}{\lambda_2(\mathcal{L}_A)} \sqrt{\frac{n_{\max}(\log n + r)}{L}}$ and consider inequality $e^{-g}\kappa\sqrt{n + r} \leq \tilde{f}_d$. The solution is given by $g \geq \log \kappa + \frac{1}{2}\log(n + r) - \log \tilde{f}_d$ and the inequality holds as long as $g \geq \kappa + 6 \log n + 3 \log \kappa$ since $L \leq \max\{1, \kappa\} e^{3\kappa_E} n^8$. Take $g = \kappa + 5 \log n + \log \kappa$, then as long as

   $$T\rho \geq 2n_{\max} n g,$$

it holds that $T\rho > \frac{1}{2}g(\rho + n_{\max})$, then $\|\theta^{(T)} - \hat{\theta}_\rho\|_2 \leq c\exp(-g)\kappa\sqrt{n+r}$ is smaller than $\tilde{C}_d\tilde{f}_d$ for some constant $\tilde{C}_d$. Since $T = \lfloor \kappa^2 e^{3\kappa_E} n^6 \rfloor$ and $\rho \geq \frac{c_\rho}{\kappa}\sqrt{\frac{n_{\max}}{\kappa^2 e^{4\kappa_E} n^8}}$, we have

$$T\rho \geq c_\rho \kappa e^{\kappa_E} n^2 \sqrt{n_{\max}} \geq 2n_{\max} n g.$$

In conclusion, we have

$$\|\theta^{(T)} - \hat{\theta}_\rho\|_\infty \leq \|\theta^{(T)} - \hat{\theta}_\rho\|_2 \leq \tilde{C}_d\tilde{f}_d.$$

The arguments above also hold with $\tilde{f}_a = \frac{e^{\kappa_E}}{\lambda_2(\mathcal{L}_A)}\sqrt{\frac{n_{\max}(n+r)}{L}}$, i.e., we have $\|\theta^{(T)} - \hat{\theta}_\rho\|_2 \leq \tilde{C}_a\tilde{f}_a$ for some constant $\tilde{C}_a$.

Step 2. Bound $\|\theta^{(T)} - \theta^*\|_\infty$ by a leave-one-out argument.

Denote $\psi(x) = \frac{1}{1+e^{-x}}$ and $r = \log\kappa + \kappa_E$, and define the leave-one-out negative log-likelihood as

$$\begin{aligned} \ell_n^{(m)}(\theta) = &\sum_{1\leq i<j\leq n: i,j\neq m} A_{ij}\left[\bar{y}_{ij}\log\frac{1}{\psi(\theta_i - \theta_j)} + (1-\bar{y}_{ij})\log\frac{1}{1-\psi(\theta_i-\theta_j)}\right] \\ &+ \sum_{j\in[n]\setminus\{m\}} A_{mj}\left[\psi\left(\theta_m^* - \theta_j^*\right)\log\frac{1}{\psi(\theta_m - \theta_j)} + \psi\left(\theta_j^* - \theta_m^*\right)\log\frac{1}{\psi(\theta_j - \theta_m)}\right], \end{aligned} \tag{2}$$

so the leave-one-out gradient descent sequence is, for $t = 0, 1, \ldots$,

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

We initialize both sequences by $\theta^{(0)} = \theta^{(0,m)} = \theta^*$ and use step size $\eta = \frac{1}{\rho + n_{\max}}$. By assumption 2, $\lambda_2(\mathcal{L}_\mathbf{A}) > 0$, so we can let $f_a = C_a\frac{e^{\kappa_E}}{\lambda_2(\mathcal{L}_A)}\left[\sqrt{\frac{n_{\max}(n+r)}{L}} + \rho\kappa(\theta^*)\sqrt{n}\right]$, $f_b = 10e^{\kappa_E}\frac{\sqrt{n_{\max}}}{n_{\min}}f_a$, $f_c = C_c\frac{e^{\kappa_E}}{\lambda_2(\mathcal{L}_A)}\sqrt{\frac{n_{\max}(\log n + r)}{L}}$, $f_d = f_b + f_c$ with sufficiently large constant $C_c > 0$ and $C_a \gg C_c$. By assumption 1, we have $f_c + f_d \leq 0.1$. We will show in Lemma 4, 5, 6, 7 that for all $0 \leq t \leq T = \lfloor \kappa^2 e^{3\kappa_E} n^6 \rfloor$

$$\begin{aligned} \|\theta^{(t)} - \theta^*\|_2 &\leq f_a, \\ \max_{m\in[n]} |\theta_m^{(t,m)} - \theta_m^*| &\leq f_b, \\ \max_{m\in[n]} \|\theta^{(t,m)} - \theta^{(t)}\|_2 &\leq f_c, \\ \|\theta^{(t)} - \theta^*\|_\infty &\leq f_d. \end{aligned} \tag{3}$$

When $t = 0$, (3) holds since $\theta^{(0)} = \theta^{(0,m)} = \theta^*$. By Lemma 4, 5, 6, 7, and a union bound, we know that (3) holds for all $0 \leq t \leq T = \lfloor \kappa^2 e^{3\kappa_E} n^6 \rfloor$ with probability at least $1 - O(n^{-4})$. Therefore, using the result in step 1, we have

$$\|\hat{\theta}_\rho - \theta^*\|_\infty \leq \|\hat{\theta}_\rho - \theta^{(T)}\|_\infty + \|\theta^{(T)} - \theta^*\|_\infty \leq 2f_d.$$

As a byproduct, we have

$$\|\hat{\theta}_\rho - \theta^*\|_2 \leq \|\hat{\theta}_\rho - \theta^{(T)}\|_2 + \|\theta^{(T)} - \theta^*\|_2 \leq 2f_a$$

□

Lemma 1. *With probability at least $1 - O(\kappa^{-2}e^{-3\kappa_E}n^{-10})$ the gradient of the regularized log-likelihood satisfies*

$$\|\nabla\ell_\rho(\theta^*)\|_2^2 \lesssim \frac{n_{\max}(n+r)}{L} + \rho\kappa(\theta^*)\sqrt{n}.$$

In particular, for $\rho \asymp \frac{1}{\kappa(\theta^*)}\sqrt{\frac{n_{\max}}{L}}$, we have $\|\nabla\ell_\rho(\theta^*)\|_2^2 \lesssim \frac{n_{\max}(n+r)}{L}$.

Proof. Triangle inequality gives

$$\|\nabla\ell_\rho(\theta^*)\|_2 \leq \|\nabla\ell_0(\theta^*)\|_2 + \rho\|\theta^*\|_2.$$

By definition of $\kappa(\theta^*)$, we have $\|\theta^*\|_2 \le \sqrt{n}\,\kappa(\theta^*)$. For the first term, by Lemma 8 we have

$$\|\nabla \ell_0(\theta^*)\|_2^2 = \sum_{i=1}^{n} \left[\sum_{j \in \mathcal{N}(i)} [\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)] \right]^2 \le C_1 \frac{n_{\max}(n+r)}{L}.$$

□

Lemma 2. *Let* $\kappa_E(x) = \max_{(i,j)\in E} |x_i - x_j|$*, then* $\forall \theta \in \mathbb{R}^n$*,*

$$\begin{aligned} \lambda_{\max}(\nabla^2 \ell_\rho(\theta; y)) &\le \rho + \frac{1}{2} n_{\max}, \\ \lambda_2(\nabla^2 \ell_\rho(\theta; y)) &\ge \rho + \frac{1}{4e^{\kappa_E(\theta)}} \lambda_2(\mathcal{L}_A). \end{aligned} \tag{4}$$

In particular, we have

$$\lambda_2(\nabla^2 \ell_\rho(\theta; y)) \ge \rho + \frac{1}{4e^{\kappa_E(\theta^*)} e^{2\|\theta - \theta^*\|_\infty}} \lambda_2(\mathcal{L}_A).$$

Proof. Use the fact that

$$\nabla^2 \ell_0(\theta; y) = \sum_{(i,j)\in E} \frac{e^{\theta_i} e^{\theta_j}}{(e^{\theta_i} + e^{\theta_j})^2} (\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^\top,$$

and $\forall (i,j) \in E$, $\frac{1}{4\exp(\kappa_E(\theta))} \le \frac{e^{\theta_i} e^{\theta_j}}{(e^{\theta_i} + e^{\theta_j})^2} \le \frac{1}{4}$, $\kappa_E(x_1) \le \kappa_E(x_2) + 2\|x_1 - x_2\|_\infty$. In addition, the largest eigenvalue of graph Laplacian satisfies (Corollary 3.9.2 in Brouwer and Haemers (2012))

$$\lambda_{\max}(\mathcal{L}_A) \le \max_{(i,j)\in E} (n_i + n_j) \le 2n_{\max}. \tag{5}$$

□

Lemma 3. *Provided that* (3) *holds, then*

$$\begin{aligned} \max_{m \in [n]} \|\theta^{(t+1,m)} - \theta^*\|_\infty &\le f_c + f_d, \\ \max_{m \in [n]} \|\theta^{(t+1,m)} - \theta^*\|_2 &\le f_c + f_a. \end{aligned} \tag{6}$$

Proof. By triangle inequality, we have

$$\begin{aligned} \max_{m \in [n]} \|\theta^{(t,m)} - \theta^*\|_\infty &\le \max_{m \in [n]} \|\theta^{(t,m)} - \theta^{(t)}\|_\infty + \|\theta^{(t)} - \theta^*\|_\infty \le f_c + f_d \\ \max_{m \in [n]} \|\theta^{(t,m)} - \theta^*\|_2 &\le \max_{m \in [n]} \|\theta^{(t,m)} - \theta^{(t)}\|_2 + \|\theta^{(t)} - \theta^*\|_2 \le f_c + f_a. \end{aligned}$$

□

Lemma 4. *Suppose* (3) *holds, and the step size satisfies* $0 < \eta \le \frac{1}{\rho + n_{\max}}$*. If*

$$f_d \le 0.1 \text{ and } f_a \ge C_a \frac{e^{\kappa_E(\theta^*)}}{\lambda_2(\mathcal{L}_A)} \left[\sqrt{\frac{n_{\max}(n+r)}{L}} + \rho\kappa(\theta^*)\sqrt{n} \right],$$

for some large constant C_a*, then with probability at least* $1 - O(\kappa^{-2} e^{-3\kappa_E} n^{-10})$ *we have*

$$\|\theta^{(t+1)} - \theta^*\|_2 \le f_a.$$

Proof. By the form of the gradient descent, we have that

$$\begin{aligned}\theta^{(t+1)} - \theta^* &= \theta^{(t)} - \eta\nabla\ell_\rho(\theta^{(t)}) - \theta^* \\ &= \theta^{(t)} - \eta\nabla\ell_\rho(\theta^{(t)}) - [\theta^* - \eta\nabla\ell_\rho(\theta^*)] - \eta\nabla\ell_\rho(\theta^*) \\ &= \left[I_n - \eta\int_0^1 \nabla^2\ell_\rho(\theta(\tau))\mathrm{d}\tau\right](\theta^{(t)} - \theta^*) - \eta\nabla\ell_\rho(\theta^*),\end{aligned}$$

where $\theta(\tau) = \theta^* + \tau(\theta^{(t)} - \theta^*)$. Letting $H = \int_0^1 \nabla^2\ell_\rho(\theta(\tau))\mathrm{d}\tau$, by the triangle inequality,

$$\|\theta^{(t+1)} - \theta^*\|_2 \le \|(I_n - \eta H)(\theta^{(t)} - \theta^*)\|_2 + \eta\|\nabla\ell_\rho(\theta^*)\|_2. \tag{7}$$

Setting $\kappa_E(x) = \max_{(i,j)\in E}|x_i - x_j|$, then, for sufficiently small ϵ, we have that

$$\kappa_E(\theta(\tau)) \le \kappa_E(\theta^*) + 2\|\theta^{(t)} - \theta^*\|_\infty \le \kappa_E(\theta^*) + \epsilon. \tag{8}$$

as long as

$$2f_d \le \epsilon. \tag{9}$$

Then, by Lemma 2 and setting $\epsilon = 0.2$, for any $\tau \in [0,1]$,

$$\rho + \frac{\lambda_2(\mathcal{L}_A)}{10e^{\kappa_E(\theta^*)}} \le \rho + \frac{\lambda_2(\mathcal{L}_A)}{8e^{\kappa_E(\theta^*)}e^\epsilon} \le \lambda_2(\nabla^2\ell_\rho(\theta(\tau))) \le \lambda_{\max}(\nabla^2\ell_\rho(\theta(\tau))) \le \rho + \frac{1}{2}n_{\max}. \tag{10}$$

Since $1_n^\top(\theta^{(t)} - \theta^*) = 0$, we obtain that

$$\|(I_n - \eta H)(\theta^{(t)} - \theta^*)\|_2 \le \max\{|1 - \eta\lambda_2(H)|, |1 - \eta\lambda_{\max}(H)|\}\|\theta^{(t)} - \theta^*\|_2. \tag{11}$$

By (10) and the fact that $\eta \le \frac{1}{\rho + n_{\max}}$, we get

$$\|(I_n - \eta H)(\theta^{(t)} - \theta^*)\|_2 \le (1 - \frac{\eta\lambda_2(\mathcal{L}_A)}{10e^{\kappa_E(\theta^*)}})\|\theta^{(t)} - \theta^*\|_2. \tag{12}$$

By Lemma 1 and the induction hypothesis, we have

$$\|\theta^{(t+1)} - \theta^*\|_2 \le (1 - \frac{\eta\lambda_2(\mathcal{L}_A)}{10e^{\kappa_E(\theta^*)}})f_a + C\eta\left[\sqrt{\frac{n_{\max}(n+r)}{L}} + \rho\kappa(\theta^*)\sqrt{n}\right] \le f_a \tag{13}$$

as long as

$$f_a \ge C_a\frac{e^{\kappa_E(\theta^*)}}{\lambda_2(\mathcal{L}_A)}\left[\sqrt{\frac{n_{\max}(n+r)}{L}} + \rho\kappa(\theta^*)\sqrt{n}\right] \tag{14}$$

for some large constant C_a. □

Lemma 5. *Suppose* (3) *holds and assume that*

1. $f_a = C_a\frac{e^{\kappa_E(\theta^*)}}{\lambda_2(\mathcal{L}_A)}\left[\sqrt{\frac{n_{\max}(n+r)}{L}} + \rho\kappa(\theta^*)\sqrt{n}\right]$, $f_c = C_c\frac{e^{\kappa_E(\theta^*)}}{\lambda_2(\mathcal{L}_A)}\sqrt{\frac{n_{\max}(\log n + r)}{L}}$ *with* $C_a \gg C_c$.
2. $\frac{n_{\min}}{10e^{\kappa_E(\theta^*)}}f_b \ge \frac{3\sqrt{n_{\max}}}{4}f_a$.
3. $f_c + f_d \le 0.1$.

then as long as the step size satisfies $0 < \eta \le \frac{1}{\rho + n_{\max}}$, *with probability at least* $1 - O(\kappa^{-2}e^{-3\kappa_E}n^{-10})$ *we have*

$$\max_{m\in[n]}\left|\theta_m^{(t+1,m)} - \theta_m^*\right| \le f_b.$$

Proof. Recall that the gradient descent step for leave-one-out estimator $\theta^{(m)}$ is defined as

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

where

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

Direct calculations give

$$\begin{aligned} [\nabla \ell_n^{(m)}(\theta)]_m = & \sum_{j \in [n] \setminus \{m\}} A_{mj} \left[\psi(\theta_m^* - \theta_j^*)(\psi(\theta_m^* - \theta_j^*) - 1) + (1 - \psi(\theta_m^* - \theta_j^*))\psi(\theta_m - \theta_j) \right] \\ = & \sum_{j \in [n] \setminus \{m\}} A_{mj} \left[-\psi(\theta_m^* - \theta_j^*) + \psi(\theta_m - \theta_j) \right]. \end{aligned}$$

Thus, we have

$$\theta_m^{(t+1,m)} - \theta_m^* = \left(1 - \eta\rho - \eta \sum_{j \in [n] \setminus \{m\}} A_{mj} \psi'(\xi_j) \right) (\theta_m^{(t,m)} - \theta_m^*) - \rho\eta\theta_m^* + \eta \sum_{j \in [n] \setminus \{m\}} A_{mj} \psi'(\xi_j)(\theta_j^{(t,m)} - \theta_j^*),$$

where ξ_j is a scalar between $\theta_m^* - \theta_j^*$ and $\theta_m^{(t,m)} - \theta_j^{(t,m)}$. Notice that $\psi'(x) = \frac{e^x}{(1+e^x)^2} \le \frac{1}{4}$ for any $c \in \mathbb{R}$, thus by Cauchy-Schwartz inequality we have

$$| \sum_{j \in [n] \setminus \{m\}} A_{mj} \psi'(\xi_j)(\theta_j^{(t,m)} - \theta_j^*)| \le \frac{1}{4} \sqrt{n_{\max}} \|\theta^{(t,m)} - \theta^*\|_2. \tag{15}$$

Also, since $\eta \le \frac{1}{\rho + n_{\max}}$,

$$1 - \eta\rho - \eta \sum_{j \in [n] \setminus \{m\}} A_{mj} \psi'(\xi_j) \ge 1 - \eta\rho - \eta \frac{n_{\max}}{4} \ge 0.$$

Therefore,

$$0 \le 1 - \eta\rho - \eta \sum_{j \in [n] \setminus \{m\}} A_{mj} \psi'(\xi_j) \le 1 - \eta n_{\min} \min_{j \in \mathcal{N}(m)} \psi'(\xi_j).$$

Since ξ_j is a scalar between $\theta_m^* - \theta_j^*$ and $\theta_m^{(t,m)} - \theta_j^{(t,m)}$, we have

$$\begin{aligned} \max_{j \in \mathcal{N}(m)} |\xi_j| \le & \max_{j \in \mathcal{N}(m)} |\theta_m^* - \theta_j^*| + \max_{j \in \mathcal{N}(m)} |\theta_m^* - \theta_j^* - (\theta_m^{(t,m)} - \theta_j^{(t,m)})| \\ \le & \kappa_E(\theta^*) + 2\|\theta^{(t,m)} - \theta^*\|_\infty \le \kappa_E(\theta^*) + \epsilon \end{aligned}$$

as long as

$$\|\theta^{(t,m)} - \theta^*\|_\infty \le f_c + f_d \le \epsilon/2. \tag{16}$$

Let $\epsilon = 0.2$, then $e^\epsilon \le 5/4$ and

$$\psi'(\xi_j) = \frac{e^{\xi_j}}{(1 + e^{\xi_j})^2} = \frac{e^{-|\xi_j|}}{(1 + e^{-|\xi_j|})^2} \ge \frac{e^{-|\xi_j|}}{4} \ge \frac{1}{4e^{\epsilon + \kappa_E(\theta^*)}} \ge \frac{1}{5e^{\kappa_E(\theta^*)}}.$$

By triangle inequality we get

$$\begin{aligned} |\theta_m^{(t+1,m)} - \theta_m^*| \le & \left(1 - \frac{\eta n_{\min}}{10e^{\kappa_E(\theta^*)}} \right) |\theta_m^{(t,m)} - \theta_m^*| + \rho\eta\|\theta^*\|_\infty + \frac{\eta\sqrt{n_{\max}}}{4} \|\theta^{(t,m)} - \theta^*\|_2 \\ \le & f_b - \frac{\eta n_{\min}}{10e^{\kappa_E(\theta^*)}} f_b + \eta\rho\kappa(\theta^*) + \eta\frac{\sqrt{n_{\max}}}{4}(f_a + f_c) \le f_b \end{aligned} \tag{17}$$

as long as

$$\frac{n_{\min}}{10e^{\kappa_E(\theta^*)}}f_b \geq \rho\kappa(\theta^*) + \frac{\sqrt{n_{\max}}}{4}(f_a + f_c).$$

