SUPPLEMENT TO: WORST-CASE VS AVERAGE-CASE DESIGN FOR ESTIMATION FROM PARTIAL PAIRWISE COMPARISONS

BY ASHWIN PANANJADY*, CHENG MAO†, VIDYA MUTHUKUMAR*, MARTIN J. WAINWRIGHT*,†, AND THOMAS A. COURTADE*

Departments of EECS* and Statistics†, UC Berkeley
Department of Mathematics, MIT‡

In the first few sections of the appendix, we provide proofs for lemmas used in the main body of the paper. In Appendix F, we analyze the minimax denoising error of the estimation problem.

Recall our assumptions throughout that \( n \geq 2 \), and that \( c, c' \) are used to denote universal constants that may change from line to line.

APPENDIX A: PROOFS OF SUPPORTING CLAIMS FOR THEOREM 1

We now prove the claimed lower bound (3.1a) on \( \mathcal{A}(G) \). This lower bound can be split into the following two claims:

\[
\text{(61a) } \quad \mathcal{A}(G) \geq \frac{1}{n^2} \alpha(G)(\alpha(G) - 1), \quad \text{and}
\]

\[
\text{(61b) } \quad \mathcal{A}(G) \geq \frac{1}{n^2} \beta(G^c).
\]

We use a different argument to establish each claim.

Proof of claim (61a): Recall from Section 1 the definition of the largest independent set. Without loss of generality, let the largest independent set be given by \( I = \{v_1, \ldots, v_\alpha\} \). Assign item \( i \) to vertex \( v_i \) for \( i \in [\alpha] \), where we have used \( \alpha \) as shorthand for \( \alpha(G) \). Now we choose permutations \( \pi \) and \( \pi' \) so that

- \( \pi(i) = i \) for \( i \in [\alpha] \),
- \( \pi'(i) = \alpha - i + 1 \) for \( i \in [\alpha] \),
- \( \pi(i) = \pi'(i) \) for \( i \in \{\alpha + 1, \ldots, n\} \).

Note that last step is possible because \( \pi([\alpha]) = \pi'([\alpha]) \). Moreover, define the matrices \( M = M_{NS}(\pi, 1/2) \) and \( M' = M_{NS}(\pi', 1/2) \). Note that by construc-
tion, we have ensured that $M(G) = M'(G)$. We then have

$$n^2 A(G) \geq \sum_{(i,j) \notin E} (M_{ij} - M'_{ij})^2 = ||M - M'||_F^2 = 2KT(\pi, \pi') = \alpha(\alpha - 1),$$

which completes the proof.

**Proof of claim** (61b): Recall the definition of a maximum biclique from Section 1. Since the complement graph $G^c$ has a biclique with $\beta(G^c)$ edges, the graph $G$ has two disjoint sets of vertices $V_1$ and $V_2$ with $|V_1||V_2| = \beta(G^c)$ that do not have edges connecting one to the other. We now pick the two permutations $\pi$ and $\pi'$ so that

- the permutation $\pi$ ranks items from $V_1$ as the top $|V_1|$ items, and ranks items from $V_2$ as the next $|V_2|$ items;
- the permutation $\pi'$ ranks items from $V_2$ as the top $|V_2|$ items, and ranks items from $V_1$ as the next $|V_1|$ items;
- the permutations $\pi$ and $\pi'$ agree with each other apart from the above constraints.

As before, we define $M = M_{NS}(\pi, 1/2)$ and $M' = M_{NS}(\pi', 1/2)$, and again, we have $M(G) = M'(G)$. The relative orders of items have been interchanged across the biclique, so it holds that $2KT(\pi, \pi') = \beta(G^c)$, which completes the proof.

**APPENDIX B: TWO LEMMAS USED IN AVERAGE-CASE THEOREMS**

We state and prove the following two lemmas, both of which were used to prove multiple results in the average case. The first lemma bounds the performance of the permutation estimator $\hat{\pi}_{ASP}$ for a general SST matrix, and is thus of independent interest.

**Lemma B.1.** For any matrix $M^* \in C_{SST}$, the permutation estimator $\hat{\pi}_{ASP}$ satisfies

$$(62a) \quad ||\hat{\pi}_{ASP}(M^*) - M^*||_F^2 \leq 4(n - 1)||\tau^* - \hat{\tau}||_1,$$

and if additionally, $M^* \in C_{NS}(\lambda^*)$, we have

$$(62b) \quad ||\hat{\pi}_{ASP}(M^*) - M^*||_F^2 \leq 8\lambda^*(n - 1)||\tau^* - \hat{\tau}||_1.$$
In addition, the score estimates satisfy the bounds

\[ \mathbb{E}[\|\tau^* - \hat{\tau}\|_1] \leq c \sum_{v \in V} \frac{1}{\sqrt{d_v}}, \quad \text{and} \]

\[ \Pr \left\{ \|\tau^* - \hat{\tau}\|_1 \geq c \sqrt{\log n} \sum_{v \in V} \frac{1}{\sqrt{d_v}} \right\} \leq n^{-10}. \]

Note that Lemma B.1 implies the bound on permutation recovery (3.2b), since for a matrix \( M^* \in \mathbb{C}_{\text{NS}}(\lambda^*) \), we have \( 8\lambda^2 KT(\hat{\pi}_{\text{ASP}}, \pi^*) = \|\hat{\pi}_{\text{ASP}}(M^*) - M^*\|_F^2 \).

Our second lemma is a type of rearrangement inequality.

**Lemma B.2.** Let \( \{a_u\}_{u=1}^n \) be an increasing sequence of positive numbers and let \( \{b_u\}_{u=1}^n \) be a decreasing sequence of positive numbers. Then we have

\[ \left( \sum_{u=1}^n a_u \right) \left( \sum_{u=1}^n b_u \right) \geq n \sum_{u=1}^n a_u b_u. \]

**B.1. Proof of Lemma B.1.** Assume without loss of generality that \( \pi^* = \text{id} \). We begin by applying Hölder’s inequality to obtain

\[ \|\hat{\pi}_{\text{ASP}}(M^*) - M^*\|_F^2 \leq \|\hat{\pi}_{\text{ASP}}(M^*) - M^*\|_\infty \|\hat{\pi}_{\text{ASP}}(M^*) - M^*\|_1. \]