By assumption, $f_a = C_a \frac{e^{\kappa_E(\theta^*)}}{\lambda_2(\mathcal{L}_A)}\left[\sqrt{\frac{n_{\max}(n+r)}{L}} + \rho\kappa(\theta^*)\sqrt{n}\right]$, $f_c = C_c \frac{e^{\kappa_E(\theta^*)}}{\lambda_2(\mathcal{L}_A)}\sqrt{\frac{n_{\max}(\log n+r)}{L}}$ with $C_a \gg \max\{C_c, 1\}$, so

$$f_a \gg f_c, \text{ and } \frac{\sqrt{n_{\max}}}{4}f_a \gg \frac{n_{\max}}{\lambda_2(\mathcal{L}_A)}\left[\sqrt{\frac{n+r}{L}} + \sqrt{\frac{n}{n_{\max}}}\rho\kappa(\theta^*)\right] \geq \rho\kappa(\theta^*).$$

Therefore, a sufficient condition for $|\theta_m^{(t+1,m)} - \theta_m^*| \leq f_b$ is

$$\frac{n_{\min}}{10e^{\kappa_E(\theta^*)}}f_b \geq \frac{3\sqrt{n_{\max}}}{4}f_a,$$

which is satisfied by our assumption. □

Lemma 6. *Suppose* (3) *holds with* $f_c = C_c \frac{e^{\kappa_E(\theta^*)}}{\lambda_2(\mathcal{L}_A)}\sqrt{\frac{n_{\max}(\log n+r)}{L}}$ *for some sufficiently large constant* C_c*,* $f_d = f_b + f_c$*, then as long as the step size satisfies* $0 < \eta \leq \frac{1}{\rho + n_{\max}}$*, with probability at least* $1 - O(\kappa^{-2}e^{-3\kappa_E}n^{-10})$ *we have*

$$\max_{m\in[n]} \|\theta^{(t+1,m)} - \theta^{(t)}\|_2 \leq f_c.$$

Proof. By the update rules, we have

$$\begin{aligned}
\theta^{(t+1)} - \theta^{(t+1,m)} =& \theta^{(t)} - \eta\nabla\ell_\rho(\theta^{(t)}) - \left[\theta^{(t,m)} - \eta\nabla\ell_\rho^{(m)}(\theta^{(t,m)})\right] \\
=& \theta^{(t)} - \eta\nabla\ell_\rho(\theta^{(t)}) - \left[\theta^{(t,m)} - \eta\nabla\ell_\rho(\theta^{(t,m)})\right] \\
& - \eta\left[\nabla\ell_\rho(\theta^{(t,m)}) - \nabla\ell_\rho^{(m)}(\theta^{(t,m)})\right] \\
=& v_1 - v_2,
\end{aligned}$$

where

$$v_1 = \left[I_n - \eta\int_0^1 \nabla^2\ell_\rho(\theta(\tau))\mathrm{d}\tau\right](\theta^{(t)} - \theta^{(t,m)}), \quad v_2 = \eta\left[\nabla\ell_\rho(\theta^{(t,m)}) - \nabla\ell_\rho^{(m)}(\theta^{(t,m)})\right]$$

Now following the same arguments towards (12), as long as $\eta \leq \frac{1}{\rho + n_{\max}}$, we can get

$$\|v_1\|_2 \leq (1 - \frac{\eta\lambda_2(\mathcal{L}_A)}{10e^{\kappa_E(\theta^*)}})\|\theta^{(t)} - \theta^{(t,m)}\|_2.$$

For v_2, we know that

$$\begin{aligned}
\frac{1}{\eta}v_2 =& \sum_{i\in[n]\setminus\{m\}} \left\{A_{mi}\left[\psi(\theta_i^{(t,m)} - \theta_m^{(t,m)}) - \bar{y}_{im}\right] - A_{mi}\left[\psi(\theta_i^{(t,m)} - \theta_m^{(t,m)}) - \psi(\theta_i^* - \theta_m^*)\right]\right\}(\boldsymbol{e}_i - \boldsymbol{e}_m) \\
=& \sum_{i\in[n]\setminus\{m\}} A_{mi}\left[\psi(\theta_i^* - \theta_m^*) - \bar{y}_{im}\right](\boldsymbol{e}_i - \boldsymbol{e}_m).
\end{aligned}$$

By the form of the derivatives and Lemma 8, we know that with probability at least $1 - O(n^{\kappa^{-2}e^{-3\kappa_E}n^{-10}})$,

$$\begin{aligned}
\|\frac{1}{\eta}v_2\|_2^2 =& \left[\sum_{i\in[n]\setminus\{m\}} A_{im}\left(\bar{y}_{im} - \psi\left(\theta_i^* - \theta_m^*\right)\right)\right]^2 + \sum_{i\in[n]\setminus\{m\}} A_{im}\left(\bar{y}_{im} - \psi\left(\theta_i^* - \theta_m^*\right)\right)^2 \\
\lesssim& \frac{n_{\max}(\log n + r)}{L} + \frac{\log n + n_{\max} + r}{L} \lesssim \frac{n_{\max}(\log n + r)}{L}.
\end{aligned} \tag{18}$$

Therefore, we have

$$\begin{aligned}\|\theta^{(t+1)} - \theta^{(t+1,m)}\|_2 &\leq \|v_1\|_2 + \|v_2\|_2 \\ &\leq (1 - \frac{\eta\lambda_2(\mathcal{L}_A)}{10e^{\kappa_E(\theta^*)}})\|\theta^{(t)} - \theta^{(t,m)}\|_2 + C\eta\sqrt{\frac{n_{\max}(\log n + r)}{L}} \\ &\leq (1 - \frac{\eta\lambda_2(\mathcal{L}_A)}{10e^{\kappa_E(\theta^*)}})f_c + C\eta\sqrt{\frac{n_{\max}(\log n + r)}{L}} \leq f_c,\end{aligned} \tag{19}$$

where the last inequality is due to the fact that C_c is a sufficiently large constant by our assumption and

$$\frac{\eta\lambda_2(\mathcal{L}_A)}{30e^{\kappa_E(\theta^*)}}f_c \geq C\eta\sqrt{\frac{n_{\max}(\log n + r)}{L}} \Leftarrow f_c = C_c\frac{e^{\kappa_E(\theta^*)}}{\lambda_2(\mathcal{L}_A)}\sqrt{\frac{n_{\max}(\log n + r)}{L}}. \tag{20}$$

□

Lemma 7. *Suppose* (3) *holds and $f_d \geq f_b + f_c$, then with probability at least $1 - O(\kappa^{-2}e^{-3\kappa_E}n^{-10})$ we have*

$$\|\theta^{(t+1)} - \theta^*\|_\infty \leq f_d.$$

Proof. By Lemma 5 and Lemma 6 we have

$$\begin{aligned}|\theta_m^{(t+1)} - \theta_m^*| &\leq |\theta_m^{(t+1)} - \theta_m^{(t+1,m)}| + |\theta_m^{(t+1,m)} - \theta_m^*| \\ &\leq \|\theta^{(t+1)} - \theta^{(t+1,m)}\|_2 + |\ \theta_m^{(t+1,m)} - \theta_m^*| \leq f_c + f_b \leq f_d,\end{aligned}$$

since $f_d \geq f_b + f_c$ by our assumption. □

Lemma 8. *With probability at least $1 - O(\kappa^{-2}e^{-3\kappa_E}n^{-10})$ it holds that*

$$\begin{aligned}\max_{i\in[n]}\left[\sum_{j\in\mathcal{N}(i)}[\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)]\right]^2 &\leq C\frac{(\log n + r)\cdot n_{\max}}{L}, \\ \sum_{i=1}^n\left[\sum_{j\in\mathcal{N}(i)}[\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)]\right]^2 &\leq C\frac{(n + r)\cdot n_{\max}}{L}. \\ \max_{i\in[n]}\sum_{j\in\mathcal{N}(i)}[\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)]^2 &\leq C\frac{\log n + n_{\max} + r}{L},\end{aligned} \tag{21}$$

where $r = \log\kappa + \kappa_E$.

Proof. To prove the first inequality, notice that by Hoeffding's inequality we have

$$\mathbb{P}\left(\sum_{j<i}A_{ij}[\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)] \geq \sqrt{\frac{8n_{\max}(\log n + r)}{L}}\right) \leq 2\exp(-\frac{2L}{n_{\max}}\cdot\frac{8n_{\max}(\log n + r)}{L}) = 2\kappa^{-2}e^{-3\kappa_E}n^{-12},$$

where $r = \log\kappa + \kappa_E$. By union bound we know that on an event B with probability at least $1 - \kappa^{-2}e^{-3\kappa_E}n^{-10}$ we have $\forall i \in [n]$, $\left[\sum_{j<i}A_{ij}[\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)]\right]^2 < \frac{8n_{\max}(\log n + r)}{L}$.

Next we prove the second inequality. Consider the unit ball $\mathcal{S} = \{v \in \mathbb{R}^n : \sum_{i\in[n]}v_i^2 = 1\}$ in $\mathbb{R}^n$. By Lemma 5.2 of Vershynin (2011), we can pick a subset $\mathcal{U} \subset \mathcal{S}$ so that $\log|\mathcal{U}| \leq cn$ and for any $v \in \mathcal{S}$, there exists a vector $u \in \mathcal{U}$ such that $\|u - v\|_2 \leq \frac{1}{2}$. For a given $v \in \mathcal{S}$, pick $u \in \mathcal{U}$ such that $\|u - v\|_2 \leq \frac{1}{2}$ and we have

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

Taking maximum over v and the left hand side can achieve $\sqrt{\sum_{i=1}^{n}\left[\sum_{j\in\mathcal{N}(i)}[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)]\right]^2}$, thus we have

$$\begin{aligned}\sqrt{\sum_{i=1}^{n}\left[\sum_{j\in\mathcal{N}(i)}[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)]\right]^2} &\leq 2\max_{u\in\mathcal{U}}\sum_{i=1}^{n}u_i\left[\sum_{j\in\mathcal{N}(i)}[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)]\right]\\ &= 2\max_{u\in\mathcal{U}}\sum_{i<j}A_{ij}(u_i-u_j)\left[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)\right].\end{aligned}$$

To apply Hoeffding's inequality and union bound, we should account for $|\mathcal{U}|\leq e^{cn}$, so we can get that with probability at least $1-\kappa^{-2}e^{-3\kappa_E}n^{-10}$ it holds that

$$\begin{aligned}\sum_{i=1}^{n}\left[\sum_{j\in\mathcal{N}(i)}[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)]\right]^2 &\leq C_1\frac{1}{L}\left[(\log n+n+r)\sum_{i<j}A_{ij}(u_i-u_j)^2\right]\\ &\leq C_1\frac{(\log n+n+r)\lambda_{\max}(\mathcal{L}_A)}{L},\end{aligned}$$

where $r=\log\kappa+\kappa_E$. Since $\lambda_{\max}(\mathcal{L}_A)\leq 2n_{\max}$, the second inequality is proved.

Next we prove the third inequality. For each $i\in[n]$, let $\mathcal{V}_i:=\{v\in\mathbb{R}^{n-1}:\sum_{j\neq i}A_{ij}v_j^2\leq 1\}$. By Lemma 5.2 of Vershynin (2011), we can pick a subset $\mathcal{U}_i\subset\mathcal{V}_i$ so that $\log|\mathcal{U}_i|\leq 2\sum_{j\neq i}A_{ij}$ and for any $v\in\mathcal{V}_i$, there exists a vector $u\in\mathcal{U}_i$ such that $\|u-v\|_2\leq\frac{1}{2}$. For a given $v\in\mathcal{V}_i$, pick $u\in\mathcal{U}_i$ such that $\|u-v\|_2\leq\frac{1}{2}$ and we have

$$\begin{aligned}\sum_{j\in\mathcal{N}(i)}v_{ij}[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)] &= \sum_{j\in\mathcal{N}(i)}u_{ij}[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)]+\sum_{j\in\mathcal{N}(i)}(v_{ij}-u_{ij})[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)]\\ &\leq \sum_{j\in\mathcal{N}(i)}u_{ij}[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)]+\frac{1}{2}\sqrt{\sum_{j\in\mathcal{N}(i)}[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)]^2}\end{aligned}$$

Taking maximum over v and the left hand side can achieve $\sqrt{\sum_{j\in\mathcal{N}(i)}[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)]^2}$, thus we have

$$\sqrt{\sum_{j\in\mathcal{N}(i)}[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)]^2}\leq 2\max_{u\in\mathcal{U}_i}\sum_{j\in\mathcal{N}(i)}u_{ij}[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)].$$

Therefore,

$$\sqrt{\max_{i\in[n]}\sum_{j\in\mathcal{N}(i)}[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)]^2}\leq 2\max_{i\in[n]}\max_{u\in\mathcal{U}_i}\sum_{j\in\mathcal{N}(i)}u_{ij}[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)].$$

Now a straightforward application of Hoeffding's inequality and union bound gives

$$\max_{i\in[n]}\sum_{j\in\mathcal{N}(i)}[\bar{y}_{ij}-\psi(\theta_i^*-\theta_j^*)]^2\leq C\frac{1}{L}\left[\log n+n_{\max}+\kappa_E+\log\kappa\right]$$

with probability at least $1-O(\kappa^{-2}e^{-3\kappa_E}n^{-10})$. □

Corollary 2 (Erdös-Rényi graph)**.** Suppose that the comparison graph comes from an Erdös-Rényi graph $ER(n,p)$. Assume that $\mathbf{1}_n^\top\boldsymbol{\theta}^*=0$, $\kappa\leq n$, $\kappa_E\leq\log n$, $L\leq n^8e^{4\kappa_E}\max\{1,\kappa\}$, $np>C_1\log n$, and $L\geq C_2\,\max\{1,\kappa\}e^{4\kappa_E}n/\log^2 n$ for some sufficiently large constants $C_1,C_2>0$. Set $\rho=c_\rho/\kappa\sqrt{n_{max}/L}$. Then $\mathbf{1}_n^\top\hat{\boldsymbol{\theta}}_\rho=0$, and with probability at least $1-O(n^{-4})$, it holds that

$$\begin{aligned}\|\hat{\boldsymbol{\theta}}_\rho-\boldsymbol{\theta}^*\|_\infty &\lesssim e^{2\kappa_E}\sqrt{\frac{1}{np^2L}}+e^{\kappa_E}\sqrt{\frac{\log n}{npL}},\\ \|\hat{\boldsymbol{\theta}}_\rho-\boldsymbol{\theta}^*\|_2 &\lesssim e^{\kappa_E}\sqrt{\frac{1}{pL}}.\end{aligned}\tag{22}$$

Proof of Corollary 2. For an $ER(n,p)$ graph $\mathcal{G}$ with $p \geq c\frac{\log n}{n}$ for some larege $c > 0$, it holds with probability at least $1 - O(n^{-10})$ that $\mathcal{G}$ is connected, and

$$\frac{1}{2}np \leq n_{\min} \leq n_{\max} \leq 2np,$$

and

$$\lambda_2\left(\mathcal{L}_A\right) = \min_{u \neq 0 : \mathbf{1}_n^\top u = 0} \frac{u^\top \mathcal{L}_A u}{\|u\|^2} \geq \frac{np}{2}.$$

The proof can be seen in either Chen et al. (2019) or Chen et al. (2020). Thus, by a union bound, we can replace the corresponding quantities in upper bounds in Theorem 1 and get the high probability bounds in Corollary 2. □

Proposition 3 (Path graph)**.** Suppose the comparison graph is a path graph $([n], E)$ with $E = \{(i, i+1)\}_{i \in [n-1]}$ and for each i, item i and item $i+1$ are compared $L_{i,i+1}$ times such that $\min_i L_{i,i+1} > ce^{2\kappa_E} n \log n$ for some universal constant c, then with probability at least $1 - n^{-4}$, the vanilla MLE $\hat{\boldsymbol{\theta}}_0$ satisfies

$$\|\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}^*\|_\infty \lesssim \sqrt{\sum_{i=1}^{n-1} \frac{\exp(2|\theta_i^* - \theta_{i+1}^*|)\log n}{L_{i,i+1}}}. \tag{23}$$

In particular, when $L_{i,i+1} = L$ for all $i \in [n-1]$, we have

$$\|\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}^*\|_\infty \lesssim e^{\kappa_E}\sqrt{\frac{n \log n}{L}}, \|\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}^*\|_2 \lesssim e^{\kappa_E} n \sqrt{\frac{\log n}{L}} \tag{24}$$

Proof of Proposition 3. Consider a path graph with edge set $E = \{(i, i+1) : i \in [n-1]\}$. Let M_{ij} be the number of wins of item i against item j. For the ease of notations, in this proof we use $\hat{\theta}$ to denote the vanilla MLE $\hat{\boldsymbol{\theta}}_0$. We know that

$$\nabla \ell\left(\theta\right)_i = \sum_{j \in \mathcal{N}(i)} \left(\frac{M_{ij}}{M_{ij} + M_{ji}} - \frac{\exp(\theta_i - \theta_j)}{1 + \exp(\theta_i - \theta_j)} \right).$$

Thus the vanilla MLE solving $\nabla \ell(\theta) = 0$ is given by

$$\hat{\theta}_{i+1} - \hat{\theta}_i = \log M_{i+1,i} - \log M_{i,i+1} := \log R_{i+1,i},$$

where $M_{ij} := \#\{i \text{ beats } j\}$ and $R_{ij} := \frac{M_{ij}}{M_{ji}}$.

Let $\hat{\theta}_1 = 0$ and we can get

$$\hat{\theta}_1 = 0,\ \hat{\theta}_{i+1} = \sum_{j=1}^{i} (\log R_{j+1,j} - \log r_{j+1,j}),\ i = 1, \cdots, n-1,$$

where $\log R_{j+1,j} := \log M_{j+1,j} - \log M_{j,j+1}$, $\log r_{j+1,j} = \theta_{j+1}^* - \theta_j^*$. Shifting θ^* to make $\theta_1^* = 0$ and we have

$$|\hat{\theta}_{i+1} - \theta_{i+1}^*| = \left| \sum_{j=1}^{i} (\log R_{j+1,j} - \log r_{j+1,j}) \right| \tag{25}$$

Let $F_{ij} := \frac{M_{ij}}{L_{ij}}$ with $L_{ij} = M_{ij} + M_{ji}$ and we have

$$\log R_{ij} - \log r_{ij} = \log \frac{1 - F_{ji}}{F_{ji}} - \log \frac{1 - p_{ji}}{p_{ji}}.$$

Now using the Talyor expansion of $f(x) = \log(\frac{1}{x} - 1)$ at p_{ji}, we can get

$$\log R_{ij} - \log r_{ij} = -v_{ij}(F_{ij} - p_{ij}) + \frac{1}{2}\frac{1 - 2z_{ji}}{z_{ji}^2(1 - z_{ji})^2}(F_{ij} - p_{ij})^2,$$

where $v_{ij} := [p_{ij}(1-p_{ij})]^{-1} = 2 + r_{ij} + r_{ji} = v_{ji} \le 4\exp(|\theta_i^* - \theta_j^*|)$, and z_{ji} is a number between p_{ji} and F_{ji}. We know that $v_{ij}(F_{ij} - p_{ij})$ is a subgaussian-$\frac{v_{ij}^2}{4L_{ij}}$ variable. In particular, for the path graph, by a standard sub-Gaussian tail bound on the first term, we can show that with probability at least $1 - n^{-9}$, it holds for $i = 1, \cdots, n-1$ simultaneously and some constant $C > 0$ that

$$|\sum_{j=1}^{i} v_{j,j+1}(F_{j,j+1} - p_{j,j+1})| \le C\sqrt{\sum_j \frac{v_{j,j+1}^2 \log n}{L_{j,j+1}}}.$$

Using Chernoff's method we can show a slightly sharper bound that if $\min_j L_{ij} > 25e^{2\kappa_E}\log n$ then with probability at least $1 - n^{-9}$,

$$\max_j |F_{ij} - p_{ij}| \le C\sqrt{\frac{\log n}{L_{ij}v_{ij}}}.$$

Since $e^{\kappa_E} = \max_{(i,j)\in E} \frac{p_{ji}}{p_{ij}} \ge \frac{1}{\frac{1}{p_{ij}p_{ji}}p_{ij}^2} = \frac{1}{v_{ij}p_{ij}^2}$, the last inequality and the condition on L_{ij} imply that $|F_{ij} - p_{ij}| \le \min\{p_{ij}, p_{ji}\}/5$ simultaneously, and consequently $z_{ji} \in [0.8p_{ji}, 1.2p_{ji}]$. Therefore, the second term can be bounded as

$$\frac{1}{2}\frac{1 - 2z_{ji}}{z_{ji}^2(1 - z_{ji})^2}(F_{ij} - p_{ij})^2 \le c\frac{1}{p_{ij}^2 p_{ji}^2}(F_{ij} - p_{ij})^2 \le c\frac{v_{ij}\log n}{L_{ij}}.$$

In particular, for the path graph, the summation of the second term is controlled by $\sum_j \frac{v_{j,j+1}\log n}{L_{j,j+1}}$. Thus, when

$$\min_j L_{j,j+1} > ce^{\kappa_E} n\log n,$$

for some sufficiently large constant c, the summation of the second term is negligible compared to the bound on the summation of the first term $\sqrt{\sum_j \frac{v_{j,j+1}^2 \log n}{L_{j,j+1}}}$. Therefore, the error (25) can be bounded as

$$|\hat{\theta}_{i+1} - \theta_{i+1}^*| \le C\sqrt{\sum_{j=1}^{i} \frac{\exp(2|\theta_j^* - \theta_{j+1}^*|)\log n}{L_{j,j+1}}},$$

which implies that $\|\hat{\theta} - \theta^*\|_\infty \le C\sqrt{\sum_{i=1}^{n-1} \frac{\exp(2|\theta_i^* - \theta_{i+1}^*|)\log n}{L_{i,i+1}}}$.

In the special case $L_{i,i+1} = L$ for all i, we have $\|\hat{\theta} - \theta^*\|_\infty \le C\sqrt{\frac{\exp(2\kappa_E)n\log n}{L}}$ and $\|\hat{\theta} - \theta^*\|_2 \le n\sqrt{\frac{\exp(2\kappa_E)\log n}{L}}$. □

In Shah (2016) they prove the lower bound $e^{-\kappa}\frac{n}{\sqrt{L}}$ for ℓ_2 error under path graph, while in our paper we prove the lower bound $e^{-\kappa}\sqrt{\frac{n}{L}}$ for ℓ_∞ error under path graph. Thus the upper bound above on the closed-form MLE achieves the minimax lower bound up to a $\sqrt{\log n}$ and $\exp(2\kappa_E)$ factor.

Proposition 4 (General tree graph)**.** Suppose the graph is a tree graph $([n], E)$ where item i and j are compared L_{ij} times such that $\min_{i,j} L_{ij} > ce^{2\kappa_E}n\log n$ for some universal constant c. Then with probability at least $1 - n^{-4}$, the vanilla MLE $\hat{\boldsymbol{\theta}}_0$ satisfies

$$\|\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}^*\|_\infty \lesssim \sqrt{\max_{i_1,i_2} \sum_{(i,j)\in \mathrm{path}(i_1,i_2)} \frac{\exp(2|\theta_i^* - \theta_j^*|)\log n}{L_{ij}}}. \tag{26}$$

In particular, when $L_{i,j} = L$, we have

$$\|\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}^*\|_\infty \lesssim e^{\kappa_E}\sqrt{\frac{D\log n}{L}}, \|\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}^*\|_2 \lesssim e^{\kappa_E}\sqrt{\frac{Dn\log n}{L}} \tag{27}$$

Proof of Proposition 4. Suppose $\mathcal{G} = ([n], E)$ is a tree graph and let D be the diameter of the graph, i.e., $D = \max_{i,j\in[n]} |\mathrm{path}(i,j)|$ where $\mathrm{path}(i,j) := \{(i,i_1),(i_2,i_3),\cdots,(i_m,j)\} \subseteq E$ is the shortest path from i to j. For instance, for complete graph and star graph, $D = 1$; for path graph, $D = n - 1$; for a complete binary tree with depth or height h, $D = 2h$.

The key property of a tree graph is that it has three equivalent definitions: 1. it is a maximal loop-free graph; 2. it is a minimal connected graph; 3. it is a simple graph with $|E| = |V| - 1$. By the first definition, it can be seen that for any two nodes i and j, there is exactly one path from i to j, otherwise there will be a loop. Using this property, for a fixed item i_0, the MLE equation

$$\nabla \ell \left(\theta\right)_i = \sum_{j \in \mathcal{N}(i)} \left(\frac{M_{ij}}{M_{ij} + M_{ji}} - \frac{\exp(\theta_i - \theta_j)}{1 + \exp(\theta_i - \theta_j)} \right) = 0, \; \forall i \in [n]$$

can be solved by

$$\hat{\theta}_{i_0} = 0, \quad \hat{\theta}_l = \sum_{(i,j) \in \mathrm{path}(i_0, l)} (\log R_{ji} - \log r_{ji}), \; l \neq i_0.$$

Note that, the choice of i_0 is not crucial for the ℓ_∞ bound, as such shifting would at most lead to an inflation on the ℓ_∞ error by 2. In fact, if we let $\hat{\boldsymbol{\epsilon}}$ with $\hat{\epsilon}_{i_0} = 0$ be the entry-wise error when we choose $\hat{\theta}_{i_0} = \theta_{i_0} = 0$, then after shifting $\hat{\boldsymbol{\theta}}$ to get $\tilde{\boldsymbol{\theta}}$ with $\tilde{\theta}_{i_1} = \theta_{i_1} = 0$ for some $i_1 \neq i_0$ we will get $\|\tilde{\boldsymbol{\epsilon}}\|_\infty \leq \|\hat{\boldsymbol{\epsilon}}\|_\infty + \|\hat{\epsilon}_{i_1} \mathbf{1}_n\|_\infty \leq 2\|\hat{\boldsymbol{\epsilon}}\|_\infty$.

Following the same argument in the proof of Proposition 3, we can show that when $\min_{(i,j)\in E} L_{i,j} > e^{2\kappa_E} n \log n$, we have

$$\|\hat{\theta} - \theta^*\|_\infty \lesssim \sqrt{\max_{i_1, i_2} \sum_{(i,j) \in \mathrm{path}(i_1, i_2)} \frac{\exp(2|\theta_i^* - \theta_j^*|) \log n}{L_{ij}}}.$$

Again, the overall ℓ_∞ error is determined by the extreme values of $|\theta_i^* - \theta_j^*|$ and $1/L_{ij}$.

To simplify the bound, we can consider the uniform sampling scheme under which $L_{ij} = L$ for all $(i,j) \in E$. By the definition $D = \max_{i,j \in [n]} |\mathrm{path}(i,j)|$, we have

$$\|\hat{\theta} - \theta^*\|_\infty \lesssim \sqrt{\frac{\exp(2\kappa_E) D \log n}{L}}, \tag{28}$$

and as a corollary,

$$\|\hat{\theta} - \theta^*\|_2 \lesssim \sqrt{\frac{\exp(2\kappa_E) D n \log n}{L}}. \tag{29}$$

In particular, for a path graph, $D = n - 1$ and the two bounds become the same bounds in Proposition 3.

For a star graph, $D = 1$ and the bounds become

$$\|\hat{\theta} - \theta^*\|_\infty \lesssim \sqrt{\frac{\exp(2\kappa_E) \log n}{L}}, \quad \|\hat{\theta} - \theta^*\|_2 \lesssim \sqrt{\frac{\exp(2\kappa_E) n \log n}{L}},$$

which are both sharp up to a $\log n$ factor according to the lower bounds in Shah et al. (2016) and our work. In addition, for the star graph, one can actually show with much simpler arguments that the condition on L can be relaxed to $L > ce^{2\kappa_E} \log n$, because in a star graph essentially we are just comparing $n - 1$ pairs separately. □

A.3 PROOF IN SECTION 3

Let $F(t) = \frac{1}{1+e^{-t}}$, then the Bradley-Terry model can be written as $p_{ij} = F[(w_i^* - w_j^*)/\sigma]$ with $\sigma = 1$. We have $\max_{t\in[0,2\kappa/\sigma]} F'(t) = F'(0) = 1/4$, so ζ in Lemma 13 satisfies

$$\zeta = ce^{\frac{2\kappa}{\sigma}}.$$

Moreover, we denote $\mathcal{W} := \{\boldsymbol{\theta}^* \mid \|\boldsymbol{\theta}^*\|_\infty \leq B\}$. Since $\mathbf{1}_n^\top \boldsymbol{\theta}^* = 0$, we have $B \asymp \kappa := \max_{i,j} |\theta_i^* - \theta_j^*|$ (in general, there is no more information, but for some special cases, e.g., when entries of $\boldsymbol{\theta}^*$ are equal-spaced, we have $\kappa = 2B$). In what follows in this section, we still use quantity σ for generality, but keep in mind that in our setting $\sigma = 1$. We still use B to differ it from κ.

A.3.1 Background and Required Results

Before writing our proof of Theorem 5 we note down specific results from (Shah et al., 2016) and other sources that we will use repeatedly.

Definition 9 (Pairwise $(\delta, \beta)-$packing set from (Shah et al., 2016))**.** Let $\theta \in \mathbb{R}^n$ be a parameter to be estimated, as indexed over a class of probability distributions $\mathcal{P} := \{\mathbb{P}_\theta \mid \theta \in \mathcal{W}\}$, and let $\rho : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ be a pseudo-metric. Suppose there exist a finite set of M vectors $\left\{\theta^1, \ldots, \theta^M\right\}$ such that the following conditions hold:

$$\min_{j,k\in[M],j\neq k} \rho\left(\theta^j, \theta^k\right) \geq \delta \text{ and } \frac{1}{\binom{M}{2}} \sum_{j,k\in[M],j\neq k} D_{\mathrm{KL}}\left(\mathbb{P}_{\theta^j} \| \mathbb{P}_{\theta^k}\right) \leq \beta$$

Then we refer to $\left\{\theta^1, \ldots, \theta^M\right\}$ as $(\delta, \beta)-$packing set.

Lemma 10 (Fano minimax lower bound)**.** *Suppose that we can construct a $(\delta, \beta)-$packing set with cardinality M, then the minimax risk is lower bounded as:*

$$\inf_{\widehat{\theta}} \sup_{\theta^*\in\mathcal{W}} \mathbb{E}\left[\rho\left(\widehat{\theta}, \theta^*\right)\right] \geq \frac{\delta}{2}\left(1 - \frac{\beta + \log 2}{\log M}\right)$$

Proof. See (Yu, 1997, Lemma 3) for details. □

Lemma 11 (Equivalence of $\|\cdot\|_\infty$ and $\|\cdot\|_2$ norms)**.** *Given any vector $\theta \in \mathbb{R}^n$, with $n \in \mathbb{N}$ fixed, the following inequalities holds:*

$$\frac{1}{\sqrt{n}} \|\theta\|_2 \leq \|\theta\|_\infty \leq \|\theta\|_2$$

Proof. The result is standard e.g. see (Wendland, 2018, Proposition 2.10) for a more general version and proof. For the sake of completeness we provide a direct proof of the equivalent statement $\|\theta\|_\infty \leq \|\theta\|_2 \leq \sqrt{n}\|\theta\|_\infty$ as follows:

$$\|\theta\|_\infty = \max_{i\in[n]} |\theta_i| = \sqrt{\max_{i\in[n]} |\theta_i|^2} \leq \sqrt{\sum_i^n {\theta_i}^2} \leq \|\theta\|_2$$

$$\|\theta\|_2 = \sqrt{\sum_i^n {\theta_i}^2} \leq \sqrt{\sum_i^n \max_{i\in[n]} |\theta_i|^2} = \sqrt{n \max_{i\in[n]} |\theta_i|^2} = \sqrt{n}\|\theta\|_\infty$$

This proves the lower and upper inequalities respectively, as required. □

Remark 1. We note that the above inequalities are tight i.e. have optimal constants. In the case of the upper bound consider $\theta = \mathbf{1}_n = (1, \ldots, 1)^\top \in \mathbb{R}^n$. So $\|\theta\|_\infty = 1$ and $\|\theta\|_2 = \sqrt{n} = \sqrt{n}\|\theta\|_\infty$ showing the tightness of the upper bound. In the case of the lower bound, consider $\theta = \mathbf{e}_1$ i.e. WLOG the first standard basis vector in $\mathbb{R}^n$. We then have that $\|\theta\|_\infty = \|\theta\|_2 = 1$, which shows tightness in the lower bound.

Lemma 12 (Lemma 7 in (Shah et al., 2016)). *For any $\alpha \in \left(0, \frac{1}{4}\right)$, there exists a set of $M(\alpha)$ binary vectors $\left\{z^1, \ldots, z^{M(\alpha)}\right\} \subset \{0,1\}^n$ such that*

$$\alpha n \leq \left\|z^j - z^k\right\|_2^2 \leq n \quad \textit{for all } j \neq k \in [M(\alpha)], \textit{ and} \tag{30}$$

$$\langle e_1, z^j \rangle = 0 \quad \textit{for all } j \in [M(\alpha)] \tag{31}$$

Lemma 13 (Lemma 8 in (Shah et al., 2016)). *For any pair of quality score vectors θ^j and θ^k, and for*

$$\zeta := \frac{\max_{x \in [0, 2B/\sigma]} F'(x)}{F(2B/\sigma)(1 - F(2B/\sigma))}$$

we have

$$D_{\mathrm{KL}}\left(\mathbb{P}_{\theta^j} \| \mathbb{P}_{\theta^k}\right) \leq \frac{N_{comp}\zeta}{\sigma^2}\left(\theta^j - \theta^k\right)^\top L\left(\theta^j - \theta^k\right) =: \frac{N_{comp}\zeta}{\sigma^2}\left\|\theta^j - \theta^k\right\|_L^2$$

Lemma 14 (Lemma 14 in (Shah et al., 2016)). *The Laplacian matrix $\tilde{\mathcal{L}}_{\mathbf{A}}$ satisfies the trace constraints:*

$$\operatorname{tr}(\tilde{\mathcal{L}}_{\mathbf{A}}) = 2 \tag{32}$$

$$\operatorname{tr}(\tilde{\mathcal{L}}_{\mathbf{A}}^{\dagger}) \geq \frac{n^2}{4} \tag{33}$$

Proof. See (Shah et al., 2016, Lemma 14) for details. □

The challenge - constructing a suitable pairwise packing set that meets Definition 9. The main tool to use here is the Varshamov-Gilbert Lemma

A.3.2 Sketch of Lower Bound Proof

In brief, we seek a minimax lower bound proof of the same form as (Shah et al., 2016, Theorem 2(a)), except in our case we choose our packing set norm to the ℓ_∞ norm, rather than the ℓ_2^2 in (Shah et al., 2016, Theorem 2(a)). We leverage their construction directly in two main ways. First, we use the slightly modified version of Fano's Lemma that enables the (δ, β)-packing set to be constructed for the ℓ_∞ norm, not the ℓ_∞^2 norm, consistent with our high probability upper bound per Lemma 10. Second, we can switch out their use of the (δ, β)-packing set in the ℓ_2 norm to the (δ', β')-packing set in the ℓ_∞ norm. This is done by using the topological equivalence of norms in finite dimensions per Lemma 11, which is shown to be tight in the dimension per Remark 1. For the sake of clarity, we use much of the same wording as the proof from (Shah et al., 2016, Appendix B), for the convenience of the reader.

A.3.3 Lower Bound Proof - Part I

Our proof follows directly the approach taken from (Shah et al., 2016, Section B.1). The normalized Laplacian $\tilde{\mathcal{L}}_{\mathbf{A}}$ of the comparison graph is symmetric and positive-semidefinite. We can thus decompose this via diagonalization as $L = U^\top \Lambda U$ where $U \in \mathbb{R}^{n \times n}$ is an orthonormal matrix, and Λ is a diagonal matrix of nonnegative eigenvalues $\Lambda_{jj} = \lambda_j(L)$ for each $j \in [n]$. Similar to (Shah et al., 2016, Section B.1) we first prove that the minimax risk is lower bounded by $c\sigma^2 \frac{n^2}{N_{\text{comp}}}$.