In the case where \( M^* \in \mathbb{C}_{\text{NS}}(\lambda^*) \), we have \( \|M^*_{\hat{\pi}_{\text{ASP}}(i)} - M^*_i\|_\infty \leq 2\lambda^* \); in the general case \( M^* \in \mathbb{C}_{\text{SS}} \), we have \( \|M^*_{\hat{\pi}_{\text{ASP}}(i)} - M^*_i\|_\infty \leq 1 \). Next, if \( M^*_{\hat{\pi}_{\text{ASP}}} \) denotes the matrix obtained from permuting the rows of \( M^* \) by \( \hat{\pi}_{\text{ASP}} \), then it holds that

\[ \|\hat{\pi}_{\text{ASP}}(M^*) - M^*\|_1 \leq \|\hat{\pi}_{\text{ASP}}(M^*) - M^*_{\hat{\pi}_{\text{ASP}}}\|_1 + \|M^*_{\hat{\pi}_{\text{ASP}}} - M^*\|_1 \]

\[ = 2 \sum_{i=1}^n \|M^*_{\hat{\pi}_{\text{ASP}}(i)} - M^*_i\|_1, \]
where the equality follows from the condition $M^*_{ij} + M^*_{ji} = 1$. We also have
\[
\sum_{i=1}^{n} \| M^*_{\pi_{ASP}(i)} - M^*_{i} \|_1^{(i)} = (n - 1) \sum_{i=1}^{n} | \tau^*_{\pi_{ASP}(i)} - \tau^*_i | \\
= (n - 1) \sum_{i=1}^{n} | \tau^*_i - \tau^*_i | \\
\leq (n - 1) \left[ \sum_{i=1}^{n} | \tau^*_i - \tau^*_{\pi_{ASP}(i)} | + \sum_{i=1}^{n} | \tau^*_{\pi_{ASP}(i)} - \tau^*_i | \right] \\
\leq (n - 1) \left[ \sum_{i=1}^{n} | \tau^*_i - \tau^*_i | + \sum_{i=1}^{n} | \tau^*_i - \tau^*_i | \right] \\
= 2(n - 1)\| \tau^* - \tau^*_i \|_1,
\]
where step (i) is due to monotonicity along each column of $M^*$, and step (ii) follows from the $\ell_1$-rearrangement inequality (see, e.g., Example 2 in the paper [9]), using the fact that both sequences \{\tau^*_i\}_{i=1}^{n} and \{\tau^*_{\pi_{ASP}(i)}\}_{i=1}^{n} are sorted in decreasing order. Combining the last three displays yields the claimed bounds (62a) and (62b).

In order to prove the second part of the lemma, it suffices to show that the random variable $\| \tau^* - \tau^*_i \|_1$ is sub-Gaussian with parameter $cS$, where $S := \sum_{v \in V} 1/\sqrt{d_v}$. Let $\sigma : \{n\} \rightarrow V$ be the uniform random assignment of items to vertices with $\sigma(A) = \emptyset$, and let $D_i$ denote the random degree $d_{\sigma(i)} = \sum_{j \neq i} O_{ij}$ of item $i$. Note that conditioned on the event $\sigma(i) = v$, the difference between a score and its empirical version can be written as
\[
\tau^*_i - \tau^*_i = \left( \frac{1}{d_v} \sum_{j : \sigma(j) = v} M^*_{ij} - \frac{1}{n - 1} \sum_{j \neq i} M^*_{ij} \right) + \frac{1}{d_v} \sum_{j : \sigma(j) = v} W_{ij},
\]
where $\sim$ denotes the presence of an edge between two vertices. The term $\frac{1}{d_v} \sum_{j : \sigma(j) = v} M^*_{ij}$ is the empirical mean of $d_v$ numbers chosen uniformly at random without replacement from the set $\{ M^*_{ij} \}_{j \neq i}$, while $\frac{1}{n - 1} \sum_{j \neq i} M^*_{ij}$ is the true expectation. Moreover, $W_{ij}$ represents independent, zero-mean noise bounded within the interval $[-1, 1]$. Consequently, applying Hoeffding's inequality for sampling without replacement [1, Proposition 1.2] and the standard Hoeffding bound [6] to the two parts respectively, we obtain
\[
Pr \left\{ | \tau^*_i - \tau^*_i | \geq t \mid \sigma(i) = v \right\} \leq 4 \exp(-c d_v t^2).
\]
Replacing $t$ by $t/\sqrt{d_v}$, we see that conditioned on the event $\sigma(i) = v$, the random variable $\sqrt{d_v} | \tau^*_i - \tau^*_i |$ is sub-Gaussian with a constant parameter $c'$,
or equivalently,

\[ \mathbb{E} \left[ \exp \left( t \sqrt{D_i} | \hat{\tau}_i - \tau^*_i | \right) \mid \sigma(i) = v \right] \leq \exp( ct^2 ). \]  

(64)

Since \( S = \sum_{i=1}^{n} 1/\sqrt{D_i} \), Jensen’s inequality implies that

\[ \mathbb{E} \left[ \exp \left( t \sum_{i=1}^{n} | \hat{\tau}_i - \tau^*_i | \right) \right] \]

\[ \leq \mathbb{E} \left[ \sum_{i=1}^{n} \frac{1}{\sqrt{S}} \exp \left( t S \sqrt{D_i} | \hat{\tau}_i - \tau^*_i | \right) \right] \]

\[ = \sum_{i=1}^{n} \frac{1}{S} \sum_{v \in V} \Pr \{ \sigma(i) = v \} \mathbb{E} \left[ \frac{1}{\sqrt{D_i}} \exp \left( t S \sqrt{D_i} | \hat{\tau}_i - \tau^*_i | \right) \mid \sigma(i) = v \right] \]

\[ \leq \sum_{i=1}^{n} \frac{1}{S} \sum_{v \in V} \frac{1}{n \sqrt{d_v}} \exp(cs^2 t^2) \]

\[ = \exp(cs^2 t^2), \]

where the last inequality follows from equation (64). Therefore, the random variable \( || \hat{\tau} - \tau^* ||_1 \) is sub-Gaussian with parameter \( cs \), as claimed. \( \square \)

**B.2. Proof of Lemma B.2.** For any increasing sequence \( \{a_u\} \) and decreasing sequence \( \{b_u\} \), the rearrangement inequality (see, e.g., Example 2 in the paper [9]) guarantees that

\[ \sum_{u=1}^{n} a_u b_u \leq \sum_{u=1}^{n} a_u b_{\pi(u)} \quad \text{for any permutation } \pi. \]

This inequality implies that

\[ \frac{1}{n} (\sum_{u=1}^{n} a_u)(\sum_{u=1}^{n} b_u) = \frac{1}{n} \sum_{v=1}^{n} \sum_{u=1}^{n} a_u b_{\pi(v)(u)} \geq \frac{1}{n} \sum_{v=1}^{n} \sum_{u=1}^{n} a_u b_u = \sum_{u=1}^{n} a_u b_u, \]

where \( \pi(v)(u) := (u + v) \mod n \) and we have used the rearrangement inequality for each of these permutations. \( \square \)

**APPENDIX C: PROOFS OF SUPPORTING LEMMAS FOR THEOREM 2**

In this section, we provide the proof of Lemma 1, which was stated in the proof of Theorem 2.
C.1. Proof of Lemma 1. We fix \( i, j \in [n] \) with \( i < j \) and condition on the event that \( \sigma(i) = u \) and \( \sigma(j) = v \) throughout the proof. First, note that the bound stated is trivially true if one of the vertices \( u \) or \( v \) has degree 1, by adjusting the constant appropriately. Hence, we assume for the rest of the proof that \( d_u, d_v \geq 2 \). Define the quantity

\[
\bar{\Delta}_{ji} = 2\lambda^* \frac{j-i-1}{n-2}.
\]

We divide the rest of our analysis into two cases.