Fix scalars $\alpha \in (0, \frac{1}{4})$ and $\delta > 0$, with values to be specified later. Obtain set of vectors on the Boolean Hypercube $\{0,1\}^n$ i.e. $\left\{z^1, \ldots, z^{M(\alpha)}\right\}$ given by Lemma 12, where $M(\alpha)$ is set to be

$$M(\alpha) := \left\lfloor \exp\left\{\frac{n}{2}(\log 2 + 2\alpha \log 2\alpha + (1 - 2\alpha)\log(1 - 2\alpha))\right\}\right\rfloor. \tag{34}$$

Define another set of vectors of the same cardinality $\left\{\theta^j \mid j \in [M(\alpha)]\right\}$ via $\theta^j := \frac{\delta}{\sqrt{n}} U^\top P z^j$, where P is a permutation matrix. The permutation matrix P has the constraint that it keeps the first coordinate constant i.e. $P_{11} = 1$. By construction for each $j \neq k$ we have that

$$\left\|\theta^j - \theta^k\right\|_\infty \overset{(i)}{\geq} \frac{1}{\sqrt{n}} \left\|\theta^j - \theta^k\right\|_2 \overset{(ii)}{=} \frac{1}{\sqrt{n}} \left(\frac{\delta}{\sqrt{n}} \left\|z^j - z^k\right\|_2 \right) \overset{(iii)}{\geq} \delta \sqrt{\frac{\alpha}{n}} \tag{35}$$

Here (i) follows from Lemma 11. Additionally (ii) follows since $\theta^j := \frac{\delta}{\sqrt{n}} U^\top P z^j$. In the case of the final inequality (iii), we have $\frac{\delta^2}{n^2} \left\|z^j - z^k\right\|_2^2 \geq \frac{\alpha n \delta^2}{n^2} = \frac{\alpha \delta^2}{n}$. Since the set $\left\{z^1, \ldots, z^{M(\alpha)}\right\}$ are binary vectors with a minimum Hamming distance at least αn using Equation (30). Consider, any distinct $j, k \in [M(\alpha)]$, then for some subset $\{i_1, \ldots, i_r\} \subseteq \{2, \ldots, n\}$ with $\alpha n \leq r \leq n$ it must follow that

$$\left\|\theta^j - \theta^k\right\|_{\tilde{\mathcal{L}}_{\mathbf{A}}}^2 = \frac{\delta^2}{n} \left\|U^\top P z^j - U^\top P z^k\right\|_{\tilde{\mathcal{L}}_{\mathbf{A}}}^2 = \frac{\delta^2}{n} \left\|z^j - z^k\right\|_\Lambda^2 = \frac{\delta^2}{n} \sum_{m=1}^{r} \lambda_{i_m}(\tilde{\mathcal{L}}_{\mathbf{A}})$$

The last part follows since Λ is a diagonal matrix of non-negative eigenvalues with $\Lambda_{ii} = \lambda_j(L)$. Now for given $\{a_2, \ldots, a_n\}$ such that $\alpha n \leq \sum_{i=2}^n a_i \leq n$ we have that

$$\frac{1}{\binom{M(\alpha)}{2}} \sum_{j \neq k} \left\|\theta^j - \theta^k\right\|_{\tilde{\mathcal{L}}_{\mathbf{A}}}^2 = \frac{\delta^2}{n} \sum_{i=2}^{n} a_i \lambda_i(\tilde{\mathcal{L}}_{\mathbf{A}})$$

The permutation matrix P is chosen such that the last $n-1$ coordinates are permuted to have $a_1 \geq \ldots \geq a_n$ and keep the n^{th} coordinate fixed. By this particular choice, and using the fact that $\operatorname{tr}(\tilde{\mathcal{L}}_{\mathbf{A}}) = 2$ we have that:

$$\frac{1}{\binom{M(\alpha)}{2}} \sum_{j \neq k} \left\|\theta^j - \theta^k\right\|_{\tilde{\mathcal{L}}_{\mathbf{A}}}^2 = \frac{\delta^2}{n} \frac{n}{n-1} \operatorname{tr}(\tilde{\mathcal{L}}_{\mathbf{A}}) \leq \frac{2\delta^2}{n} \operatorname{tr}(\tilde{\mathcal{L}}_{\mathbf{A}}) = \frac{4\delta^2}{n}$$

Now by the choice of P above, we have that for every choice of $j \in [M(\alpha)]$

$$\langle \tilde{\mathcal{L}}_{\mathbf{A}}, \theta^j \rangle = \frac{\delta}{\sqrt{n}} \mathbf{e}_1^\top P z^j = \mathbf{e}_1^\top z^j = 0$$

where the last equality follows from Equation (31). Now the basic condition needs to be verified i.e. did the θ^j we chose satisfy the boundedness constraint, to ensure that $\theta^j \in \mathcal{W}_B$? Setting $\delta^2 = 0.01 \frac{\sigma^2 n^2}{4 N_{\text{comp}} \zeta}$, it indeed follows that $\left\|\theta^j\right\|_\infty \leq \frac{\delta}{\sqrt{n}} \left\|z^j\right\|_2 \overset{(i)}{\leq} \delta \overset{(ii)}{\leq} B$. Here (i) follows since $z^j \in \{0,1\}^n$. Furthermore (ii) follows from our choice of δ and our assumption that $N_{\text{comp}} \geq \frac{c\sigma^2 \operatorname{tr}\left(\tilde{\mathcal{L}}_{\mathbf{A}}^\dagger\right)}{\zeta B^2}$ with $c = 0.002$, where Lemma 14 guarantees that $N_{\text{comp}} \geq \frac{c\sigma^2 n^2}{4\zeta B^2}$. We have thus verified that each vector θ_j also satisfies the boundedness constraint $\left\|\theta^j\right\|_\infty \leq B$, which is required for membership in $\mathcal{W}_B$. Finally by Lemma 13 we have that:

$$D_{\text{KL}}\left(\mathbf{P}_{\theta^j} \| \mathbf{P}_{\theta^k}\right) \leq \frac{N_{\text{comp}} \zeta}{\sigma^2} \frac{4\delta^2}{n} = 0.01n \tag{36}$$

To summarize, we have now constructed a (δ', β')-packing set with respect to the norm $\rho\left(\theta^j, \theta^k\right) := \left\|\theta^j - \theta^k\right\|_\infty$, where $\delta' = \delta\sqrt{\frac{\alpha}{n}}$ from Equation (35), and $\beta' = 0.01d$ from Equation (36).

Finally we have by substituting (δ', β') into the pairwise Fano's lower bound (Lemma 10) that:

$$\sup_{\theta^* \in \mathcal{W}_B} \mathbb{E}\left[\left\|\widetilde{\theta} - \theta^*\right\|_\infty\right] \geq \delta \sqrt{\frac{\alpha}{n}} \left(1 - \frac{0.01n + \log 2}{\log M(\alpha)}\right) = c\sigma \sqrt{\frac{n}{\zeta N_{\text{comp}}}} \left(1 - \frac{0.01n + \log 2}{\log M(\alpha)}\right)$$

which yields the claim, after appropriate substitution of δ and setting $\alpha = 0.01$.

For the case of $n \leq 9$, consider the set of the three n-length vectors $z^1 = (0, \ldots, -1)$, $z^2 = (0, \ldots, 0)$ and $z^3 = (0, \ldots, 0)$. Construct the packing set $\left(\theta^1, \theta^2, \theta^3\right)$ from these three vectors $\left(z^1, z^2, z^3\right)$ as done above for the case of $n > 9$. From the calculations made for the general case above, we have for all pairs $\min_{j \neq k} \left\|\theta^j - \theta^k\right\|_\infty^2 \geq \frac{1}{9} \min_{j \neq k} \left\|\theta^j - \theta^k\right\|_2^2 \geq \frac{\delta^2}{81}$ and $\max_{j,k} \left\|\theta^j - \theta^k\right\|_{\tilde{\mathcal{L}}_{\mathbf{A}}}^2 \leq 4\delta^2$, and as a result $\max_{j,k} D_{\text{KL}}\left(\mathbf{P}_{\theta^j} \| \mathbf{P}_{\theta^k}\right) \leq \frac{4 N_{\text{comp}} \zeta \delta^2}{\sigma^2}$. Choosing $\delta^2 = \frac{\sigma^2 \log 2}{8 N_{\text{comp}} \zeta}$ and applying the pairwise Fano's lower bound (Lemma 10) yields the claim.

For the other case, the lower bound in terms of $\lambda_i(\tilde{\mathcal{L}}_{\mathbf{A}})$, the argument is similar, and we end up with an extra factor of $\frac{1}{n'}$.

A.3.4 Lower Bound Proof - Part II

Given an integer $n' \in \{2, \ldots, n\}$, and constants $\alpha \in (0, \frac{1}{4})$, $\delta > 0$, define the integer:

$$M'(\alpha) := \left\lfloor \exp\left\{ \frac{n'}{2} (\log 2 + 2\alpha \log 2\alpha + (1 - 2\alpha) \log(1 - 2\alpha)) \right\} \right\rfloor \tag{37}$$

Applying Lemma 12 using n' as the dimension results in a subset $\left\{ z^1, \ldots, z^{M'(\alpha)} \right\}$ of the Boolean hypercube $\{0,1\}^{n'}$, with specified properties. We then define a finite set of size $M'(\alpha)$, of n-length vectors $\left\{ \widetilde{\theta}^1, \ldots, \widetilde{\theta}^{M'(\alpha)} \right\}$ using:

$$\widetilde{\theta}^j = \left[0 \left(z^j\right)^\top 0 \cdots 0 \right]^\top \qquad \text{for each } j \in [M(\alpha)]$$

For each $j \in [M(\alpha)]$, let us define $\theta^j := \frac{\delta}{\sqrt{n'}} U^\top \sqrt{\Lambda^\dagger}\, \widetilde{\theta}^j$. For the first standard basis vector $\mathbf{e}_1 \in \mathbb{R}^n$, we then have that $\langle \mathbf{1}_n, \theta^j \rangle = \frac{\delta}{\sqrt{n'}} \tilde{\mathcal{L}}_{\mathbf{A}}^\top U^\top \sqrt{\Lambda^\dagger}\, \widetilde{\theta}^j = 0$. Here the main fact used is $\tilde{\mathcal{L}}_{\mathbf{A}} \mathbf{1}_n = 0$. Additionally we have that for any $j \neq k$, we have that:

$$\left\| \theta^j - \theta^k \right\|_\infty^2 \geq \frac{1}{n} \left\| \theta^j - \theta^k \right\|_2^2 = \frac{1}{n} \frac{\delta^2}{n'} \left(\widetilde{\theta}^j - \widetilde{\theta}^k \right)^\top \Lambda^\dagger \left(\widetilde{\theta}^j - \widetilde{\theta}^k \right) \geq \frac{1}{n} \frac{\delta^2}{n'} \sum_{i=\lceil (1-\alpha) n' \rceil}^{n'} \frac{1}{\lambda_i} \tag{38}$$

Now, setting $\delta^2 = 0.01 \frac{\sigma^2 n'}{N_{\text{comp}} \zeta}$ results in:

$$\left\| \theta^j \right\|_\infty \leq \frac{\delta}{\sqrt{n'}} \left\| \sqrt{\Lambda^\dagger}\, \widetilde{\theta}^j \right\|_2 \overset{(i)}{\leq} \frac{\delta}{\sqrt{n'}} \sqrt{\operatorname{tr}\left(\Lambda^\dagger\right)} \overset{(ii)}{=} \frac{\delta}{\sqrt{n'}} \sqrt{\operatorname{tr}\left(L^\dagger\right)} \overset{(iii)}{\leq} B \tag{39}$$

where inequality (i) follows from the fact that z^j has entries in $\{0,1\}$; step (ii) follows because the matrices $\sqrt{\Lambda^\dagger}$ and $\sqrt{\tilde{\mathcal{L}}_{\mathbf{A}}^\dagger}$ have the same eigenvalues; and inequality (iii) follows from our choice of δ and our assumption $N_{\text{comp}} \geq \frac{c\sigma^2 \operatorname{tr}\left(\tilde{\mathcal{L}}_{\mathbf{A}}^\dagger\right)}{\zeta B^2}$ on the sample size with $c = 0.01$. We have thus verified that each vector θ^j also satisfies the boundedness constraint $\left\| \theta^j \right\|_\infty \leq B$, as required for membership in $\mathcal{W}_B$. Furthermore, for any pair of distinct vectors in this set, we have:

$$\left\| \theta^j - \theta^k \right\|_{\tilde{\mathcal{L}}_{\mathbf{A}}}^2 = \frac{\delta^2}{n'} \left\| z^j - z^k \right\|_2^2 \leq \delta^2$$

By Lemma 13 we have that:

$$D_{\text{KL}}\left(\mathbf{P}_{\theta^j} \| \mathbf{P}_{\theta^k} \right) \leq \frac{N_{\text{comp}} \zeta}{\sigma^2} \left\| \theta^j - \theta^k \right\|_{\tilde{\mathcal{L}}_{\mathbf{A}}}^2 = 0.01 n' \tag{40}$$

Finally we have by substituting (δ', β') into the pairwise Fano's lower bound (Lemma 10) that:

$$\sup_{\theta^* \in \mathcal{W}_B} \mathbb{E}\left[\left\| \widetilde{\theta} - \theta^* \right\|_\infty \right] \geq \frac{\frac{\delta}{\sqrt{n' d}} \sqrt{\sum_{i=\lceil (1-\alpha) n' \rceil}^{n'} \frac{1}{\lambda_i}}}{2} \left(1 - \frac{0.01 n' + \log 2}{\log M'(\alpha)} \right)$$

Substituting our choice of δ and setting $\alpha = 0.01$ proves the claim for $n' > 9$.

For the case of $n' \leq 9$. Consider the packing set of the three n'-length vectors $\theta^1 = \delta U \sqrt{\Lambda^\dagger}\, (0, 1, \ldots, 0)$, $\theta^2 = -\theta^1$ and $\theta^3 = (0, \ldots, 0)$. Then we have for all pairs $\min_{j \neq k} \left\| \theta^j - \theta^k \right\|_\infty^2 \geq \frac{1}{9} \min_{j \neq k} \left\| \theta^j - \theta^k \right\|_2^2 \geq \frac{\delta^2}{9 \lambda_2(L)}$ and $\max_{j,k} \left\| \theta^j - \theta^k \right\|_{\tilde{\mathcal{L}}_{\mathbf{A}}}^2 \leq 4\delta^2$, and as a result $\max_{j,k} D_{\text{KL}}\left(\mathbf{P}_{\theta^j} \| \mathbf{P}_{\theta^k} \right) \leq \frac{4 N_{\text{comp}} \zeta \delta^2}{\sigma^2}$. Choosing $\delta^2 = \frac{\sigma^2 \log 2}{8 N_{\text{comp}} \zeta}$ and applying the pairwise Fano's lower bound (Lemma 10) yields the claim.

A.4 ADDITIONAL EXPERIMENTS

In this section, we show some additional results of experiments. The section has three parts:

1. Experiments related to Section 4.
2. Extra comparisons as a supplement to Section 5.
3. Experiments to illustrate that in some cases κ_E is still loose and point out a potential future direction.

A.4.1 Subadditivity

In this section, we show some simulation results illustrating the advantage of using subadditivity property in the estimation of the BTL model, as we discuss in Section 4.

Island graph. In this setting, we consider the Island graph with n nodes described in Example 7 and denote the node set of the k-th island as V_k. Suppose we get the estimator $\hat{\boldsymbol{\theta}}^{(k)} \in \mathbb{R}^{|V_k|}$ $(k > 1)$ based on k-th Island V_k with the augmented version $\tilde{\boldsymbol{\theta}}^{(k)} \in \mathbb{R}^n$ such that the subvector $\tilde{\boldsymbol{\theta}}^{(k)}(V_k) = \hat{\boldsymbol{\theta}}^{(k)}$. We can define the ensemble estimator add-MLE in the following way: first shift $\tilde{\boldsymbol{\theta}}^{(k)}$ and get

$$\check{\boldsymbol{\theta}}^{(k)} = \hat{\boldsymbol{\theta}}^{(k)} + s_k, s_k = s_{k-1} + \tilde{\boldsymbol{\theta}}^{(k-1)}(n_{\text{island}}) - \tilde{\boldsymbol{\theta}}^{(k)}(n_{\text{overlap}}), s_0 = 0.$$

Then construct $\tilde{\boldsymbol{\theta}}^{add} \in \mathbb{R}^n$ such that for all k, $\tilde{\boldsymbol{\theta}}^{add}(V_k) = \check{\boldsymbol{\theta}}^{(k)}$. At last, we centerize $\check{\boldsymbol{\theta}}^{(k)}$ and get

$$\hat{\boldsymbol{\theta}}^{add} = \tilde{\boldsymbol{\theta}}^{add} - \mathbf{1}_n \cdot \frac{1}{n} \mathbf{1}_n^\top \tilde{\boldsymbol{\theta}}^{add}.$$

For the ease of implementation and precise description of the performance, we construct the true parameter $\boldsymbol{\theta}^*$ by first set $\boldsymbol{\theta}^*(i) = \boldsymbol{\theta}^*(1) + (i-1)\delta$ such that $\operatorname{avg}(\boldsymbol{\theta}^*) = 0$, and then shift $\boldsymbol{\theta}^*_{(k)} := \boldsymbol{\theta}^*(V_k)$ by $s_k := -(k-1)s$ and call $s \in \mathbb{R}$ the shifting coefficient. Figure 1 shows that add-MLE outperforms the joint-MLE, where "shift" in the right panel is the shifting coefficient s. Notice that to save space and for the ease of understanding, we transform "shift" to "diff" in the Section 5 so that "diff" shows the difference in the average: $\operatorname{avg}(\boldsymbol{\theta}^*_{(k-1)}) - \operatorname{avg}(\boldsymbol{\theta}^*_{(k)})$. We set $L = 10$ for all pairs.

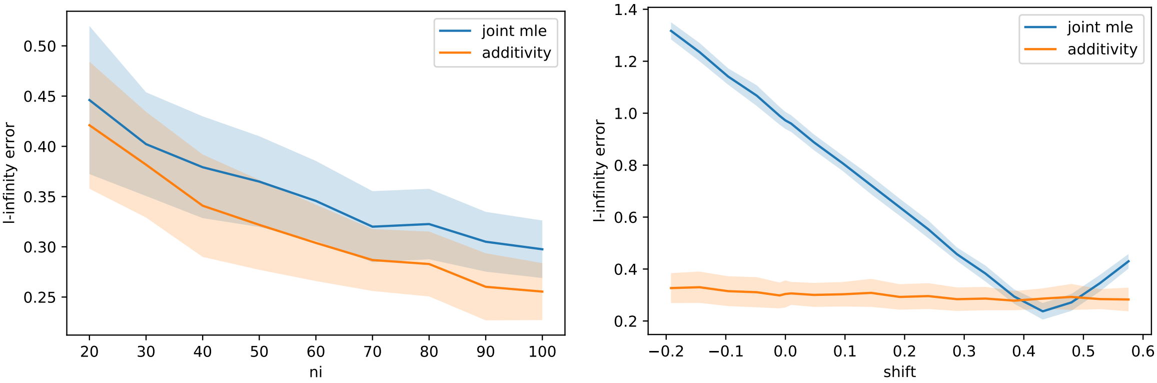

Figure 1: Estimation errors given by joint-MLE and add-MLE on Island graphs. The y-axis is $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_\infty$. In the left panel, the x-axis is the size of islands n_{island} while $n_{overlap} = 5$, $k = 3$, $L = 10$ in all cases. In the right panel, the x-axis is s, the shifting coefficient while $n_{island} = 50$, $n_{overlap} = 5$, $k = 5$, $L = 10$. The lines show the average of 100 trials with the standard deviation shown as the shaded area.

Barbell graph. In this setting, a barbell graph $\mathcal{G}_{barbell}([n], E)$ of two equal-sized complete subgraphs $\mathcal{G}_1, \mathcal{G}_2$ linked by a set of bridge edges $E_{bridge} = \{(i,j) : i \in \mathcal{G}_1, j \in \mathcal{G}_2, (i,j) \in E\}$ are generated as is discussed in Example 8. Note that the vertex sets V_1, V_2 of $\mathcal{G}_1, \mathcal{G}_2$ are disjoint and $V_1 \cup V_2 = [n]$. We set $L = 10$ for all pairs. Again, two methods of estimation are compared:

- Joint-MLE. A single regularized MLE $\hat{\boldsymbol{\theta}}^{joint}$ is fitted on $\mathcal{G}_{barbell}$.

- Add-MLE. Two regularized MLE's $\hat{\boldsymbol{\theta}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$ are fitted separately on $\mathcal{G}_1$ and $\mathcal{G}_2$. Then for each $e = (i, j) \in E_{bridge}$, we calculate

$$\hat{d}_e := \log(\frac{\hat{p}_e}{1 - \hat{p}_e}),$$

where $\hat{p}_e = \text{clip}(\frac{win_{ij}}{win_{ij}+loss_{ij}}, p_{up}, p_{lb})$ for $e = (i, j)$ with two constants $0 < p_{lb} < p_{up} < 1$ for regularity. We take $p_{lb} = 0.1, p_{up} = 0.9$. Then for $e = (i, j) \in E_{bridge}$, the shifting constant $\hat{s}_e$ is defined as

$$\hat{s}_e := \hat{d}_e - (\tilde{\theta}_i^{(1)} - \tilde{\theta}_j^{(2)}),$$

where $\tilde{\boldsymbol{\theta}}^{(i)}$ is the augmented version of $\hat{\boldsymbol{\theta}}^{(1)}$ satisfying $\tilde{\boldsymbol{\theta}}^{(i)}(V_i) = \hat{\boldsymbol{\theta}}^{(1)}$. Then the average $\hat{s}_E$ is calculated via $\hat{s}_E := \frac{1}{|E|}\sum_{e \in E} \hat{s}_e$ and the add-MLE $\hat{\boldsymbol{\theta}}^{add}$ is constructed via

$$\tilde{\theta}_i^{add} := \begin{cases} \tilde{\theta}_i^{(1)}, i \in \mathcal{G}_1, \\ \tilde{\theta}_i^{(2)} - \hat{s}_E, i \in \mathcal{G}_2, \end{cases}$$

and $\hat{\boldsymbol{\theta}}^{add} = \tilde{\boldsymbol{\theta}}^{add} - \mathbf{1}_n \cdot \frac{1}{n}\mathbf{1}_n^\top \tilde{\boldsymbol{\theta}}^{add}$.