Case 1, \((u, v) \notin E(G)\). When the vertices \( u \) and \( v \) are not connected, we have

\[
\begin{align*}
\bar{\tau}_j &:= \mathbb{E}[\hat{\tau}_j] = \frac{1}{2} + \lambda^* \left( \frac{n-j-j-2}{n-2} \right) \\
\bar{\tau}_i &:= \mathbb{E}[\hat{\tau}_i] = \frac{1}{2} + \lambda^* \left( \frac{n-i-1-i-2}{n-2} \right),
\end{align*}
\]

and it can be verified that \( \bar{\tau}_i - \bar{\tau}_j = \bar{\Delta}_{ji} \). Consequently, we have

\[
\Pr\{\hat{\pi}_{\text{ASP}}(j) < \hat{\pi}_{\text{ASP}}(i) \mid \sigma(i) = u, \sigma(j) = v\} = \Pr\{\hat{\tau}_j > \hat{\tau}_i \mid \sigma(i) = u, \sigma(j) = v\} \leq \Pr\left\{ |\hat{\tau}_j - \hat{\tau}_i| > \frac{\sqrt{d_u}}{\sqrt{d_u} + \sqrt{d_v}} \bar{\Delta}_{ji} \mid \sigma(i) = u, \sigma(j) = v\right\} + \Pr\left\{ |\hat{\tau}_i - \hat{\tau}_i| > \frac{\sqrt{d_v}}{\sqrt{d_u} + \sqrt{d_v}} \bar{\Delta}_{ji} \mid \sigma(i) = u, \sigma(j) = v\right\}
\]

\[
\leq 4 \exp\left\{-c \frac{d_u d_v}{(\sqrt{d_u} + \sqrt{d_v})^2} \bar{\Delta}^2_{ji}\right\},
\]

where the last step follows from the Hoeffding bound for sampling without replacement in conjunction with the standard Hoeffding bound for bounded independent noise, by an argument similar to that of equation (63).

Case 2, \((u, v) \in E(G)\). When the vertices \( u \) and \( v \) are connected, we have

\[
\begin{align*}
\bar{\tau}_j &:= \mathbb{E}[\hat{\tau}_j] = \frac{1}{2} + \frac{d_v-1}{d_v} \lambda^* \left( \frac{n-j-j-2}{n-2} \right) - \frac{1}{d_v} \lambda^* \\
\bar{\tau}_i &:= \mathbb{E}[\hat{\tau}_i] = \frac{1}{2} + \frac{d_u-1}{d_u} \lambda^* \left( \frac{n-i-1-i-2}{n-2} \right) + \frac{1}{d_u} \lambda^*,
\end{align*}
\]

and it can be verified that \( \bar{\tau}_i - \bar{\tau}_j \geq \bar{\Delta}_{ji} \).
Now, however, we must apply the Hoeffding bound for sampling without replacement to $d_u - 1$ and $d_v - 1$ random variables, respectively. Recalling that $d_u, d_v \geq 2$, we have

$$\Pr \left\{ \tilde{\pi}_{\text{ASP}}(j) < \tilde{\pi}_{\text{ASP}}(i) \mid \sigma(i) = u, \sigma(j) = v \right\}$$

$$= \Pr \left\{ \tilde{\tau}_j > \tilde{\tau}_i \mid \sigma(i) = u, \sigma(j) = v \right\}$$

$$\leq \Pr \left\{ \frac{|\tilde{\tau}_j - \tilde{\tau}_j|}{\sqrt{d_u} + \sqrt{d_u}} \tilde{\Delta}_{ji} \mid \sigma(i) = u, \sigma(j) = v \right\}$$

$$+ \Pr \left\{ \frac{|\tilde{\tau}_i - \tilde{\tau}_i|}{\sqrt{d_v} + \sqrt{d_v}} \tilde{\Delta}_{ji} \mid \sigma(i) = u, \sigma(j) = v \right\}$$

$$\leq 4 \exp \left\{ -c \frac{(d_u - 1)(d_v - 1)}{(\sqrt{d_u} - 1 + \sqrt{d_v} - 1)^2} \tilde{\Delta}_{ji}^2 \right\}$$

(68)

$$\leq 4 \exp \left\{ -c' \frac{d_ud_v}{(\sqrt{d_u} + \sqrt{d_v})^2} \tilde{\Delta}_{ji}^2 \right\}.$$  

We use the shorthand $L_{uv}$ to denote the LHS of equation (5.2). Having established the bounds (67) and (68), we now combine them to derive that

$$L_{uv} \leq \frac{1}{n(n-1)} \sum_{j=2}^{n-1} \sum_{i<j} 4 \exp \left\{ -c \frac{d_ud_v}{(\sqrt{d_u} + \sqrt{d_v})^2} (j-i-1)^2 \frac{(\lambda^*)^2}{(n-2)^2} \right\}$$

$$\leq \frac{4}{n(n-1)(n-1)} \sum_{m=1}^{n} \exp \left\{ - \frac{d_ud_v}{(\sqrt{d_u} + \sqrt{d_v})^2} m^2 \frac{(\lambda^*)^2}{(n-2)^2} \right\},$$

where we have used $m = j - i$, and noted that there are at most $n - 1$ repetitions of each distinct value of $j - i$ in the sum over $j > i$.

Defining $\psi(q) = \sum_{m=1}^{\infty} q^m$, we recall the following theta function identity\(^1\) for $ab = \pi$ (see, for instance, equation (2.3) in Yi [10]):

$$\sqrt{a} \left( 1 + 2\psi(e^{-a^2}) \right) = \sqrt{b} \left( 1 + 2\psi(e^{-b^2}) \right).$$

\(^1\)For the rest of this subsection, $\pi$ denotes the universal constant.
Using the identity by setting \( a^2 = c \frac{d_ud_v}{\sqrt{d_u} + \sqrt{d_v}} \frac{(\lambda^*)^2}{n^2} \) yields

\[
L_{uv} \leq \frac{c}{n} \frac{n \sqrt{d_u} + \sqrt{d_v}}{\sqrt{d_u}d_v} \left( 1 + 2 \sum_{m=1}^{\infty} \exp \left\{ -\pi^2 \left( \frac{\sqrt{d_u} + \sqrt{d_v}}{d_u d_v} \right)^2 m^2 \frac{n^2}{(\lambda^*)^2} \right\} \right)
\]

\[
\leq \frac{c}{\lambda^*} \frac{\sqrt{d_u} + \sqrt{d_v}}{\sqrt{d_ud_v}} \left( 1 + 2 \sum_{m=1}^{\infty} \exp \left\{ -\pi^2 \left( \frac{\sqrt{d_u} + \sqrt{d_v}}{d_ud_v} \right)^2 \frac{n^2}{(\lambda^*)^2} \right\} \right)
\]

(69) \quad \leq \frac{c}{\lambda^*} \frac{\sqrt{d_u} + \sqrt{d_v}}{\sqrt{d_ud_v}} \left( 1 + \sum_{m=1}^{\infty} \exp \left\{ -16\pi^2 nm \right\} \right),

where in the last step, we have used the fact that \( \lambda^* \leq 1/2 \), and that \( \frac{(\sqrt{d_u} + \sqrt{d_v})^2}{d_u d_v} \geq 4/n \). Bounding the geometric sum by a universal constant yields the required result. \( \square \)

**APPENDIX D: PROOFS OF SUPPORTING LEMMAS FOR THEOREM 3**

In this section, we prove Lemma 3, which was used in the proof of part (a) of Theorem 3 in the main text. We also provide the proof of part (b).

**D.1. Theorem 3, part (a).**

D.1.1. **Proof of Lemma 3.** Fix a graph \( G \) with degree sequence \( \{d_v\}_{v \in V} \), and introduce the shorthand \( S = \sum_{v \in V} 1/\sqrt{d_v} \). For some parameter \( k \) to be chosen, define the graph \( G' \) on the same vertex set to be the disjoint union of one clique of size \( c_1 \lfloor \sqrt{|E|} \rfloor \), \( c_2 k \) cliques of size \( \lfloor n/k \rfloor \) and \( c_3 S \) cliques of size 2, where \( c_1, c_2 \) and \( c_3 \) are constants to be determined such that the sizes of each clique are integers. The number of vertices remains the same, so that

(70) \quad n = c_1 \lfloor \sqrt{|E|} \rfloor + c_2 k \lfloor n/k \rfloor + 2c_3 S.

The number of edges of \( G' \) is

\[
|E'| = \left( \frac{c_1 \lfloor \sqrt{|E|} \rfloor}{2} \right) + c_2 k \left( \frac{\lfloor n/k \rfloor}{2} \right) + c_3 S \asymp |E| + \frac{n^2}{k},
\]

where the last approximation holds because \( S \leq n \leq 2|E| \). Moreover, let

\[
S' = \sum_{v \in V} \frac{1}{\sqrt{d'_v}} = \frac{c_1 \lfloor \sqrt{|E|} \rfloor}{\sqrt{c_1 \lfloor \sqrt{|E|} \rfloor} - 1} + \frac{c_2 k}{\sqrt{\lfloor n/k \rfloor} - 1} + c_3 S \asymp \sqrt{nk} + S,
\]
where the last approximation holds since $|E|^{1/4} \leq \sqrt{n} \leq S$.

In order to guarantee that $|E'| \asymp |E|$ and $S' \asymp S$, we need to choose an integer $k$ so that $n^2/k \leq c|E|$ and $\sqrt{nk} \leq cS$, or equivalently

$$\frac{n^2}{c|E|} \leq k \leq \frac{c^2S^2}{n}.$$  

Such an integer $k$ exists if $|E|S^2 \geq n^3$. Indeed, applying Lemma B.2 twice (with $a_u = d_u(u)$ and $b_u = 1/\sqrt{d_u(u)}$ the first time and $a_u = \sqrt{d_u(u)}$ and $b_u = 1/\sqrt{d_u(u)}$ the second time, where $\{d_u(u)\}_{u=1}^n$ is the degree sequence in ascending order), we obtain that

$$|E|S^2 = \left(\sum_{v \in V} d_v\right) \left(\sum_{v \in V} \frac{1}{\sqrt{d_v}}\right)^2 \geq n\left(\sum_{v \in V} \sqrt{d_v}\right) \left(\sum_{v \in V} \frac{1}{\sqrt{d_v}}\right) \geq n^3.$$  

With $k$ selected, it is easy to choose $c_1, c_2$ and $c_3$ so that inequality (70) holds, since each of $\sqrt{|E|}, k[n/k]$ and $S$ is no larger than $n$. The issue of integrality can be taken care of by constant-order adjustment of these numbers, so the proof is complete. \hfill \Box

**D.1.2. Proof of Lemma 4.** For $\ell \in [n_j]$ and $\ell \leq s \leq n - n_j + \ell$, define a bivariate function

$$p(\ell, s) := \binom{n_j}{\ell - 1} \binom{n - n_j}{s - \ell} \binom{n}{s - 1}^{-1}.$$  

Note that for any fixed $s$, the function $\ell \mapsto p(\ell, s)$ is the probability mass function of the hypergeometric distribution that describes the probability of $\ell - 1$ successes in $s - 1$ draws without replacement from a population of size $n$ with $n_j$ successes. Hence, its maximum is attained at $\ell = \left\lfloor s \frac{n_j}{n} + 1 \right\rfloor$. Now we consider the index set

$$\mathcal{I} = \left\{ (\ell, s) : \left[\frac{n_j}{3} \right] \leq \ell \leq \left[\frac{2n_j}{3}\right], \left[\frac{n_j}{3}\right] \leq \frac{s n_j + 1}{n + 2} \leq \left[\frac{2n_j}{3}\right] \right\},$$

and notice that $\mathcal{I} \subset \left[\frac{n_j}{3}, \frac{2n_j}{3}\right] \times \left[\frac{n}{5}, \frac{4n}{5}\right]$. The range of interest $[2n_j/5] \leq \ell \leq [3n_j/5]$ and $|s - \mu| \leq n/\sqrt{n_j}$, is contained within the set $\mathcal{I}$, since $\mu = \ell \frac{n+1}{n_j+1}$. Moreover, inequality (5.8) ensures that $\Pr\{\pi(i) = s\} \leq p(\ell, s)\frac{c n_j}{n}$ for $(\ell, s) \in \mathcal{I}$. Thus, in order to complete the proof, it suffices to prove that $p(\ell, s) \leq c/\sqrt{n_j}$ for $(\ell, s) \in \mathcal{I}$, and it suffices to consider $(\ell, s)$ such that $\ell = \left\lfloor s \frac{n_j}{n+2} + 1 \right\rfloor$ since each function $\ell \mapsto p(\ell, s)$ attains its maximum at such a pair $(\ell, s)$. 

Toward this end, we use Stirling’s approximation [4] to obtain
\begin{equation}
 p(\ell, s) \leq c_2 \frac{\sqrt{n_j(n-n_j)(s-1)(n-s+1)}}{\sqrt{\ell-1}(n_j-\ell+1)(s-\ell)(n-n_j-s+\ell)n}
\end{equation}
\begin{equation}
 = \frac{n_j^{n_j}(n-n_j)^{n-n_j}(s-1)^{s-1}(n-s+1)^{n-s+1}}{(\ell-1)^{\ell-1}(n_j-\ell+1)^{n_j-\ell+1}(s-\ell)^{s-\ell}(n-n_j-s+\ell)^{n-n_j-s+\ell}n^n}.
\end{equation}

Since the factor in line (71) scales as \(1/\sqrt{n_j}\) for \((\ell, s) \in \mathcal{I}\), it remains to bound the factor in line (72) by a universal constant. This follows from lengthy yet standard approximations which we briefly describe here. Assume that \(s = \ell \frac{n+2}{n+1}\); the extension to the general case is easy. We first group together
\[
\left[\frac{n_j(s-1)}{(\ell-1)n}\right]^{\ell-1} = \left[\frac{n_j(n\ell+n-2n-1)}{(\ell-1)n}\right]^{\ell-1} = \left[1 + \frac{1 + \left(2n_j - n_j - \ell n\right)}{(n_j n + n)}\right]^{\ell-1},
\]
which is bounded by a constant for \((\ell, s) \in \mathcal{I}\) considering that \(\lim_{m \to \infty}(1 + \frac{a}{m})^m = e^a\). Then, we group together the terms
\[
\left[\frac{n_j(n-s+1)}{(n_j-\ell+1)n}\right]^{n_j-\ell+1}, \left[\frac{(n-n_j)(s-1)}{(s-\ell)n}\right]^{s-\ell}, \left[\frac{(n-n_j)(n-s+1)}{(n-n_j-s+\ell)n}\right]^{n-n_j-s+\ell},
\]
respectively, and a similar argument yields that each term is bounded by a constant.