Notice that we slightly change the way of constructing the add-MLE in Section 4 to exploit multiple random bridge edges in this setting.

Similar to the Island graph case, we let $\mathcal{G}_1$ be the complete graph of node $\{1, \ldots, \frac{n}{2}\}$ and $\mathcal{G}_2$ the complete graph of nodes $\{\frac{n}{2} + 1, \ldots, n\}$. To set the true parameter, we let

$$\theta_i^* = \begin{cases} \theta_1^* + (i-1)\delta, i \leq \frac{n}{2}, \\ -s + \theta_1^* + (i-1)\delta, i > \frac{n}{2}, \end{cases}$$

where $s \in \mathbb{R}$ is the shifting coefficient.

As we can see from Figure 2, the performance of the add-MLE is more stable than that of the joint-MLE, while ensuring better or similar ℓ_∞ estimation error.

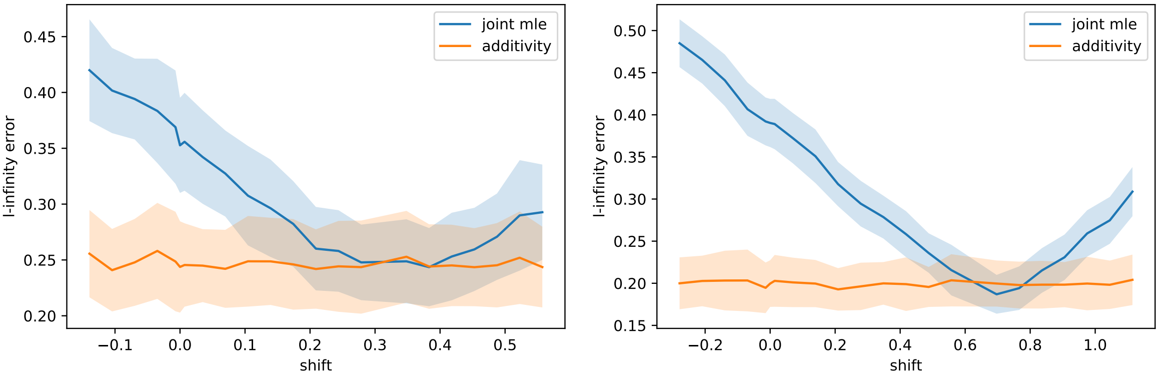

Figure 2: Estimation errors given by joint-MLE and add-MLE on Barbell graphs with 100 nodes(left) and 200 nodes (right). Two equal-sized complete subgraphs are linked by 10 (left) and 20 (right) randomly sampled bridge edges. The y-axis is $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_\infty$ and the x-axis is s, the shifting coefficient. $L = 100$ for bridge edges and $L = 10$ for non-bridge edges. The lines show the average of 100 trials with the standard deviation shown as the shaded area.

A.4.2 Extra comparisons

A one-sentence summary of this section is, we demonstrate our discussions in Section 5 by real cases that κ_E can give much tighter upper bounds than κ.

We consider a k-banded graph where comparisons are made only for pairs with difference in indices smaller or equal to k. That is, the edge set of the comparison graph is $E_k := \{(i,j) : |i-j| \le k\}$. We consider two settings, $k = \sqrt{n}$ and $k = n/\log n$, and in each setting, we set $\kappa = \log(n)$, $L = 10$ and $\theta_i^* = \theta_1^* + (i-1)\delta$ for $i > 1$ with $\delta = \kappa/(n-1)$.

We compare the real ℓ_2-error of the regularized MLE and three upper bounds: the upper bound for the ℓ_2-error provided by our paper using κ and κ_E, and the one provided by Shah et al. (2016). As is shown in Figure 3, κ_E gives tighter upper bounds than κ in the setting of banded comparison graph. It should be noted that all curves are going up because for a banded graph, L needs to be sufficiently large to guarantee $o(1)$ ℓ_∞ error, and we set $L = 10$ simply for illustration of the effectiveness of κ_E versus κ here.

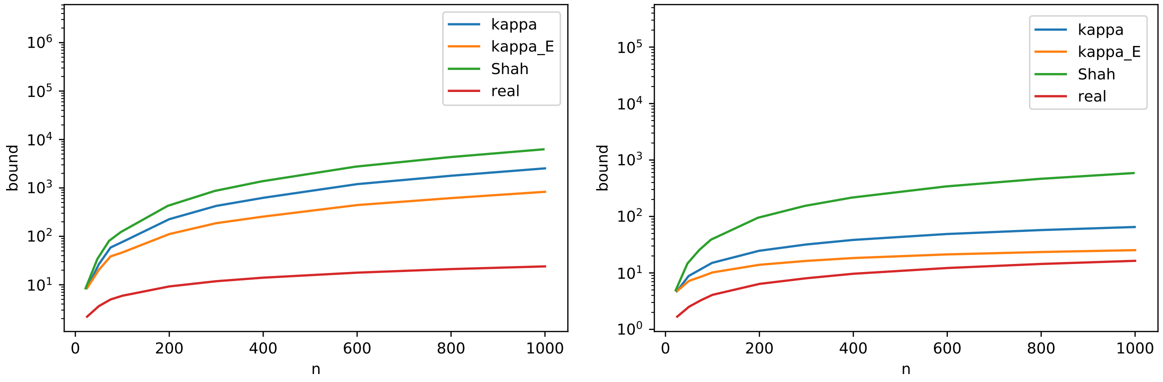

Figure 3: Comparison under the k-banded graph with $k = \sqrt{n}$ (left) and $k = n/\log(n)$ (right). Each point on lines is an average of 20 trials. The two curves of upper bounds kappa and Shah are manually shifted downwards so that they started at the same level with the curve for kappa_E, to remove the affect of the choice of constant (though all leading constants are set to be 1 here) and make it easier to compare the increasing rate of the bound.

A.4.3 Cases where κ_E is loose

Consider a path graph of n nodes and edge set $\{(i, i+1) : i \in [n-1]\}$. Assume $\theta_i^* = \theta_1^* + (i-1)\delta$, then $\kappa_E = \delta$ and $\kappa = (n-1)\delta$. In this case, a factor of e^{κ_E} gives tighter control than e^{κ}. However, if we add one edge $(1, n)$ into the graph, κ_E becomes $(n-1)\delta$ and our upper bound will increase a lot, which is counter-intuitive because the newly-added 1 out of n edges should not affect the estimation accuracy too much. In other words, the bound gets looser after the new edge is added.

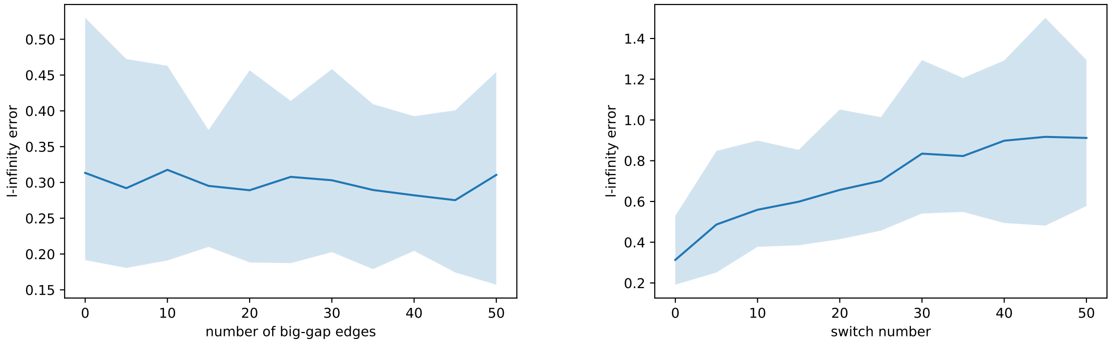

Figure 4: Experiment on the path graph. Left: the ℓ_∞ error of the regularized MLE when adding new edges with big performance gaps to the path graph. Right: the ℓ_∞ error of the regularized MLE when switch some pairs of the performance parameter $\{\theta_i^*\}$ so that there are more big-gap edges while keeping the algebraic connectivity. The results show that when the proportion of big-gap edges is small, the estimation error would not be affected a lot. Both experiments are under $\kappa \approx 6.9$, equal-gap θ^*, $n = 200$ and $L = 5000$ and based on 40 trials for each hyperparameter, with 0.05 and 0.95 quantiles shown by the shaded area.

The reason is that κ_E itself is not enough to provide tight control across all comparison graphs. For instance, to avoid such loose cases, one also needs to take into account the proportion of edges with big performance gaps compared to edges with small performance gaps.

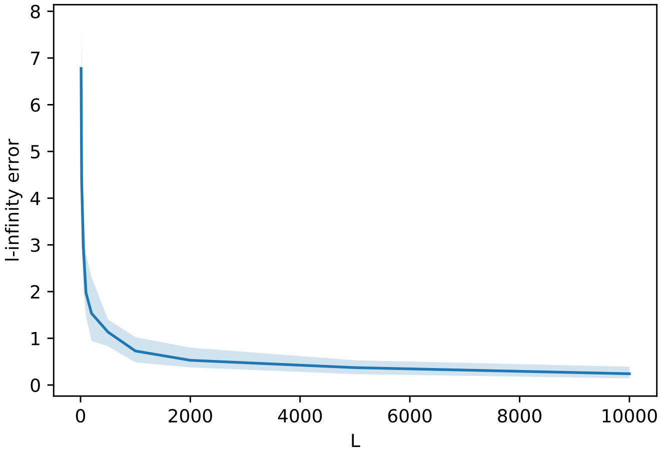

Figure 5: Experiment on the path graph, under $\kappa \approx 6.9$, equal-gap θ^*, $n = 200$ and based on 40 trials for each hyperparameter L, with 0.05 and 0.95 quantiles shown by the shaded area.

Figure 4 shows some numerical results on the impact of big-gap edges. In the left panel, we add edges to the top-right and bottom-left corner of the adjacency matrix so that their performance gaps are big. In the right panel, we switch some pairs of parameters $\{\theta_i^*\}$ so that there are more big-gap edges while keeping the algebraic connectivity. The first switch will switch $\theta_1^*, \theta_{[n/2]+1}^*$. The i-th switch will switch the pair $\theta_{2i-1}^*, \theta_{[n/2]+2i-1}^*$. As an example, taking $n = 8$, $\theta_1^* = 0$, and $\delta = 1$, 2 switches will make parameters change as

$$(1, 2, 3, 4, 5, 6, 7, 8) \longrightarrow (5, 2, 7, 4, 1, 6, 3, 8). \tag{41}$$

In the left panel, as new big-gap edges come in, the algebraic connectivity of the graph gets larger as well, keeping estimation errors in the same level. In the right panel, as we keep the algebraic connectivity constant, the impact of the proportion of big-gap edges is shown more clearly.

Another thing we need to point out is that for cases like path graphs, even if we ignore the factor of κ, bounding estimation error itself is hard due to the poor connectivity of the graph, as is argued in Shah et al. (2016) (their upper bound for both path and cycle graphs is not optimal). As we have shown in Section 2.2, even for d-regular graphs with relatively small d, the algebraic connectivity is not big enough and our upper bound cannot match the lower bound. But as the comparison graph gets denser and more regular, κ_E will get closer to κ, making the difference between e^{κ_E} and e^{κ} not as dramatic, although in finite sample phase the difference can still be big because it is an exponential factor.

The last thing we want is that the estimation itself (without providing a theoretical tight upper bound for the estimation error) under the path graph is hard: one need a huge L to make accurate estimation when n is big, as is shown in Figure 5.

A.5 OTHER SUPPORTING RESULTS

Proposition 6. (Subadditivity) Let I_1, I_2, I_3 be three subsets of $[n]$ such that $\cup_{j=1}^{3} I_j = [n]$ and, for each $j \neq k$, $I_j \not\subseteq I_k$ and for $i = 1, 2$, $I_i \cap I_3 \neq \emptyset$. Assume that the sub-graphs induced by the I_j's are connected. Let $\boldsymbol{\theta}^*$ be the vector of preference scores in the BTL model over n items and $\hat{\boldsymbol{\theta}}_{(j)}$ be the MLE of $\boldsymbol{\theta}^*_{(j)}$ for the BTL model involving only items in I_j, $j = 1, 2, 3$, with augmented versions $\tilde{\boldsymbol{\theta}}_{(j)} \in \mathbb{R}^n$ such that $\tilde{\boldsymbol{\theta}}_{(j)}(I_j) = \hat{\boldsymbol{\theta}}_{(j)}$. Take two nodes $t_1 \in I_1 \cap I_3$, $t_2 \in I_2 \cap I_3$, and let $\delta_3 = \tilde{\boldsymbol{\theta}}_{(1)}(t_1) - \tilde{\boldsymbol{\theta}}_{(3)}(t_1)$, $\delta_2 = \tilde{\boldsymbol{\theta}}_{(3)}(t_2) - \tilde{\boldsymbol{\theta}}_{(2)}(t_2)$. An ensemble MLE $\hat{\boldsymbol{\theta}} \in \mathbb{R}^n$ is a vector such that $\hat{\boldsymbol{\theta}}(I_1) = \hat{\boldsymbol{\theta}}_{(1)}$, $\hat{\boldsymbol{\theta}}(S_2) = \hat{\boldsymbol{\theta}}_{(2)}(S_2) + \delta_3 + \delta_2$, and $\hat{\boldsymbol{\theta}}(S_3) = \tilde{\boldsymbol{\theta}}_{(3)}(S_3) + \delta_3$, where $S_2 = I_2 \setminus I_1$ and $S_3 = I_3 \setminus (I_1 \cup I_2)$. It holds for any ensemble MLE $\hat{\boldsymbol{\theta}}$ that

$$\frac{1}{4} d_\infty(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}^*) \le d_\infty(\hat{\boldsymbol{\theta}}_{(1)}, \boldsymbol{\theta}^*_{(1)}) + d_\infty(\hat{\boldsymbol{\theta}}_{(2)}, \boldsymbol{\theta}^*_{(2)}) + d_\infty(\hat{\boldsymbol{\theta}}_{(3)}, \boldsymbol{\theta}^*_{(3)}), \tag{42}$$

where $d_\infty(\mathbf{v}_1, \mathbf{v}_2) := \|(\mathbf{v}_1 - \mathbf{1}^\top \mathrm{avg}(\mathbf{v}_1)) - (\mathbf{v}_2 - \mathbf{1}^\top \mathrm{avg}(\mathbf{v}_2))\|_\infty$, where $\mathrm{avg}(\mathbf{x}) := \frac{1}{n}\mathbf{1}_n^\top \mathbf{x}$.

Proof of Proposition 6. First, for each $i = 1, 2, 3$, we may shift $\hat{\boldsymbol{\theta}}_{(i)}$ and $\boldsymbol{\theta}^*_{(i)}$ separately by constant vectors to ensure that so that

$$\mathrm{avg}(\hat{\boldsymbol{\theta}}_{(i)}) = \frac{1}{|I_i|}\mathbf{1}_{|I_i|}^\top \hat{\boldsymbol{\theta}}_{(i)} = 0 \quad \text{and} \quad \mathrm{avg}(\boldsymbol{\theta}^*_{(i)}) = 0,$$

so that

$$d_\infty(\hat{\boldsymbol{\theta}}_{(i)}, \boldsymbol{\theta}^*_{(i)}) = \|\hat{\boldsymbol{\theta}}_{(i)} - \boldsymbol{\theta}^*_{(i)}\|_\infty. \tag{43}$$

Next, for each $i = 1, 2, 3$, let $\tilde{\boldsymbol{\theta}}_{(i)} \in \mathbb{R}^n$ be the augmented version of $\hat{\boldsymbol{\theta}}_{(i)} \in \mathbb{R}^{|I_i|}$, given by

$$\tilde{\boldsymbol{\theta}}_{(i)}(I_i) = \hat{\boldsymbol{\theta}}_{(i)}, \quad \tilde{\boldsymbol{\theta}}_{(i)}(j) = 0,\ j \notin I_i,$$

where $\mathbf{v}(j)$ refers to the j-th entry of vector $\mathbf{v}$. Similarly, we define

$$\tilde{\boldsymbol{\theta}}^*_{(i)}(I_i) = \boldsymbol{\theta}^*_{(i)}, \quad \tilde{\boldsymbol{\theta}}^*_{(i)}(j) = 0,\ j \notin I_i,$$

Step 1. We now define the ensemble MLE $\hat{\boldsymbol{\theta}}$ and show the subadditivity property. The idea is to first fix $\tilde{\boldsymbol{\theta}}_{(1)}$, and then shift the entries of $\tilde{\boldsymbol{\theta}}_{(2)}$ and $\tilde{\boldsymbol{\theta}}_{(3)}$ with coordinates in I_2 and I_3 respectively to comply with the differences in the common entries of I_1, I_3 and I_2, I_3.