D.2. Proof of Theorem 3, part (b). Given a parameter space \(\Theta\), a set \(\mathcal{P} = \{\theta_1, \theta_2, \ldots, \theta_{|\mathcal{P}|}\}\) is said to be a \(\delta\)-packing in the metric \(\rho\) if \(\rho(\theta_i, \theta_j) > \delta\) for all \(i \neq j\). The lower bound of part (b) is based on the following packing lemma for the set of permutations in Kendall’s tau distance. We note that a similar lemma was proved by Barg and Mazumdar [2].

**Lemma D.1.** For some positive constant \(c_1\), there exists an \(c_1 n^2\)-packing \(\mathcal{P}\) of the set of permutations in the Kendall’s tau distance such that \(\log |\mathcal{P}| \geq n\).

Consider the random observation model with graph \(G = (V, E)\), where \(E\) denotes the random edge set of observations. We denote by \(Q_M\) the law of the random observation noisy sorting model with underlying matrix \(M = M_{NS}(\pi, \lambda)\). We require the following lemma.
**Lemma D.2.** Let $P_{M,G}$ denote the law of the noisy sorting model with underlying matrix $M \in \mathbb{C}_{NS}(\lambda)$ for $\lambda \in [0,1/4]$ and comparison graph $G$. Suppose that the entries of two matrices $M, M' \in \mathbb{C}_{NS}(\lambda)$ differ in $s$ edges of the graph $G$. Then the KL divergence is bounded as

$$ \text{KL}(P_{M,G}, P_{M',G}) \leq 9\lambda^2 s. \quad (73) $$

Note that conditional on any instance of $E$, Lemma D.2 guarantees that

$$ \text{KL}(P_{M,G}, P_{M',G}) \leq 9\lambda^2 \left| \{(i,j) \in E : i < j, M_{i,j} \neq M'_{i,j} \} \right|, $$

where $P_{M,G}$ denotes the model for fixed graph $G$. Hence taking expectation over the random edge set yields the upper bound

$$ \text{KL}(Q_M, Q_{M'}) \leq 9\lambda^2 \sum_{i<j} \frac{2|E|}{n(n-1)} = 9\lambda^2 |E|, $$

valid for any $M, M' \in \mathbb{C}_{NS}(\lambda)$.

Note that $\|M - M'\|_{F}^2 = 8\lambda^2 \text{KT}(\pi, \pi')$ for $M = M_{NS}(\pi, \lambda)$ and $M' = M_{NS}(\pi', \lambda)$. Hence Fano's inequality applied to the packing given by Lemma D.1 yields that

$$ \inf_{\hat{M}} \sup_{M^* \in \mathbb{C}_{NS}} \mathbb{E} \left[ \|\hat{M} - M^*\|_{F}^2 \right] \geq 8\lambda^2 c_1 n^2 \left( 1 - \frac{9\lambda^2 |E| + \log 2}{n} \right). $$

The proof is completed by choosing $\lambda^2 = c_2 n/|E|$ for a sufficiently small constant $c_2$. \hfill \square

It remains to prove Lemmas D.1 and D.2.

**D.2.1. Proof of Lemma D.1.** The inversion table $b = (b_1, \ldots, b_n)$ of a permutation $\pi$ has entries defined by

$$ b_i = \sum_{j=i+1}^{n} \mathbf{1}\{\pi(i) > \pi(j)\} \text{ for each } i \in [n]. $$

We refer the reader to Mahmoud [7] and references therein for background on inversion tables. By definition, we have $b_i \in \{0, 1, \ldots, n-i\}$ and $\text{KT}(\pi, \text{id}) = \sum_{i=1}^{n} b_i$ where $\text{id}$ denotes the identity permutation. In fact, the set of tables $b$ satisfying $b_i \in \{0, 1, \ldots, n-i\}$ is bijective to the set of permutations via this relation [7]. This bijection aids in counting permutations with constraints.

Denote by $B(\text{id}, r)$ the set of permutations that are within Kendall’s tau distance $r$ of the identity $\text{id}$. We seek an upper bound on $|B(\text{id}, r)|$. Every
\( \pi \in B(\text{id}, r) \) corresponds to an inversion table \( b \) such that \( \sum_{i=1}^{n} b_i \leq r \). If \( b_i \) is only required to be a nonnegative integer, then the number of \( b \) satisfying \( \sum_{i=1}^{n} b_i \leq r \) is bounded by \( \binom{n+r}{n} \). After taking logarithms, this yields a bound

\[
\log |B(\text{id}, r)| \leq n \log(1 + r/n) + n.
\]

Let \( \mathcal{P} \) be a maximal \( c_1 n^2 \)-packing of the set of permutations, which is necessarily also a \( c_1 n^2 \)-covering of that set. Then the family \( \{B(\pi, c_1 n^2)\}_{\pi \in \mathcal{P}} \) covers all permutations. By the right-invariance of the Kendall's tau distance under composition, the above bound yields \( \log |B(\pi, c_1 n^2)| \leq n \log(1+c_1 n) + n \) for each \( \pi \). Since there are \( n! \) permutations in total, we conclude that \( \log |\mathcal{P}| \geq \log(n!) - n \log(1 + c_1 n) - n \geq n \) for a sufficiently small constant \( c_1 \).

D.2.2. Proof of Lemma D.2. The KL divergence between Bernoulli observations has the form

\[
\begin{align*}
\text{KL}(\text{Ber}(1/2 + \lambda), \text{Ber}(1/2 - \lambda)) &= \text{KL}(\text{Ber}(1/2 - \lambda), \text{Ber}(1/2 + \lambda)) \\
&= (1/2 + \lambda) \log \frac{1/2 + \lambda}{1/2 - \lambda} + (1/2 - \lambda) \log \frac{1/2 - \lambda}{1/2 + \lambda} \\
&= 2 \lambda \log \frac{1/2 + \lambda}{1/2 - \lambda} \\
&\leq 9 \lambda^2 \quad \text{for all } \lambda \in [0, 1/4],
\end{align*}
\]

where the last inequality follows by some simple algebra.

Note that the KL divergence between a pair of product distributions is equal to the sum of the KL divergences between individual pairs. Since \( M \) and \( M' \) differ in \( s \) entries on the graph \( G \) and the Bernoulli observations are independent for different edges, we see that \( \text{KL}(P_{M,G}, P_{M',G}) \leq 9 \lambda^2 s \). \( \square \)

APPENDIX E: PROOFS OF SUPPORTING LEMMAS FOR THEOREM 4

In this section, we provide proofs of Lemmas 5, 6, and 7, which were used in the proof of Theorem 4.

E.1. Proof of Lemma 5. Our proof relies crucially on the fact that one of the two sets is a block.
For a fixed integer \( k \), we condition on the event \(|B \cap E_2| = k\). Note that
\[
E_2 = \pi(E) = \{(i, j) : (\pi(i), \pi(j)) \in E\},
\]
where \( \pi \) is a uniform random permutation, and \( E \) is a fixed instance of \( E_2 \).

For any pair of tuples \((i, j), (k, \ell)\) \( \in B \), consider the permutation \( \tilde{\pi} \) defined by

- \( \tilde{\pi}(i) = k \), \( \tilde{\pi}(k) = i \), \( \tilde{\pi}(j) = \ell \) and \( \tilde{\pi}(\ell) = j \);
- \( \tilde{\pi}(m) = m \) for \( m \neq i, j, k \) or \( \ell \).

Note that right-composition by \( \tilde{\pi} \) is clearly a bijection between the sets
\[
\{\pi : (i, j) \in \pi(E)\} \quad \text{and} \quad \{\pi : (k, \ell) \in \pi(E)\}.
\]
Therefore, we have
\[
|\{\pi : (i, j) \in E_2\}| = |\{\pi : (k, \ell) \in E_2\}|.
\]
A counting argument then completes the proof. Indeed, conditioned on the event \(|B \cap E_2| = k\), we have
\[
\sum_{(i, j) \in B} \Pr\{(i, j) \in E_2\} = \mathbb{E}\left[ \sum_{(i, j) \in B} \mathbf{1}\{(i, j) \in E_2\} \right] = k,
\]
which implies that \( \Pr\{(i, j) \in E_2\} = \frac{k}{|B|} \).

\[\textbf{E.2. Proof of Lemma 6.}\]

Fix an individual block \( B \) of dimensions \( h \times w \), and let \( E = E_2 \) for notational convenience. Define the random variable \( Y = |B \cap E| + 1 \) so that \((|B \cap E| + 1)^{-1} \leq 2/Y\). Hence we require a bound on the quantity \( \mathbb{E}[Y^{-1}] \). Toward this end, we write
\[
Y = 1 + \sum_{(i, j) \in B} \mathbf{1}\{(i, j) \in E\}, \quad \text{and}
\]
\[
Y^2 = 1 + 2 \sum_{(i, j) \in B} \mathbf{1}\{(i, j) \in E\} + \sum_{(i, j, (i', j')) \in B} \mathbf{1}\{(i, j), (i', j') \in E\}.
\]

Note that for \( (i, j), (i', j') \in B \) where \( i \neq i' \) and \( j \neq j' \), we have
\[
\Pr\{(i, j) \in E\} = \frac{2|E|}{n(n - 1)},
\]
\[
\Pr\{(i, j), (i', j') \in E\} = \frac{\sum_{v \in V} d_v(d_v - 1)}{n(n - 1)(n - 2)}, \quad \text{and}
\]
\[
\Pr\{(i, j), (i', j') \in E\} = \frac{4|E|^2 - 2 \sum_{v \in V} d_v(d_v - 1) - 2|E|}{n(n - 1)(n - 2)(n - 3)}.
\]
Hence, we can compute the first two moments of \( Y \) as

\[
\mathbb{E}[Y] = 1 + \sum_{(i,j) \in B} \mathbb{P}\{(i,j) \in E\} = 1 + \frac{2hw|E|}{n(n-1)}, \quad \text{and}
\]

\[
\mathbb{E}[Y^2] = 1 + 2 \sum_{(i,j) \in B} \mathbb{P}\{(i,j) \in E\} + \sum_{(i,j),(i',j') \in B} \mathbb{P}\{(i,j),(i',j') \in E\}
= 1 + \frac{4hw|E|}{n(n-1)} + \frac{2hw|E|}{n(n-1)}
+ \left[ hw(w-1) + wh(h-1) \right] \frac{\sum_{v \in V} d_v(d_v - 1)}{n(n-1)(n-2)}
+ h(h-1)w(w-1) \frac{4|E|^2 - 2 \sum_{v \in V} d_v(d_v - 1) - 2|E|}{n(n-1)(n-2)(n-3)},
\]

where for the last step we split into cases according to whether \( i = i' \) or \( j = j' \). Therefore, the variance \( \text{var}(Y) \) is equal to

\[
\mathbb{E}[Y^2] - \mathbb{E}[Y]^2
= \frac{2hw|E|}{n(n-1)} + \left[ hw(w-1) + wh(h-1) \right] \frac{\sum_{v \in V} d_v(d_v - 1)}{n(n-1)(n-2)}
+ h(h-1)w(w-1) \frac{4|E|^2 - 2 \sum_{v \in V} d_v(d_v - 1) - 2|E|}{n(n-1)(n-2)(n-3)} - \frac{4h^2w^2|E|^2}{n^2(n-1)^2}.
\]

We note that

\[
\frac{h(h-1)w(w-1)}{n(n-1)(n-2)(n-3)} = \frac{h^2w^2}{n^2(n-1)^2}
\leq \frac{2h^2w^2}{n^2(n-1)^2(n-2)(n-3)},
\]

where in the last step, we have used the fact that the quantity above is maximized when \( h = w \), and that \( 2 \leq h + w \leq n \) by the construction of the blocks.

Combining the pieces, we conclude that \( \text{var}(Y) \) is bounded by

\[
\frac{c hw|E|}{n^2} + c(hw^2 + wh^2) \frac{\sum_{v \in V} d_v^2}{n^3} + c \frac{h^2w^2|E|^2}{n^6}
\leq 2c \frac{hw|E|}{n^2} + c(hw^2 + wh^2) \frac{\sum_{v \in V} d_v^2}{n^3}.
\]
where the inequality holds because $h \leq n$, $w \leq n$ and $|E| \leq n^2$. Using the fact that $Y \geq 1$ and applying Chebyshev's inequality, we obtain

$$
\mathbb{E}[Y^{-1}] \leq \Pr \left\{ Y \leq \frac{\mathbb{E}[Y]}{2} \right\} + \frac{2}{\mathbb{E}[Y]} \leq \frac{4}{\mathbb{E}[Y]^2} \text{var}(Y) + \frac{2}{\mathbb{E}[Y]}
$$

$$
\leq c \frac{n^4}{h^2w^2|E|^2} \left[ \frac{hw|E|}{n^2} + (hw^2 + wh^2) \frac{\sum_{v \in V} d_v^2}{n^3} \right] + c \frac{n^2}{hw|E|}
$$

$$
= 2c \frac{n^2}{hw|E|} + cn \frac{h + w}{hw} \sum_{v \in V} d_v^2.
$$

Now the above bound yields

$$
\mathbb{E}\left| \frac{B}{Y} \right| \leq 2c \frac{n^2}{|E|} + cn (h + w) \sum_{v \in V} d_v^2.
$$

Note that there are at most $m^2 = (n/S)^2$ blocks in total and the sum of $h$ over $m - 1$ off-diagonal blocks vertically is bounded by $n$ (similarly for $w$). Thus we conclude that

$$
\mathbb{E} \sum_{B \in B(b)} \frac{|B|}{|B \cap E| \vee 1} \leq c \frac{m^2n^2}{|E|} + c mn^2 \sum_{v \in V} d_v^2.
$$

In order to complete the proof, it suffices to show that

$$
\frac{n^2}{|E|} \left( \sum_{v \in V} \frac{1}{\sqrt{d_v}} \right)^{-2} + n \left( \sum_{v \in V} \frac{1}{\sqrt{d_v}} \right)^{-1} \sum_{v \in V} d_v^2 \leq c \frac{n}{n \sum_{v \in V} \frac{1}{\sqrt{d_v}}}.
$$