Let $S_1 = I_1$, $S_2 = I_2 \setminus I_1$ (note that we don't put any constraint on $I_1 \cap I_2$), $S_3 = I_3 \setminus (I_1 \cup I_2)$. We allow S_3 to be $\emptyset$, but by the assumption that $I_j \not\subseteq I_k$, we have $S_2 \neq \emptyset$. Since $\cup_{j=1}^{3} I_j = [n]$ and $I_i \cap I_3 \neq \emptyset$ for $i = 1, 2$, we have $\cup_{j=1}^{3} S_j = [n]$ and $S_i \cap S_k = \emptyset$ for any $i \neq k$. Pick an arbitrary $t_1 \in I_1 \cap I_3$, $t_2 \in I_2 \cap I_3$, and let

$$\delta_3 = \tilde{\boldsymbol{\theta}}_{(1)}(t_1) - \tilde{\boldsymbol{\theta}}_{(3)}(t_1),\ \delta_2 = \tilde{\boldsymbol{\theta}}_{(3)}(t_2) - \tilde{\boldsymbol{\theta}}_{(2)}(t_2).$$

Notice that since there is no constraint on $I_1 \cap I_2$, t_1 can be equal to t_2. Moreover, define $\check{\boldsymbol{\theta}}_{(3)}$ and $\check{\boldsymbol{\theta}}_{(2)}$ by

$$\check{\boldsymbol{\theta}}_{(3)}(j) = \begin{cases} \tilde{\boldsymbol{\theta}}_{(3)}(j) + \delta_3, & j \in S_3, \\ 0, & j \notin S_3, \end{cases} \qquad \text{and} \quad \check{\boldsymbol{\theta}}_{(2)}(j) = \begin{cases} \tilde{\boldsymbol{\theta}}_{(2)}(j) + \delta_3 + \delta_2, & j \in S_2, \\ 0, & j \notin S_2, \end{cases}$$

respectively. Letting

$$\hat{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}_{(1)} + \check{\boldsymbol{\theta}}_{(2)} + \check{\boldsymbol{\theta}}_{(3)}, \tag{44}$$

it can be seen that

$$\hat{\boldsymbol{\theta}}(j) = \begin{cases} \tilde{\boldsymbol{\theta}}_{(1)}(j), & j \in S_1, \\ \tilde{\boldsymbol{\theta}}_{(2)}(j) + \delta_3 + \delta_2, & j \in S_2, \\ \tilde{\boldsymbol{\theta}}_{(3)}(j) + \delta_3, & j \in S_3. \end{cases}$$

Step 2. Let $\delta_3^* = \tilde{\boldsymbol{\theta}}^*_{(1)}(t_1) - \tilde{\boldsymbol{\theta}}^*_{(3)}(t_1)$, $\delta_2^* = \tilde{\boldsymbol{\theta}}^*_{(3)}(t_2) - \tilde{\boldsymbol{\theta}}^*_{(2)}(t_2)$. Define a new true parameter $\check{\boldsymbol{\theta}}^*$ by shifting $\boldsymbol{\theta}^*$ via

$$\check{\boldsymbol{\theta}}^*(j) = \begin{cases} \tilde{\boldsymbol{\theta}}^*_{(1)}(j), & j \in S_1, \\ \tilde{\boldsymbol{\theta}}^*_{(2)}(j) + \delta_3^* + \delta_2^*, & j \in S_2, \\ \tilde{\boldsymbol{\theta}}^*_{(3)}(j) + \delta_3^*, & j \in S_3. \end{cases}$$

It can be verified that $\check{\boldsymbol{\theta}}^* = \boldsymbol{\theta}^* + (\tilde{\boldsymbol{\theta}}^*(t_1) - \boldsymbol{\theta}^*(t_1))\mathbf{1}_n$, i.e., $\check{\boldsymbol{\theta}}^*$ is a shift of $\boldsymbol{\theta}^*$. To show this, it suffices to show that for all $j \in [n]$,

$$\check{\boldsymbol{\theta}}^*(j) - \boldsymbol{\theta}^*(j) = \tilde{\boldsymbol{\theta}}^*(t_1) - \boldsymbol{\theta}^*(t_1). \tag{45}$$

In fact, for $j \in S_1$, since $\tilde{\boldsymbol{\theta}}^*_{(1)}(I_1)$ is a shift of $\boldsymbol{\theta}^*(I_1)$, Equation (45) holds immediately by the definition of $\check{\boldsymbol{\theta}}^*$. For $j \in S_3$, we have

$$\check{\boldsymbol{\theta}}^*(j) - \boldsymbol{\theta}^*(j) = \tilde{\boldsymbol{\theta}}^*_{(3)}(j) + \tilde{\boldsymbol{\theta}}^*_{(1)}(t_1) - \tilde{\boldsymbol{\theta}}^*_{(3)}(t_1) - \boldsymbol{\theta}^*(j).$$

Since $t_1 \in I_3$ and $\tilde{\boldsymbol{\theta}}^*_{(1)}(I_3)$ is a shift of $\boldsymbol{\theta}^*(I_3)$, it holds that $\tilde{\boldsymbol{\theta}}^*_{(3)}(j) - \boldsymbol{\theta}^*(j) = \tilde{\boldsymbol{\theta}}^*_{(3)}(t_1) - \boldsymbol{\theta}^*(t_1)$ and Equation (45) follows. For $j \in S_2$, we have

$$\begin{aligned}\check{\boldsymbol{\theta}}^*(j) - \boldsymbol{\theta}^*(j) &= \tilde{\boldsymbol{\theta}}^*_{(2)}(j) + \tilde{\boldsymbol{\theta}}^*_{(1)}(t_1) - \tilde{\boldsymbol{\theta}}^*_{(3)}(t_1) + \tilde{\boldsymbol{\theta}}^*_{(3)}(t_2) - \tilde{\boldsymbol{\theta}}^*_{(2)}(t_2) - \boldsymbol{\theta}^*(j)\\ &= \tilde{\boldsymbol{\theta}}^*_{(2)}(j) - \boldsymbol{\theta}^*(j) + \tilde{\boldsymbol{\theta}}^*_{(3)}(t_2) - \tilde{\boldsymbol{\theta}}^*_{(2)}(t_2) + \tilde{\boldsymbol{\theta}}^*_{(1)}(t_1) - \tilde{\boldsymbol{\theta}}^*_{(3)}(t_1).\end{aligned}$$

Since $t_2 \in I_2$ and $\tilde{\boldsymbol{\theta}}^*_{(2)}(I_2)$ is a shift of $\boldsymbol{\theta}^*(I_2)$, it holds that $\tilde{\boldsymbol{\theta}}^*_{(2)}(j) - \boldsymbol{\theta}^*(j) = \tilde{\boldsymbol{\theta}}^*_{(2)}(t_2) - \boldsymbol{\theta}^*(t_2)$ and thus

$$\check{\boldsymbol{\theta}}^*(j) - \boldsymbol{\theta}^*(j) = \tilde{\boldsymbol{\theta}}^*_{(3)}(t_2) - \boldsymbol{\theta}^*(t_2) + \tilde{\boldsymbol{\theta}}^*_{(1)}(t_1) - \tilde{\boldsymbol{\theta}}^*_{(3)}(t_1)$$

Again, since $t_1, t_2 \in I_3$, we have $\tilde{\boldsymbol{\theta}}^*_{(3)}(t_2) - \boldsymbol{\theta}^*(t_2) = \tilde{\boldsymbol{\theta}}^*_{(3)}(t_1) - \boldsymbol{\theta}^*(t_1)$ and Equation (45) follows.

Step 3. Now we are ready to show the conclusion of the proposition. To analyze the error, we first notice that for any $\mathbf{v}, \mathbf{u} \in \mathbb{R}^n$ and $a \in \mathbb{R}$,

$$d_\infty(\mathbf{u}, \mathbf{v}) = d_\infty(\mathbf{u}, \mathbf{v} + a\mathbf{1}_n) \le \|\mathbf{u} - (\mathbf{v} + a\mathbf{1}_n)\|_\infty + \frac{1}{n}\sum_{i=1}^n |u_i - (v_i + a)| \le 2\|\mathbf{u} - (\mathbf{v} + a\mathbf{1}_n)\|_\infty.$$

Since $\check{\boldsymbol{\theta}}^*$ is a shift of $\boldsymbol{\theta}^*$, we have

$$d_\infty(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}^*) \le 2\|\hat{\boldsymbol{\theta}} - \check{\boldsymbol{\theta}}^*\|_\infty.$$

For $j \in S_1$,

$$|\hat{\boldsymbol{\theta}}(j) - \check{\boldsymbol{\theta}}^*(j)| = |\tilde{\boldsymbol{\theta}}_{(1)}(j) - \tilde{\boldsymbol{\theta}}^*_{(1)}(j)| \le \|\hat{\boldsymbol{\theta}}_{(1)} - \boldsymbol{\theta}^*_{(1)}\|_\infty.$$

For $j \in S_3$,

$$|\hat{\boldsymbol{\theta}}(j) - \check{\boldsymbol{\theta}}^*(j)| = |\tilde{\boldsymbol{\theta}}_{(3)}(j) - \tilde{\boldsymbol{\theta}}^*_{(3)}(j)| + |\delta_3 - \delta_3^*| \le 2\|\hat{\boldsymbol{\theta}}_{(1)} - \boldsymbol{\theta}^*_{(1)}\|_\infty + \|\hat{\boldsymbol{\theta}}_{(3)} - \boldsymbol{\theta}^*_{(3)}\|_\infty.$$

For $j \in S_2$.

$$\begin{aligned}|\hat{\boldsymbol{\theta}}(j) - \check{\boldsymbol{\theta}}^*(j)| &= |\tilde{\boldsymbol{\theta}}_{(2)}(j) - \tilde{\boldsymbol{\theta}}^*_{(2)}(j)| + |\delta_3 - \delta_3^*| + |\delta_2 - \delta_2^*|\\ &\le \|\hat{\boldsymbol{\theta}}_{(1)} - \boldsymbol{\theta}^*_{(1)}\|_\infty + 2\|\hat{\boldsymbol{\theta}}_{(2)} - \boldsymbol{\theta}^*_{(2)}\|_\infty + 2\|\hat{\boldsymbol{\theta}}_{(3)} - \boldsymbol{\theta}^*_{(3)}\|_\infty.\end{aligned}$$

Therefore by definition of $\|\cdot\|_\infty$ and Equation (43),

$$\begin{aligned}\|\hat{\boldsymbol{\theta}} - \check{\boldsymbol{\theta}}^*\|_\infty &\le 2(\|\hat{\boldsymbol{\theta}}_{(1)} - \boldsymbol{\theta}^*_{(1)}\|_\infty + \|\hat{\boldsymbol{\theta}}_{(2)} - \boldsymbol{\theta}^*_{(2)}\|_\infty + \|\hat{\boldsymbol{\theta}}_{(3)} - \boldsymbol{\theta}^*_{(3)}\|_\infty)\\ &= 2(d_\infty(\hat{\boldsymbol{\theta}}_{(1)}, \boldsymbol{\theta}^*_{(1)}) + d_\infty(\hat{\boldsymbol{\theta}}_{(2)}, \boldsymbol{\theta}^*_{(2)}) + d_\infty(\hat{\boldsymbol{\theta}}_{(3)}, \boldsymbol{\theta}^*_{(3)})),\end{aligned}$$

and we have

$$\frac{1}{4}d_\infty(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}^*) \le d_\infty(\hat{\boldsymbol{\theta}}_{(1)}, \boldsymbol{\theta}^*_{(1)}) + d_\infty(\hat{\boldsymbol{\theta}}_{(2)}, \boldsymbol{\theta}^*_{(2)}) + d_\infty(\hat{\boldsymbol{\theta}}_{(3)}, \boldsymbol{\theta}^*_{(3)}).$$

□

Lemma 15. *Given a d-Cayley graph, where $(i, j) \in E$ if and only if $i - j \equiv k(\operatorname{mod} n)$ with $-d \le k \le d$, $k \ne 0$. It satisfies $\lambda_2(\mathcal{L}_{\mathbf{A}}) \asymp d^3/n^2$.*

Proof. By definition, a d-Cayley graph is a $2d$-regular graph, and it's well known that (see, e.g. Brouwer and Haemers, 2012) the spectra of its adjacency matrix $\mathbf{A}$ is given by

$$\lambda_j(\mathbf{A}) = \sum_{k=1}^d (\zeta_j^k + \zeta_{-j}^k), \quad \zeta_j := \cos\frac{2\pi j}{n} + \sqrt{-1}\sin\frac{2\pi j}{n},$$

for $j = 1, \cdots, n$. Thus, $\lambda_2(\mathcal{L}_{\mathbf{A}}) = 2d - 2\sum_{k=1}^{d} \cos(\frac{2\pi k}{n})$. Since $\cos(\frac{2\pi k}{n}) = 1 - 2\sin^2 \frac{\pi}{n}k$ and for $k \leq d < 0.5n$, $\sin \frac{\pi}{n}k \in (0.5\frac{\pi}{n}k, \frac{\pi}{n}k)$, we have

$$\lambda_2(\mathcal{L}_{\mathbf{A}}) = c' 4\pi^2/n^2 \sum_{k=1}^{d} k^2 = cd^3/n^2,$$

for some factor $c \in (2\pi^2/3, 4\pi^2/3)$. □

A.6 SPECIAL CASES OF COMPARISON GRAPHS

By Theorem 1, for the estimator $\hat{\boldsymbol{\theta}}_\rho$ to be consistent, L needs to be sufficiently large. We can check some common types of comparison graph topologies and see in what order the necessary sample complexity $N_{\text{comp}} = |E|L$ needs to be, to achieve consistency. To simplify results, we assume $e^{2\kappa_E} \lesssim \log n$. Spectral properties of graphs listed here can be found in well-known textbooks (Brouwer and Haemers, 2012). Shah et al. (2016) provide analogous comparisons. Per Section 2.2 we include results for the d-Caley graphs, and expander graphs here. For reader convenience other results from Section 2.2 are also noted below.

Complete graph: In this case, $\lambda_2(\mathcal{L}_{\mathbf{A}}) = n_{\max} = n_{\min} = n - 1$. Thus, we need $e^{2\kappa_E} \log n/n = o(L)$. Hence $L = \Omega(1)$ and $N_{\text{comp}} = \Omega(n^2)$.

Expander graph: If the comparison graph is a d-regular expander graph with edge expansion (or Cheeger number) coefficient ϕ, then $\lambda_2(\mathcal{L}_{\mathbf{A}}) \geq \phi^2/(2d)$ (Alon et al., 2008), $n_{\max} = n_{\min} = d$. Here ϕ is defined by $\phi := \min_{|S|} e(S,T)/|S|$ where $\{S, T\}$ is a partition of the vertex set and $|S| \leq |T|$. We need $d^2 e^{2\kappa_E}/\phi^4 \cdot (e^{2\kappa_E} n \vee d \log n) = o(L)$, so $N_{\text{comp}} = \Omega(nd^3 e^{2\kappa_E}/\phi^4 \cdot (e^{2\kappa_E} n \vee d \log n))$.

Complete bipartite graph: If the comparison graph has two partitioned sets of size m_1 and m_2 such that $m_1 \leq m_2$, then $\lambda_2(\mathcal{L}_{\mathbf{A}}) = m_1$, $n_{\max} = m_2$, $n_{\min} = m_1$. We need $e^{2\kappa_E} m_2/m_1^2 \cdot [(e^{2\kappa_E} n m_2/m_1^2) \vee \log n] = o(L)$. When $m_1 = \Omega(n)$, we have $N_{\text{comp}} = \Omega(n^2)$.

d-Cayley graph: $(i, j) \in E$ if and only if $i - j \equiv k (\text{mod } n)$ with $-d \leq k \leq d$, $k \neq 0$. It is a $2d$-regular graph, and $\lambda_2(\mathcal{L}_{\mathbf{A}}) \asymp d^3/n^2$ (see Appendix A.5), $n_{\max} = n_{\min} = 2d$. Thus, we need $e^{2\kappa_E} n^4/d^6 \cdot (e^{2\kappa_E} n \vee 2d \log n) = o(L)$, so $N_{\text{comp}} = \Omega(e^{2\kappa_E} n^5/d^5 \cdot (e^{2\kappa_E} n \vee 2d \log n))$ for $d = o(n)$ and $N_{\text{comp}} = \Omega(n^2)$ for $d = \Omega(n)$.

Path or Cycle graph: When comparisons occur based on a path or cycle comparison graph, then by Proposition 3, $\|\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}^*\|_\infty \lesssim e^{\kappa_E}\sqrt{\frac{n \log n}{L}}$. Thus, we need $e^{2\kappa_E} n \log n = o(L)$ and $N_{\text{comp}} = \Omega(e^{2\kappa_E} n^2 \log n)$.

Star graph: A star graph on n node is a tree graph with diameter $D = 1$. By Proposition 4, we have $\|\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}^*\|_\infty \lesssim e^{\kappa_E}\sqrt{\frac{\log n}{L}}$. Thus, we need $e^{2\kappa_E} \log n = o(L)$ and $N_{\text{comp}} = \Omega(e^{2\kappa_E} n \log n)$.

Barbell graph: It contains two size-$n/2$ complete sub-graphs connected by 1 edge, so $\lambda_2(\mathcal{L}_{\mathbf{A}}) \asymp 1/n$, $n_{\max} = n/2$, $n_{\min} = n/2 - 1$. We need $e^{2\kappa_E} n^3 \log n = o(L)$, and $N_{\text{comp}} = e^{2\kappa_E} n^5 \log n$.

A.7 UPPER BOUND FOR UNREGULARIZED/VANILLA MLE

A.7.1 Main theorem

The unregularized or vanilla MLE is defined as

$$\hat{\boldsymbol{\theta}} := \underset{\boldsymbol{\theta}\in\mathbb{R}^n:\ \mathbf{1}_n^\top\boldsymbol{\theta}=0}{\arg\min}\ \ell(\boldsymbol{\theta};\mathbf{y}), \tag{46}$$

where $\ell(\boldsymbol{\theta};\mathbf{y})$ is the negative log-likelihood, given by

$$\ell(\boldsymbol{\theta};\mathbf{y}) := -\sum_{1\leq i<j\leq n} A_{ij}\left\{\bar{y}_{ij}\log\psi(\theta_i-\theta_j)+(1-\bar{y}_{ij})\log[1-\psi(\theta_i-\theta_j)]\right\}, \tag{47}$$

and $t\in\mathbb{R}\mapsto\psi(t)=1/[1+e^{-t}]$ the sigmoid function. To make the expressions of results simpler, we consider the parameter range $\kappa\leq n$ and $\kappa_E\leq\log n$ as discussed in the comments after Theorem 1.

Theorem 16 (Vanilla MLE)**.** *Assume the BTL model with parameter $\boldsymbol{\theta}^*=(\theta_1^*,\ldots,\theta_n^*)^\top$ such that $\mathbf{1}_n^\top\boldsymbol{\theta}^*=0$ and a comparison graph $\mathcal{G}=\mathcal{G}([n],E)$ with adjacency matrix $\mathbf{A}$, algebraic connectivity $\lambda_2(\mathcal{L}_{\mathbf{A}})$ and maximum and minimum degrees $n_{\max}$ and $n_{\min}$, respectively. Suppose that each pair of items $(i,j)\in E$ are compared L times. Let $\kappa=\max_{i,j}|\theta_i^*-\theta_j^*|$ and $\kappa_E=\max_{(i,j)\in E}|\theta_i^*-\theta_j^*|$. Assume that $\mathcal{G}$ is connected, or equivalently, $\lambda_2(\mathcal{L}_A)>0$. In addition, assume that 1. $\lambda_2(\mathcal{L}_A)^2L>Ce^{2\kappa_E}\max\{n_{\max}\log n, e^{2\kappa_E}n\frac{n_{\max}^2}{n_{\min}^2}\}$ for some large constant $C>0$, and 2. $\lambda_2(\mathcal{L}_{\mathbf{A}})\geq 2e^{2\kappa_E}n_{\max}/n_{\min}$. Then, with probability at least $1-O(n^{-5})$, the unregularized MLE $\hat{\boldsymbol{\theta}}$ from* (46) *satisfies*

$$\|\hat{\theta}-\theta^*\|_\infty \lesssim e^{\kappa_E}\sqrt{\frac{n_{\max}\log n}{Ln_{\min}^2}}+s\sqrt{\frac{n_{\max}}{L}}\left[1+\frac{e^{\kappa_E}\sqrt{\log n}}{n_{\min}}+s\sqrt{\frac{n}{n_{\max}}}\right], \tag{48}$$

where $s:=\frac{e^{2\kappa_E}n_{\max}}{\lambda_2 n_{\min}}$, and

$$\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|_2\lesssim\frac{e^{\kappa_E}}{\lambda_2(\mathcal{L}_{\mathbf{A}})}\sqrt{\frac{n_{\max}n}{L}}, \tag{49}$$

provided that $L\lesssim n^5$, $\kappa<n$, $\kappa_E\leq\log n$, and the right hand side of Equation (48) *is smaller than a sufficiently small constant $C>0$.*

Remark 2. The expression of the ℓ_∞ bound looks messy, but in the $ER(n,p)$ case it reduces to the same form as the upper bound of the regularized MLE in Corollary 2. Moreover, for vanilla MLE we require an additional pure topological assumption $\lambda_2(\mathcal{L}_{\mathbf{A}})\geq 2e^{2\kappa_E}n_{\max}/n_{\min}$, which is stringent in some sense as it exclude many graphs with small $n_{\min}$. But for such graphs with high degree heterogeneity $n_{\max}/n_{\min}$, it's reasonable that we need some regularity in our objective function. We believe that this condition can be weakened, and the ℓ_∞ upper bound can be improved or tightened by improving the proof techniques, which can be a good future direction for researchers.