Note that Lemma B.2 implies that

$$
2|E| \left( \sum_{v \in V} \frac{1}{\sqrt{d_v}} \right)^2 = \left( \sum_{v \in V} d_v \right) \left( \sum_{v \in V} \frac{1}{\sqrt{d_v}} \right)^2 \geq n^3.
$$

It follows that

$$
\frac{n^2}{|E|} \left( \sum_{v \in V} \frac{1}{\sqrt{d_v}} \right)^{-2} \leq \frac{2}{n} \leq \frac{2}{n} \sum_{v \in V} \frac{1}{\sqrt{d_v}},
$$

and that

$$
\frac{n}{n \sum_{v \in V} \frac{1}{\sqrt{d_v}}} \sum_{v \in V} d_v^2 \leq \frac{4}{n^2} \sum_{v \in V} d_v^2 \left( \sum_{v \in V} \frac{1}{\sqrt{d_v}} \right) \leq \frac{4}{n} \sum_{v \in V} \frac{1}{\sqrt{d_v}}
$$

since $d_v \leq n$. 

\[\square\]
E.3. Proof of Lemma 7. This lemma is a generalization of an approximation theorem due to Chatterjee [3] and Shah et al. [8] to the noisy and two-dimensional setting.

We use the shorthand $\mathbf{b} = \mathbf{b}_t(\hat{\tau})$ for the rest of the proof. Also define the set of placeholder elements in the partition $\mathbf{b}$ as

$$s(\mathbf{b}) = \{ i : i \text{ is smallest index in some set } I \in \mathbf{b} \}.$$ 

We are now ready to prove the lemma. Begin by writing

$$\| X - R(X, \mathbf{b}) \|^2_F = \sum_{k=1}^{n} \left\| X_k - \frac{1}{|\mathbf{b}(k)|} \sum_{j \in \mathbf{b}(k)} X_j \right\|^2_F \leq \sum_{k=1}^{n} \left\| X_k - \frac{1}{|\mathbf{b}(k)|} \sum_{j \in \mathbf{b}(k)} X_j \right\|_1 \leq \sum_{k=1}^{n} \frac{1}{|\mathbf{b}(k)|} \sum_{j \in \mathbf{b}(k)} \| X_k - X_j \|_1,$$

where step (i) follows from the fact that each entry of the difference matrix $X - R(X, \mathbf{b})$ is bounded in the interval $[-1, 1]$; step (ii) follows from Jensen’s inequality and convexity of the $\ell_1$ norm. We can now use the fact that for fixed $k$ and $j$, the quantity $X_{k\ell} - X_{j\ell}$ has the same sign for all $\ell \in [n]$ due to the monotonicity of columns of the matrix $X$, to write

$$\| X - R(X, \mathbf{b}) \|^2_F \leq \sum_{k=1}^{n} \frac{n}{|\mathbf{b}(k)|} \sum_{j \in \mathbf{b}(k)} |\tau(X)_k - \tau(X)_j| \leq \sum_{k \in s(\mathbf{b})} \frac{n}{|\mathbf{b}(k)|} \sum_{i,j \in \mathbf{b}(k)} (|\hat{\tau}_i - \tau(X)_i| + |\hat{\tau}_j - \tau(X)_j| + |\hat{\tau}_i - \hat{\tau}_j|) \leq \sum_{k \in s(\mathbf{b})} \frac{n}{|\mathbf{b}(k)|} \sum_{i,j \in \mathbf{b}(k)} t |\hat{\tau} - \tau(X)|_1 + nt.$$ 

Step (iii) uses the property of the blocking partition $\mathbf{b}$, which ensures that $|\hat{\tau}_i - \hat{\tau}_j| \leq t$ when $i$ and $j$ lie in the same block. This completes the proof.
APPENDIX F: BOUNDS ON THE MINIMAX DENOISING ERROR

As we saw in Theorem 1, the minimax risk of Frobenius norm estimation is prohibitively large for many comparison topologies. In some applications, however, it may be of interest to control the denoising error, which is the error we make on the observations seen on the edges of the graph. Accordingly, we define the quantity

$$\mathcal{E}(G, \mathcal{C}) = \inf_{\hat{M} = f(Y(G))} \sup_{M^* \in \mathcal{C}} \mathbb{E}\left[ \frac{1}{|E|} \| \hat{M} - M^* \|_2^2 \right],$$

where we have used a normalization of $|E|$ to provide an average entry-wise bound on the denoising error. The following theorem provides bounds on the minimax denoising error for fixed topologies.

To facilitate stating the lower bound, we let $L(G)$ denote the length of the longest path in the graph $G$. Moreover, we call a subgraph $S$ a star in the graph $G$ if a central vertex of $S$ is connected to all the remaining vertices, and no edge exists between vertices other than the center. Let $N(G)$ denote the number of edges in the maximum star in the graph $G$.

**Theorem F.1.** For any connected graph $G$, we have

$$\mathcal{E}(G, \mathcal{C}_{NS}) \gtrsim \frac{L(G) \vee N(G)}{|E|} \quad \text{and} \quad \mathcal{E}(G, \mathcal{C}_{SST}) \lesssim \frac{n \log^2 n}{|E|}.$$  \hspace{1cm} (74)

Again, the lower bound on the error of the noisy sorting class provides a lower bound for the SST class. Conversely, the upper bound on the error for the SST class upper bounds the error for the noisy sorting class. For many graphs used in practice, the above bounds can be evaluated to provide a tight characterization of the denoising error up to logarithmic factors.

The upper bound is obtained by the least squares estimator

$$\hat{M}_{LS} = \arg \min_{\hat{M} \in \mathcal{C}_{SST}} \|Y - M^*\|_E^2.$$  

While we do not know yet whether such an estimator is computable in polynomial time, analyzing it provides a notion of the fundamental limits of the problem. In particular, it is clear that the denoising problem is easier than Frobenius norm estimation, and we obtain consistent rates provided that the number of edges in the graph satisfies $|E| = \omega(n \log^2 n)$.

**F.1. Proof of Theorem F.1.** In this section, we prove Theorem F.1 on the denoising error rate of the problem, splitting it into proofs of the lower and upper bounds.
F.1.1. Proof of lower bound. We consider the two terms $L(G)$ and $D(G)$ in the lower bound separately. In each case, we construct a suitable local packing $\mathcal{P}$ of the parameter space $\mathbb{C}_{NS}$, and then apply Fano’s inequality. For readability, we define the packing $\mathcal{P}$ by a sequence of constraints. First, every matrix in $\mathcal{P}$ is chosen to be $M_{NS}(\pi, \lambda)$ for a fixed $\lambda$ and some permutation $\pi$, so we focus on selecting the permutations $\pi$.

Longest path lower bound. To prove the bound in terms of $L(G)$, consider any path $T$ of odd length $\ell$ in the graph $G$ without loss of generality. We index the vertices of $T$ by $1, \ldots, \ell + 1$ and pair the adjacent vertices as $(1, 2), (3, 4), \ldots, (\ell, \ell + 1)$. The rest of the vertices are indexed by $\ell + 2, \ldots, n$.