To prove Theorem 16, we need two lemmas, Lemma 17 anbd 18.

Lemma 17. *Under the setting of Theorem 16, it holds with probability at least $1-O(n^{-5})$ that*

$$\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|_\infty\leq 5. \tag{50}$$

Following Chen et al. (2020), we decompose the full negative loglikelihood function as

$$\ell_n(\boldsymbol{\theta})=\ell_n^{(-m)}\left(\boldsymbol{\theta}_{-m}\right)+\ell_n^{(m)}\left(\boldsymbol{\theta}_{-m}\right), \tag{51}$$

where $\theta_m\in\mathbb{R}$ is the m-th entry of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_{-m}\in\mathbb{R}^{n-1}$ is the subvector containing the rest of entries, and the two functions are given by

$$\ell_n^{(-m)}\left(\boldsymbol{\theta}_{-m}\right)=\sum_{1\leq i<j\leq n: i,j\neq m}A_{ij}\left[\bar{y}_{ij}\log\frac{1}{\psi\left(\theta_i-\theta_j\right)}+\left(1-\bar{y}_{ij}\right)\log\frac{1}{1-\psi\left(\theta_i-\theta_j\right)}\right],$$

$$\ell_n^{(m)}\left(\theta_m|\boldsymbol{\theta}_{-m}\right)=\sum_{j\in[n]\backslash\{m\}}A_{mj}\left[\bar{y}_{mj}\log\frac{1}{\psi\left(\theta_m-\theta_j\right)}+\left(1-\bar{y}_{mj}\right)\log\frac{1}{1-\psi\left(\theta_m-\theta_j\right)}\right].$$

Let $H^{(-m)} := \nabla^2 \ell_n^{(-m)}(\ell_{-m})$, and

$$\boldsymbol{\theta}_{-m}^{(m)} = \underset{\boldsymbol{\theta}_{-m}: \|\boldsymbol{\theta}_{-m} - \boldsymbol{\theta}_{-m}^*\|_\infty \leq 5}{\operatorname{argmin}} \ell_n^{(-m)}\left(\boldsymbol{\theta}_{-m}\right).$$

Lemma 18. *Under the setting of Theorem 16, it holds with probability at least* $1 - O(n^{-9})$ *that*

$$\max_{m \in [n]} \|\theta_{-m}^{(m)} - \theta_{-m}^* - a_m \boldsymbol{1}_{n-1}\|_2^2 \leq C \frac{e^{2\kappa_E} n_{\max} n}{L \lambda_2(\mathcal{L}_A)^2} \tag{52}$$

for some constant $C > 0$*, where* $a_m = \operatorname{avg}\left(\theta_{-m}^{(m)} - \boldsymbol{\theta}_{-m}^*\right) := \frac{1}{n-1} \mathbf{1}_{n-1}^\top \left(\theta_{-m}^{(m)} - \boldsymbol{\theta}_{-m}^*\right)$.

Proof of Theorem 16. Again we define the leave-one-out negative log-likelihood as

$$\begin{aligned} \ell_n^{(m)}(\theta) = & \sum_{1 \leq i < j \leq n: i, j \neq m} A_{ij} \left[\bar{y}_{ij} \log \frac{1}{\psi\left(\theta_i - \theta_j\right)} + (1 - \bar{y}_{ij}) \log \frac{1}{1 - \psi\left(\theta_i - \theta_j\right)} \right] \\ & + \sum_{i \in [n] \setminus \{m\}} A_{mj} \left[\psi\left(\theta_i^* - \theta_m^*\right) \log \frac{1}{\psi\left(\theta_i - \theta_m\right)} + \psi\left(\theta_m^* - \theta_i^*\right) \log \frac{1}{\psi\left(\theta_m - \theta_i\right)} \right], \end{aligned} \tag{53}$$

For the ℓ_2 bound, notice that by Taylor expansion

$$\ell_n(\hat{\theta}) = \ell_n(\theta^*) + (\hat{\theta} - \theta^*)^T \nabla \ell_n(\theta^*) + \frac{1}{2} (\hat{\theta} - \theta^*)^T H(\xi) (\hat{\theta} - \theta^*),$$

where ξ is a convex combination of $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^*$. By Lemma 17, we have $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_\infty \leq 5$ and hence $\|\xi - \boldsymbol{\theta}^*\|_\infty \leq 5$. By Lemma 2, we have $\frac{1}{2}(\hat{\theta} - \theta^*)^T H(\xi)(\hat{\theta} - \theta^*) \geq c e^{-\kappa_E} \lambda_2 \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2^2$ for some constant $c > 0$. By the fact that $\ell_n(\boldsymbol{\theta}^*) \geq \ell_n(\hat{\boldsymbol{\theta}})$, Cauchy-Schwartz inequality, and Lemma 1, we have

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2 \lesssim \frac{e^{\kappa_E}}{\lambda_2} \|\nabla \ell(\boldsymbol{\theta}^*)\|_2 \leq \frac{e^{\kappa_E}}{\lambda_2} \sqrt{\frac{n_{\max} n}{L}}.$$

For the ℓ_∞ bound, the proof can be sketched as following steps.

0. By Lemma 17, $\|\hat{\theta} - \theta^*\|_\infty \leq 5$ with probability at least $1 - O(n^{-5})$.
1. By Lemma 18, it holds with probability exceeding $1 - O(n^{-9})$ that,

$$\max_{m \in [n]} \|\theta_{-m}^{(m)} - \theta_{-m}^* - a_m \mathbf{1}_{n-1}\|_2^2 \leq C \frac{e^{2\kappa_E} n_{\max} n}{L \lambda_2(\mathcal{L}_A)^2}$$

   where $a_m = \operatorname{avg}\left(\theta_{-m}^{(m)} - \boldsymbol{\theta}_{-m}^*\right) := \frac{1}{n-1} \mathbf{1}_{n-1}^\top \left(\theta_{-m}^{(m)} - \boldsymbol{\theta}_{-m}^*\right)$.

2. Show that on the same event,

$$\begin{aligned} \max_{m \in [n]} \|\theta_{-m}^{(m)} - \widehat{\theta}_{-m} - a_m \mathbf{1}_{n-1}\|_2^2 \leq & C_1 \frac{\max_{m \in [n]} \sum_{i \in \mathcal{N}(m)} \left(\bar{y}_{mi} - \psi\left(\theta_m^* - \theta_i^*\right)\right)^2}{\lambda_2(\mathcal{L}_A)^2 e^{-2\kappa_E}} \\ & + C_1 \frac{e^{2\kappa_E} n_{\max}}{\lambda_2(\mathcal{L}_A)^2} \|\widehat{\theta} - \theta^*\|_\infty^2. \end{aligned} \tag{54}$$

   Following the same arguments towards Equation (66), we can get

$$\|\boldsymbol{\theta}_{-m}^{(m)} - \hat{\boldsymbol{\theta}}_{-m} - \bar{a}_m \mathbf{1}_{n-1}\|_2^2 \leq \frac{e^{2\kappa_E}}{c^2 \lambda_2(\mathcal{L}_A)^2} \|\nabla \ell_n^{(-m)}(\hat{\boldsymbol{\theta}}_{-m})\|_2^2.$$

   By Equation (51), for each $i \in [n] \setminus \{m\}$, we have

$$\frac{\partial}{\partial \boldsymbol{\theta}_i} \ell_n^{(-m)}\left(\boldsymbol{\theta}_{-m}\right) = \frac{\partial}{\partial \theta_i} \ell_n(\boldsymbol{\theta}) - \frac{\partial}{\partial \theta_i} \ell_n^{(m)}\left(\theta_m \mid \boldsymbol{\theta}_{-m}\right),$$

Using the fact that $\nabla \ell_n(\hat{\boldsymbol{\theta}}) = 0$, we get

$$\frac{\partial}{\partial \theta_i}\ell_n^{(-m)}\left(\boldsymbol{\theta}_{-m}\right)_{|\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = -\frac{\partial}{\partial \theta_i}\ell_n^{(m)}\left(\theta_m \mid \boldsymbol{\theta}_{-m}\right)|_{\theta=\widehat{\boldsymbol{\theta}}} = -A_{mi}\left[\bar{y}_{mi} - \psi(\widehat{\theta}_m - \widehat{\theta}_i)\right].$$

Therefore, we have

$$\begin{aligned}
\|\nabla \ell_n^{(-m)}(\widehat{\boldsymbol{\theta}}_{-m})\|_2^2 = & \sum_{i\in[n]\backslash\{m\}} A_{mi}\left[\bar{y}_{mi} - \psi(\widehat{\theta}_m - \widehat{\theta}_i)\right]^2 \\
\leq & 2\sum_{i\in[n]\backslash\{m\}} A_{mi}\left(\bar{y}_{mi} - \psi\left(\theta_m^* - \theta_i^*\right)\right)^2 \\
& + 2\sum_{i\in[n]\backslash\{m\}} A_{mi}\left[\psi\left(\theta_m^* - \theta_i^*\right) - \psi(\widehat{\theta}_m - \widehat{\theta}_i)\right]^2 \\
\leq & 2\sum_{i\in[n]\backslash\{m\}} A_{mi}\left(\bar{y}_{mi} - \psi\left(\theta_m^* - \theta_i^*\right)\right)^2 + 2\|\widehat{\theta} - \theta^*\|_\infty^2 \sum_{i\in[n]\backslash\{m\}} A_{mi} \\
\leq & 2\sum_{i\in[n]\backslash\{m\}} A_{mi}\left(\bar{y}_{mi} - \psi\left(\theta_m^* - \theta_i^*\right)\right)^2 + 4n_{\max}\|\widehat{\theta} - \theta^*\|_\infty^2.
\end{aligned}$$

Now use the fact that $\mathbf{1}_n^\top \boldsymbol{\theta}^* = \mathbf{1}_n^\top \hat{\boldsymbol{\theta}} = 0$, we have

$$\|a_m \mathbf{1}_{n-1} - \bar{a}_m \mathbf{1}_{n-1}\|_2^2 = (n-1)[\text{avg}(\widehat{\boldsymbol{\theta}}_{-m} - \boldsymbol{\theta}_{-m}^*)]^2 = \frac{(\hat{\boldsymbol{\theta}}_m - \theta_m^*)^2}{n-1} \leq \frac{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_\infty^2}{n-1}.$$

These results, together with the fact that $\lambda_2(\mathcal{L}_{\mathbf{A}}) \leq 2n_{\max} \leq 2n$, give Equation (54).

3. Show that on the same event,

$$\begin{aligned}
\|\widehat{\theta} - \theta^*\|_\infty \cdot C_4 e^{-\kappa_E} n_{\min} \leq \max_{m\in[n]} | & \sum_{i\in\mathcal{N}(m)} \left(\bar{y}_{mi} - \psi\left(\theta_m^* - \theta_i^*\right)\right)| \\
& + \sqrt{n_{\max}} \max_{m\in[n]} \|\theta_{-m}^{(m)} - \theta_{-m}^* - a_m \mathbf{1}_{n-1}\|_2 \\
& + \sqrt{n_{\max}} \max_{m\in[n]} \|\theta_{-m}^{(m)} - \widehat{\theta}_{-m} - a_m \mathbf{1}_{n-1}\|_2.
\end{aligned} \tag{55}$$

First, define two univariate functions as some proxy of gradient and hessian:

$$\begin{aligned}
g^{(m)}\left(\theta_m \mid \boldsymbol{\theta}_{-m}\right) &= \frac{\partial}{\partial \theta_m}\ell_n^{(m)}\left(\theta_m \mid \boldsymbol{\theta}_{-m}\right) = -\sum_{i\in[n]\backslash\{m\}} A_{mi}\left(\bar{y}_{mi} - \psi\left(\theta_m - \theta_i\right)\right) \\
h^{(m)}\left(\theta_m \mid \boldsymbol{\theta}_{-m}\right) &= \frac{\partial^2}{\partial \theta_m^2}\ell_n^{(m)}\left(\theta_m \mid \boldsymbol{\theta}_{-m}\right) = \sum_{i\in[n]\backslash\{m\}} A_{mi}\psi\left(\theta_m - \theta_i\right)\psi\left(\theta_i - \theta_m\right).
\end{aligned}$$

By the definition of $\hat{\boldsymbol{\theta}}$ and the shift invariance of ℓ_n, we have $\ell_n(\hat{\boldsymbol{\theta}}) \leq \ell_n(\boldsymbol{\theta})$ for any $\boldsymbol{\theta} \in \mathbb{R}^n$, thus

$$\ell_n^{(m)}(\theta_m^* \mid \hat{\boldsymbol{\theta}}_{-m}) + \ell_n^{(-m)}(\hat{\boldsymbol{\theta}}_{-m}) \geq \ell_n(\hat{\boldsymbol{\theta}}).$$

This implies

$$\begin{aligned}
\ell_n^{(m)}(\theta_m^* \mid \hat{\theta}_{-m}) \geq & \ell_n^{(m)}(\hat{\theta}_m \mid \hat{\theta}_{-m}) \\
& = \ell_n^{(m)}(\theta_m^* \mid \hat{\boldsymbol{\theta}}_{-m}) + (\hat{\theta}_m - \theta_m^*) g^{(m)}(\theta_m^* \mid \hat{\boldsymbol{\theta}}_{-m}) + \frac{1}{2}(\hat{\theta}_m - \theta_m^*)^2 h^{(m)}(\xi \mid \hat{\boldsymbol{\theta}}_{-m}),
\end{aligned}$$

where ξ is a convex combination of θ_m^* and $\hat{\theta}_m$. By Lemma 17, we have $|\xi - \theta_m^*| \leq |\hat{\theta}_m - \theta_m^*| \leq 5$. Thus for any $i \neq m$ it holds that $|\xi - \hat{\theta}_i| \leq |\xi - \theta_m^*| + |\theta_m^* - \theta_i^*| + \left|\widehat{\theta}_i - \theta_i^*\right| \leq 10 + \kappa$. By definition of $h^{(m)}$, we have $\frac{1}{2}h^{(m)}(\xi|\hat{\boldsymbol{\theta}}_{-m}) \geq c_2 e^{-\kappa_E} n_{\min}$ for some constant $c_2 > 0$. Therefore, we get

$$(\hat{\theta}_m - \theta_m^*)^2 \leq \frac{e^{2\kappa_E}}{(c_2 n_{\min})^2}|g^{(m)}(\theta_m^* \mid \hat{\boldsymbol{\theta}}_{-m})|^2. \tag{56}$$

To bound $|g^{(m)}(\theta^*_m \mid \hat{\boldsymbol{\theta}}_{-m})|$, we decompose it as

$$
\begin{aligned}
|g^{(m)}(\theta^*_m \mid \hat{\theta}_{-m})| &= \left| \sum_{i\in[n]\setminus\{m\}} A_{mi}(\bar{y}_{mi} - \psi(\theta^*_m - \widehat{\theta}_i)) \right| \\
&\leq \left| \sum_{i\in[n]\setminus\{m\}} A_{mi}(\bar{y}_{mi} - \psi(\theta^*_m - \theta^*_i)) \right| && (57) \\
&\quad + \left| \sum_{i\in[n]\setminus\{m\}} A_{mi}(\psi(\theta^*_m - \theta^*_i) - \psi(\theta^*_m - \theta^{(m)}_i + a_m)) \right| && (58) \\
&\quad + \left| \sum_{i\in[n]\setminus\{m\}} A_{mi}(\psi(\theta^*_m - \theta^{(m)}_i + a_m) - \psi(\theta^*_m - \widehat{\theta}_i)) \right|. && (59)
\end{aligned}
$$

By Cauchy-Schwartz inequality, we can bound (58) and (59) by

$$
\begin{aligned}
\left| \sum_{i\in[n]\setminus\{m\}} A_{mi}(\psi(\theta^*_m - \theta^*_i) - \psi(\theta^*_m - \theta^{(m)}_i + a_m)) \right|^2 &\leq n_{\max} \|\boldsymbol{\theta}^{(m)}_{-m} - \boldsymbol{\theta}^*_{-m} - a_m \mathbf{1}_{n-1}\|_2^2, \\
\left| \sum_{i\in[n]\setminus\{m\}} A_{mi}(\psi(\theta^*_m - \theta^{(m)}_i + a_m) - \psi(\theta^*_m - \widehat{\theta}_i)) \right|^2 &\leq n_{\max} \|\boldsymbol{\theta}^{(m)}_{-m} - \hat{\boldsymbol{\theta}}_{-m} - a_m \mathbf{1}_{n-1}\|_2^2.
\end{aligned}
$$

Plugging these bounds into Equation (56) and taking maximum over $m \in [n]$ give the desired bound (55).

4. Plug (55) back into (54) and get

$$
\begin{aligned}
\max_{m\in[n]} \|\theta^{(m)}_{-m} - \widehat{\theta}_{-m} - a_m \mathbf{1}_{n-1}\|_2^2 \lesssim & \frac{e^{2\kappa_E}}{\lambda_2(\mathcal{L}_A)^2} \max_m \sum_{i\in\mathcal{N}(m)} (\bar{y}_{mi} - \psi^*_{mi})^2 \\
& + \frac{e^{4\kappa_E}}{\lambda_2(\mathcal{L}_A)^2} \frac{n_{\max}}{n_{\min}^2} \max_{m\in[n]} | \sum_{i\in\mathcal{N}(m)} (\bar{y}_{mi} - \psi^*_{mi})|^2 \\
& + \frac{e^{4\kappa_E}}{\lambda_2(\mathcal{L}_A)^2} \frac{n_{\max}^2}{n_{\min}^2} \max_{m\in[n]} \|\theta^{(m)}_{-m} - \theta^*_{-m} - a_m \mathbf{1}_{n-1}\|_2^2 \\
& + \frac{e^{4\kappa_E}}{\lambda_2(\mathcal{L}_A)^2} \frac{n_{\max}^2}{n_{\min}^2} \max_{m\in[n]} \|\theta^{(m)}_{-m} - \hat{\theta}_{-m} - a_m \mathbf{1}_{n-1}\|_2^2.
\end{aligned}
$$

as we assume $\frac{e^{2\kappa_E}}{\lambda_2(\mathcal{L}_A)} \frac{n_{\max}}{n_{\min}} \leq \frac{1}{2}$, we have

$$
\begin{aligned}
\max_{m\in[n]} \|\theta^{(m)}_{-m} - \widehat{\theta}_{-m} - a_m \mathbf{1}_{n-1}\|_2^2 \lesssim & \frac{e^{2\kappa_E}(\log n + n_{\max})}{\lambda_2(\mathcal{L}_A)^2 L} + \frac{e^{4\kappa_E}}{\lambda_2(\mathcal{L}_A)^2} \frac{n_{\max}^2}{n_{\min}^2} \frac{\log n}{L} + \\
& \frac{e^{4\kappa_E}}{\lambda_2(\mathcal{L}_A)^2} \frac{n_{\max}^2}{n_{\min}^2} \frac{e^{2\kappa_E} n_{\max} n}{\lambda_2(\mathcal{L}_A)^2 L}.
\end{aligned} \tag{60}
$$

5. Plug (60) back into (55) and we can get

$$
\begin{aligned}
\|\widehat{\theta} - \theta^*\|_\infty^2 \lesssim & \frac{e^{2\kappa_E}}{n_{\min}^2} \frac{n_{\max} \log n}{L} + \frac{e^{2\kappa_E} n_{\max}^2}{n_{\min}^2 n} \frac{e^{2\kappa_E} n_{\max} n}{\lambda_2(\mathcal{L}_A)^2 L} \\
& + \frac{e^{2\kappa_E} n_{\max}}{n_{\min}^2 \lambda_2(\mathcal{L}_A)^2 L} \left[e^{2\kappa_E}(\log n + n_{\max}) + \frac{e^{4\kappa_E} n_{\max}^2 \log n}{n_{\min}^2} + \frac{e^{6\kappa_E} n_{\max}^3 n}{\lambda_2(\mathcal{L}_A)^2 n_{\min}^2} \right]
\end{aligned} \tag{61}
$$

One term can be reduced and the inequality becomes

$$
\|\hat{\theta} - \theta^*\|_\infty \lesssim e^{\kappa_E} \sqrt{\frac{n_{\max} \log n}{L n_{\min}^2}} + e^{2\kappa_E} \sqrt{\frac{n_{\max}^3}{L \lambda_2(\mathcal{L}_A)^2 n_{\min}^2}} \left[1 + e^{\kappa_E} \sqrt{\frac{\log n}{n_{\min}^2}} + e^{2\kappa_E} \sqrt{\frac{n_{\max} n}{\lambda_2(\mathcal{L}_A)^2 n_{\min}^2}} \right]. \tag{62}
$$

□

A.7.2 Proof of Lemmas

Proof of Lemma 17. We will use a gradient descent sequence defined by

$$\theta^{(t+1)} = \theta^{(t)} - \eta[\nabla \ell_n(\theta^{(t)}) + \rho\theta^{(t)}].$$

The leave-one-out negative log-likelihood is defined as

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

so the leave-one-out gradient descent sequence is defined as

$$\theta^{(t+1,m)} = \theta^{(t,m)} - \eta[\nabla \ell_n^{(m)}(\theta^{(t,m)}) + \rho\theta^{(t,m)}].$$

We initialize both sequences by $\theta^{(0)} = \theta^{(0,m)} = \theta^*$ and set $\rho = \frac{1}{\kappa}\sqrt{\frac{n_{\max}}{L}}$ and step size $\eta = \frac{1}{\lambda + n_{\max}}$. We will show that under the assumption $\lambda_2(\mathcal{L}_A)^2 L > Ce^{2\kappa_E}\max\{n_{\max}\log n, e^{2\kappa_E} n n_{\max}^2/n_{\min}^2\}$ for some large constant $C>0$, we have

$$\begin{aligned}\max_{m\in[n]}\|\theta^{(t,m)}-\theta^{(t)}\|_2 \le & f_1 := C_1\frac{e^{\kappa_E}}{\lambda_2}\sqrt{\frac{n_{\max}\log n}{L}} \le 1\\ \|\theta^{(t)}-\theta^*\|_2 \le & f_2 := C_2\frac{e^{\kappa_E}}{\lambda_2}\sqrt{\frac{n_{\max} n}{L}} \le \sqrt{\frac{n}{\log n}}\\ \max_{m\in[n]}|\theta_m^{(t,m)}-\theta_m^*| \le & f_3 := C_3\frac{e^{2\kappa_E}}{\lambda_2}\frac{n_{\max}}{n_{\min}}\sqrt{\frac{n}{L}} \le 1.\end{aligned} \tag{63}$$

A useful fact given that (63) holds is that

$$\|\theta^{(t,m)}-\theta^*\|_\infty \le f_1 + f_2. \tag{64}$$

We again have the Taylor expansion

$$\theta^{(t+1)} - \theta^{(t+1,m)} = [(1-\eta\rho)I_n - \eta H(\xi)](\theta^{(t)}-\theta^{(t,m)}) - \eta[\nabla\ell_n(\theta^{(t,m)}) - \nabla\ell_n^{(m)}(\theta^{(t,m)})].$$

Now by the fact that $\lambda_{\min,\perp}(H(\xi)) \ge c_0 e^{-\kappa}\lambda_2$, we have

$$\|\left((1-\eta\rho)I_n - \eta H(\xi)\right)(\theta^{(t)}-\theta^{(t,m)})\|_2 \le (1-\eta\rho - c_1\eta\lambda_2)\|\theta^{(t)}-\theta^{(t,m)}\|_2$$

for some constant $c_1 > 0$ and the other term can be bounded as

$$\begin{aligned}&\|\nabla\ell_n(\theta^{(t,m)}) - \nabla\ell_n^{(m)}(\theta^{(t,m)})\|_2^2\\ =&\left[\sum_{j\in[n]\setminus\{m\}} A_{jm}\left(\bar{y}_{jm} - \psi\left(\theta_j^*-\theta_m^*\right)\right)\right]^2 + \sum_{j\in[n]\setminus\{m\}} A_{jm}\left(\bar{y}_{jm} - \psi\left(\theta_j^*-\theta_m^*\right)\right)^2\\ \le& C_1\frac{1}{L}n_{\max}\log n + C_1\frac{1}{L}(\log n + n_{\max}).\end{aligned}$$

Therefore, for

$$\|\theta^{(t+1)}-\theta^{(t+1,m)}\|_2 \le (1-c_1\eta\lambda_2)f_1 + \eta\sqrt{2C_1\frac{1}{L}n_{\max}\log n} \le f_1$$

to hold, we need $f_1 > C\frac{e^{\kappa_E}}{\lambda_2}\sqrt{\frac{n_{\max}\log n}{L}}$ for some sufficiently large positive constant $C>0$.

Next, we bound $\|\theta^{(t+1)}-\theta^*\|$. By Tarlor expansion,

$$\theta^{(t+1)}-\theta^* = \left((1-\eta\lambda)I_n - \eta H(\xi)\right)\left(\theta^{(t)}-\theta^*\right) - \eta\lambda\theta^* - \eta\nabla\ell_n\left(\theta^*\right).$$

Equation (41) becomes

$$((1-\eta\lambda)I_n - \eta H(\xi))\,(\theta^{(t)}-\theta^*) \le (1-\eta\lambda - c_2\eta\lambda_2)\,\|\theta^{(t)}-\theta^*\|_2,$$

and equation (42) becomes

$$\|\nabla \ell_n(\theta^*)\|_2^2 = \sum_{i=1}^n \left[\sum_{j\in[n]\setminus\{i\}} A_{ij}\left(\bar{y}_{ij} - \psi\left(\theta_i^* - \theta_j^*\right)\right)\right]^2 \le C_2 \frac{nn_{\max}}{L},$$

for some constants $c_2, C_2 > 0$. Therefore,

$$\|\theta^{(t+1)}-\theta^*\|_2 \le (1-c_2\eta\lambda_2)f_2 + \eta\sqrt{C_2\frac{nn_{\max}}{L}} + \eta\lambda\|\theta^*\|_2.$$

For $\|\theta^{(t+1)}-\theta^*\| \le f_2$ to hold, we need $\frac{\lambda_2}{e^{\kappa_E}}f_2 > C\sqrt{\frac{n_{\max}n}{L}}$ for some sufficiently large constant C, which is guaranteed by the definition of f_2.

Next, we bound $|\theta_m^{(t+1,m)} - \theta_m^*|$. Note that by the definition of the gradient descent

$$\theta_m^{(t+1,m)} - \theta_m^* = \left[1-\eta\lambda - \eta\sum_{j\in[n]\setminus\{m\}} A_{mj}\psi'(\xi_j)\right](\theta_m^{(t,m)} - \theta_m^*) - \lambda\eta\theta_m^* + \eta\sum_{j\in[n]\setminus\{m\}} A_{mj}\psi'(\xi_j)(\theta_j^{(t,m)} - \theta_j^*)$$

where ξ_j is a scalar between $\theta_m^* - \theta_j^*$ and $\theta_m^{(t,m)} - \theta_j^{(t,m)}$. Since $\|\theta^{(t,m)} - \theta^*\|_\infty \le 3$, we have $|\xi_j - \theta_m^* + \theta_j^*| \le |\theta_m^* - \theta_j^* - \theta_m^{(t,m)} + \theta_j^{(t,m)}| \le 6$ and $\|\xi\|_\infty$ is bounded. Therefore,

$$\sum_{j\in[n]\setminus\{m\}} A_{mj}\psi'(\xi_j) \ge c_3 \min_{i\in[n]} n_i$$

and

$$\begin{aligned}|\sum_{j\in\mathcal{N}(m)} A_{mj}\psi'(\xi_j)(\theta_j^{(t,m)} - \theta_j^*)| &\le \sqrt{\sum_{j\in\mathcal{N}(m)}[A_{mj}\psi'(\xi_j)]^2}\sqrt{\sum_{j\in\mathcal{N}(m)}(\theta_j^{(t,m)} - \theta_j^*)^2} \\ &\le c_4\sqrt{n_{\max}}(f_1+f_2).\end{aligned}$$

for some constant $c_3, c_4 > 0$. Thus we have

$$|\theta_m^{(t+1,m)} - \theta_m^*| \le (1-c_3\eta n_{\min}) + \eta\sqrt{n_{\max}}(f_1+f_2) + \lambda\eta|\theta_m^*|.$$

For $|\theta_m^{(t+1,m)} - \theta_m^*| \le f_3$ to hold, we need $\frac{n_{\min}}{e^{\kappa_E}}f_3 > C\sqrt{n_{\max}}f_2$ for some sufficiently large constant $C > 0$, which is ensured by the definition of f_2, f_3.

As the last step, we again use the fact that $\ell_\rho(\cdot)$ is ρ-strongly convex and $(\rho + n_{\max})$-smooth (see definition in the paragraph before Equation (1)), so by Theorem 3.10 in Bubeck (2015), we have

$$\|\theta^{(t)} - \hat{\theta}_\rho\|_2 \le (1-\frac{\rho}{\rho+n_{\max}})^t\|\theta^* - \hat{\theta}_\rho\|_2. \tag{65}$$

By a union bound, Equation (63) holds for all $t \le T$ with probability at leaast $1-O(Tn^{-10})$. Triangle inequality implies that

$$\|\hat{\theta}_\rho - \theta^*\|_\infty \le \|\theta^{(T)} - \hat{\theta}_\rho\|_2 + \|\theta^{(T)} - \theta^*\|_\infty \le (1-\frac{\rho}{\rho+n_{\max}})^T\sqrt{n}\|\hat{\theta}_\rho - \theta^*\|_\infty + 2$$

Take $T = n^5$ and remember that $L \lesssim n^5$. If $\rho > n_{\max}$, then $(1-\frac{\rho}{\rho+n_{\max}})^T\sqrt{n} \le 2^{-n^5}\sqrt{n} \le 1/2$. Otherwise, since

$$(1-\frac{\rho}{\rho+n_{\max}})^T\sqrt{n} \le \exp\left(-\frac{T\rho}{\rho+n_{\max}}\right)\sqrt{n},$$

using the fact that $\kappa < n$, we have

$$(1-\frac{\rho}{\rho+n_{\max}})^T\sqrt{n} \le \exp\left(-\frac{T}{c\kappa}\sqrt{\frac{1}{n_{\max}L}}\right)\kappa\sqrt{n} \le ce^{-n}n^{3/2} \le \frac{1}{2}.$$

In conclusion, we have $\|\hat{\theta}_\rho - \theta^*\|_\infty \le \frac{1}{2}\|\hat{\theta}_\rho - \theta^*\|_\infty + 2$, thus $\|\hat{\theta}_\rho - \theta^*\|_\infty \le 4$, with probability at least $1-O(n^{-5})$. □

Proof of Lemma 18. By definition, $\boldsymbol{\theta}_{-m}^{(m)}$ is a constrained MLE on a subset of the data, thus by Taylor expansion, for ξ given by a convex combination of $\boldsymbol{\theta}_{-m}^{*}$ and $\boldsymbol{\theta}_{-m}^{(m)}$, we have

$$\begin{aligned}\ell_n^{(-m)}\left(\boldsymbol{\theta}_{-m}^{*}\right) \geq& \ell_n^{(-m)}(\boldsymbol{\theta}_{-m}^{(m)})\\ =&\ell_n^{(-m)}\left(\boldsymbol{\theta}_{-m}^{*}\right) + (\boldsymbol{\theta}_{-m}^{(m)} - \boldsymbol{\theta}_{-m}^{*} - a_m\mathbf{1}_{n-1})^T\nabla\ell_n^{(-m)}\left(\boldsymbol{\theta}_{-m}^{*}\right)\\ &+\frac{1}{2}(\boldsymbol{\theta}_{-m}^{(m)} - \boldsymbol{\theta}_{-m}^{*} - a_m\mathbf{1}_{n-1})^T H^{(-m)}(\xi)(\boldsymbol{\theta}_{-m}^{(m)} - \boldsymbol{\theta}_{-m}^{*} - a_m\mathbf{1}_{n-1}),\end{aligned}$$

where we use the invariant property of $\ell_n^{(-m)}\left(\boldsymbol{\theta}_{-m}\right)$, i.e., $\ell_n^{(-m)}\left(\boldsymbol{\theta}_{-m}\right) = \ell_n^{(-m)}\left(\boldsymbol{\theta}_{-m} + c\mathbf{1}_{n-1}\right)$. By the fact that $\|\xi - \boldsymbol{\theta}_{-m}^{*}\|_\infty \leq \|\boldsymbol{\theta}_{-m}^{(m)} - \boldsymbol{\theta}_{-m}^{*}\|_\infty \leq 5$ and Lemma 2, we have

$$(\boldsymbol{\theta}_{-m}^{(m)} - \boldsymbol{\theta}_{-m}^{*} - a_m\mathbf{1}_{n-1})^T H^{(-m)}(\xi)(\boldsymbol{\theta}_{-m}^{(m)} - \boldsymbol{\theta}_{-m}^{*} - a_m\mathbf{1}_{n-1}) \geq ce^{-\kappa_E}\lambda_2(\mathcal{L}_{A_{-m}})\|\boldsymbol{\theta}_{-m}^{(m)} - \boldsymbol{\theta}_{-m}^{*} - a_m\mathbf{1}_{n-1}\|_2^2.$$

Applying Cauchy-Schwartz inequality to the expansion and we can get

$$\|\boldsymbol{\theta}_{-m}^{(m)} - \boldsymbol{\theta}_{-m}^{*} - a_m\mathbf{1}_{n-1}\|_2^2 \leq \frac{e^{2\kappa_E}}{c^2\lambda_2(\mathcal{L}_{A_{-m}})^2}\|\nabla\ell_n^{(-m)}(\boldsymbol{\theta}_m^{*})\|_2^2.$$

Where A_{-m} is the adjacency matrix of the comparison graph with node m excluded. By the interlacing property of the eigenvalue sequences of Laplacians of graph and its induced subgraph (see, e.g. Brouwer and Haemers, 2012, Proposition 3.2.1), we have $\lambda_2(\mathcal{L}_{A_{-m}}) \geq \lambda_2(\mathcal{L}_A)$, thus

$$\|\boldsymbol{\theta}_{-m}^{(m)} - \boldsymbol{\theta}_{-m}^{*} - a_m\mathbf{1}_{n-1}\|_2^2 \leq \frac{e^{2\kappa_E}}{c^2\lambda_2(\mathcal{L}_{A})^2}\|\nabla\ell_n^{(-m)}(\boldsymbol{\theta}_m^{*})\|_2^2. \tag{66}$$

Now by Lemma 1, it holds with probability at least $1 - O(n^{-10})$ that

$$\|\boldsymbol{\theta}_{-m}^{(m)} - \boldsymbol{\theta}_{-m}^{*} - a_m\mathbf{1}_{n-1}\|_2^2 \leq C\frac{e^{2\kappa_E}n_{\max}n}{L\lambda_2(\mathcal{L}_A)^2},$$

and the conclusion is guaranteed by a union bound. □

Bibliography

Arpit Agarwal, Prathamesh Patil, and Shivani Agarwal. Accelerated Spectral Ranking. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pages 70–79, 2018. URL http://proceedings.mlr.press/v80/agarwal18b.html.

Noga Alon, Oded Schwartz, and Asaf Shapira. An Elementary Construction of Constant-Degree Expanders. *Comb. Probab. Comput.*, 17(3):319–327, May 2008. ISSN 0963-5483. doi: 10.1017/S0963548307008851. URL https://doi.org/10.1017/S0963548307008851.

Andries E. Brouwer and Willem H. Haemers. *Spectra of graphs.* Springer, 2012.

Sébastien Bubeck. Convex Optimization: Algorithms and Complexity. *Found. Trends Mach. Learn.*, 8(3–4):231–357, November 2015. ISSN 1935-8237. doi: 10.1561/2200000050. URL https://doi.org/10.1561/2200000050.

Pinhan Chen, Chao Gao, and Anderson Y. Zhang. Partial Recovery for Top-k Ranking: Optimality of MLE and Sub-Optimality of Spectral Method. *arXiv:2006.16485*, 2020. URL http://arxiv.org/abs/2006.16485v1.

Yuxin Chen, Jianqing Fan, Cong Ma, and Kaizheng Wang. Spectral method and regularized MLE are both optimal for top-K ranking. *Ann. Statist.*, 47(4):2204–2235, 2019. ISSN 0090-5364. doi: 10.1214/18-AOS1745.

Bruce E. Hajek, Sewoong Oh, and Jiaming Xu. Minimax-optimal Inference from Partial Rankings. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 1475–1483, 2014. URL http://papers.nips.cc/paper/5361-minimax-optimal-inference-from-partial-rankings.

Ruijian Han, Rougang Ye, Chunxi Tan, and Kani Chen. Asymptotic theory of sparse Bradley–Terry model. *Ann. Appl. Probab.*, 30(5):2491–2515, 2020. ISSN 1050-5164. doi: 10.1214/20-AAP1564.

Julien M. Hendrickx, Alex Olshevsky, and Venkatesh Saligrama. Minimax Rate for Learning From Pairwise Comparisons in the BTL Model. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 4193–4202. PMLR, 2020. URL `http://proceedings.mlr.press/v119/hendrickx20a.html`.

Sahand Negahban, Sewoong Oh, and Devavrat Shah. Rank Centrality: Ranking from Pairwise Comparisons. *Oper. Res.*, 65 (1):266–287, 2017. doi: 10.1287/opre.2016.1534.

Nihar B. Shah, Sivaraman Balakrishnan, Joseph Bradley, Abhay Parekh, Kannan Ramchandran, and Martin J. Wainwright. Estimation from Pairwise Comparisons: Sharp Minimax Bounds with Topology Dependence. *Journal of Machine Learning Research*, 17(58):1–47, 2016. URL `http://jmlr.org/papers/v17/15-189.html`.

Gordon Simons and Yi-Ching Yao. Asymptotics when the number of parameters tends to infinity in the Bradley-Terry model for paired comparisons. *Ann. Statist.*, 27(3):1041–1060, 06 1999. doi: 10.1214/aos/1018031267. URL `https://doi.org/10.1214/aos/1018031267`.

Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices, 2011.

Holger Wendland. *Numerical linear algebra*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2018. ISBN 978-1-316-60117-4; 978-1-107-14713-3. An introduction.

Ting Yan, Yaning Yang, and Jinfeng Xu. Sparse Paired Comparisons in the Bradley-Terry Model. *Statistica Sinica*, 22(3): 1305–1318, 2012. ISSN 10170405, 19968507. URL `http://www.jstor.org/stable/24309985`.

Bin Yu. Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam*, pages 423–435. Springer, New York, 1997.