For every instance in the packing, we require that $\pi(i) = i$ for $i \geq \ell + 2$, and

\[
\begin{align*}
\text{either } \begin{cases} 
\pi(2i - 1) = 2i - 1 \\
\pi(2i) = 2i 
\end{cases} 
\text{ or } \begin{cases} 
\pi(2i - 1) = 2i \\
\pi(2i) = 2i - 1 
\end{cases}
\end{align*}
\]

for $i \leq \frac{\ell + 1}{2}$. The Gilbert-Varshamov bound guarantees that among all $2^{\ell + 1}\frac{\ell + 1}{2}$ such permutations, there exist $2^{a\ell}$ of them, any two of which satisfy that $K_T(\pi, \pi') \geq b\ell$, for some constants $a$ and $b$.

We choose the packing $\mathcal{P}$ consisting of matrices corresponding to these permutations, so that $\|M - M'\|_E^2 \geq 8b\lambda^2\ell$ for any distinct $M, M' \in \mathcal{P}$. Moreover, the KL divergence between any two models with underlying matrices in the packing is bounded by $9\lambda^2\ell \frac{\ell + 1}{2}$ by Lemma D.2. Finally, Fano’s inequality implies that

\[
|E| \mathcal{E}(G, \mathbb{C}_{NS}) \geq 8b\lambda^2\ell \left(1 - \frac{9\lambda^2(\ell + 1)/2 + \log 2}{a\ell}\right).
\]

The conclusion then follows by choosing $\lambda$ to be a sufficiently small constant.

Maximum star lower bound. Next, we turn to the bound in terms of $N(G)$. Consider any star $S$ in the graph, and let $d = |E(S)|$ so that $|V(S)| = d + 1$. To construct the packing $\mathcal{P}$, let the vertices of $S$ form the top $d + 1$ items and choose the same ranking for the vertices of $S^c$ for each instance in the packing. Then all the matrices in the packing have the same $(i, j)$-th entry for $i \in S^c$ or $j \in S^c$. Hence the KL divergence between any two models with underlying matrices in the packing $\mathcal{P}$ is bounded by $9\lambda^2d$ by Lemma D.2.

On the other hand, all the $2^d$ assignments of values to the edges in $E(S)$

\[
\{M_{ij} : (i, j) \in S, i < j\} \in \{1/2 + \lambda, 1/2 - \lambda\}^d
\]

are possible, since there are no cycle conflicts. The Gilbert-Varshamov bound guarantees that there exist $2^{ad}$ such assignments, any two of which are separated by $bd$ in the Hamming distance, for some constants $a$ and $b$. We choose
the packing $\mathcal{P}$ consisting of matrices corresponding to these assignments, so that $\|M - M'\|_E^2 \geq 8b\lambda^2d$ for any distinct $M, M' \in \mathcal{P}$.

Finally, Fano’s inequality implies that

$$|E| E(G, \mathcal{C}_{NS}) \geq 8b\lambda^2d \left(1 - \frac{9\lambda^2d + \log 2}{ad}\right).$$

The conclusion follows by choosing $\lambda$ to be a sufficiently small constant. □

F.1.2. Proof of upper bound. As mentioned before, we obtain the upper bound by considering the estimator $\hat{M}_{LS}$. The proof follows from previous results on the full observation case [8], but we provide it for completeness. Note that for each $(i,j) \in E$, the observation model takes the form

$$Y_{ij} = M^*_ij + W_{ij},$$

where $W_{ij}$ is a zero-mean noise variable lying in the interval $[-1, 1]$.

The optimality of $\hat{M}_{LS}$ and feasibility of $M^*$ imply that we must have the basic inequality $\|Y - \hat{M}_{LS}\|_E^2 \leq \|Y - M^*\|_E^2$, which after simplification, leads to

$$\frac{1}{2} \|\Delta\|_E^2 \leq \langle \Delta, W \rangle_E,$$

(75)

where $\Delta = \hat{M}_{LS} - M^*$, and $\langle A, B \rangle_E = \sum_{(i,j) \in E} A_{ij}B_{ij}$ denotes the trace inner product restricted to the indices in $E$.

In order to establish the upper bound, we first define the class of difference matrices $\mathcal{C}_{\text{DIFF}} := \{M - M' \mid M, M' \in \mathcal{C}_{\text{SST}}\}$, as well as the associated random variable

$$Z(t) := \sup_{D \in \mathcal{C}_{\text{DIFF}} : \|D\|_E \leq t} \langle D, W \rangle_E.$$

With this notation, inequality (75) implies $\frac{1}{2} \|\Delta\|_E^2 \leq Z(\|\Delta\|_E)$. It follows from the star-shaped property\(^2\) of the set $\mathcal{C}_{\text{DIFF}}$ that the following critical inequality is satisfied for some $\delta > 0$:

$$\mathbb{E}[Z(\delta)] \leq \frac{\delta^2}{2}.$$

We are interested in the smallest such value $\delta$. In order to find it, we use Dudley’s entropy integral, for which we require a bound on the covering number of the class $\mathcal{C}_{\text{DIFF}}$. Such a bound was calculated for the Frobenius

\(^2\)A set $S$ is said to be star-shaped if $t \in S$ implies that $\alpha t \in S$ for all $\alpha \in [0, 1]$
norm by Shah et al. [8] using the results of Gao and Wellner [5]. Clearly, since $\|M_i - M_j\|_E^2 \leq \|M_i - M_j\|_F^2$, a $\delta$-covering in the Frobenius norm automatically serves as a $\delta$-covering in the edge norm $\|\cdot\|_E$. Thus, we have the following lemma.

**Lemma F.1.** [8] For every $\epsilon > 0$, we have the metric entropy bound

$$\log N(\epsilon, C_{\text{DIFF}}, \|\cdot\|_E) \leq \log N(\epsilon, C_{\text{DIFF}}, \|\cdot\|_F) \leq \frac{9n^2}{\epsilon^2} \left( \log \frac{n}{\epsilon} \right)^2 + 9n \log n.$$  

Dudley’s entropy integral then yields that for all $t > 0$, we have

$$\mathbb{E}[Z(t)] \leq c \inf_{\delta \in [0,n]} \left\{ n\delta + \int_{\delta/2}^t \sqrt{\log N(\epsilon, C_{\text{DIFF}} \cap B_E(t), \|\cdot\|_E)} \, d\epsilon \right\}$$

$$\leq c \left\{ n^{-\delta} + \int_{n^{-\delta/2}}^t \sqrt{\log N(\epsilon, C_{\text{DIFF}}, \|\cdot\|_E)} \, d\epsilon \right\}.$$

After some algebra (for details, see Shah et al. [8]), we have

$$\mathbb{E}[Z(t)] \leq c \left\{ n \log^2 n + t \sqrt{n \log n} \right\}.$$

Setting $t = c \sqrt{n \log n}$ completes the proof. 

**REFERENCES**


Department of EECS
Department of Statistics
University of California Berkeley
Berkeley, California 94720
USA
E-mail: ashwinpm@berkeley.edu
vidya.mathukumar@berkeley.edu
wainwrig@stat.berkeley.edu
courtade@berkeley.edu

Department of Mathematics
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, Massachusetts 02139
USA
E-mail: maocheng@mit.edu