APPENDIX

A. Proofs

This section is devoted to the proofs of our theoretical results–namely, Proposition 1 and Theorems 1 through 5. Throughout these and other proofs, we use the notation $\{c, c', c_0, C, C'\}$ and so on to denote positive constants whose values may change from line to line. In addition, we assume throughout that n is lower bounded by a universal constant so as to avoid degeneracies. For any square matrix $A \in \mathbb{R}^{n \times n}$, we let $\{\sigma_1(A), \ldots, \sigma_n(A)\}$ denote its singular values (ordered from largest to smallest), and similarly, for any symmetric matrix $M \in \mathbb{R}^{n \times n}$, we let $\{\lambda_1(M), \ldots, \lambda_n(M)\}$ denote its ordered eigenvalues.

A.1. Proof of Proposition 1

We will show that the matrix M^* specified in Figure 1a satisfies the conditions required by the proposition. It is easy to verify that $M^* \in \mathbb{C}_{\text{SST}}$, so that it remains to prove the approximation-theoretic lower bound (4). In order to do so, we require the following auxiliary result:

Lemma 1. *Consider any matrix M that belongs to $\mathbb{C}_{\text{PAR}}(F)$ for a valid function F. Suppose for some collection of four distinct items $\{i_1, \ldots, i_4\}$, the matrix M satisfies the inequality $M_{i_1 i_2} > M_{i_3 i_4}$. Then it must also satisfy the inequality $M_{i_1 i_3} \geq M_{i_2 i_4}$.*

We return to prove this lemma at the end of this section. Taking it as given, let us now proceed to prove the lower bound (4). For any valid F, fix an arbitrary member M of a class $\mathbb{C}_{\text{PAR}}(F)$, and let $w \in \mathbb{R}^n$ be the underlying weight vector (see the definition (2)).

Pick any item in the set of first $\frac{n}{4}$ items (corresponding to the first $\frac{n}{4}$ rows of M^*) and call this item as "1"; pick an item from the next set of $\frac{n}{4}$ items (rows) and call it item "2"; item "3" from the next set and item "4" from the final set. Our analysis proceeds by developing some relations between the pairwise comparison probabilities for these four items that must hold for every parametric model, that are strongly violated by M^*. We divide our analysis into two possible relations between the entries of M.

<u>Case I:</u> First suppose that $M_{12} \leq M_{34}$. Since $M^*_{12} = 6/8$ and $M^*_{34} = 5/8$ in our construction, it follows that

$$(M_{12} - M^*_{12})^2 + (M_{34} - M^*_{34})^2 \geq \frac{1}{256}.$$

<u>Case II:</u> Otherwise, we may assume that $M_{12} > M_{34}$. Then Lemma 1 implies that $M_{13} \geq M_{24}$. Moreover, since $M^*_{13} = 7/8$ and $M^*_{24} = 1$ in our construction, it follows that

$$(M_{13} - M^*_{13})^2 + (M_{24} - M^*_{24})^2 \geq \frac{1}{256}.$$

Aggregating across these two exhaustive cases, we find that

$$\sum_{(u,v) \in \{1,2,3,4\}} (M_{uv} - M^*_{uv})^2 \geq \frac{1}{256}.$$

Since this bound holds for any arbitrary selection of items from the four sets, we conclude that $\frac{1}{n^2} |\!|\!| M - M^* |\!|\!|_{\text{F}}^2$ is lower bounded by a universal constant $c > 0$ as claimed.

Finally, it is easy to see that upon perturbation of any of the entries of M^* by at most $\frac{1}{32}$—while still ensuring that the resulting matrix lies in $\mathbb{C}_{\text{SST}}$—the aforementioned results continue to hold, albeit with a worse constant. Every matrix in this class satisfies the claim of this proposition.

Proof of Lemma 1: It remains to prove Lemma 1. Since M belongs to the parametric family, there must exist some valid function F and some vector w that induce M (see Equation (2)). Since F is non-decreasing, the condition $M_{i_1 i_2} > M_{i_3 i_4}$ implies that

$$w_{i_1} - w_{i_2} > w_{i_3} - w_{i_4}.$$

Adding $w_{i_2} - w_{i_3}$ to both sides of this inequality yields $w_{i_1} - w_{i_3} > w_{i_2} - w_{i_4}$. Finally, applying the non-decreasing function F to both sides of this inequality gives yields $M_{i_1 i_3} \geq M_{i_2 i_4}$ as claimed, thereby completing the proof.

A.2. Proof of Theorem 1

This section is devoted to the proof of Theorem 1, including both the upper and lower bounds on the minimax risk in squared Frobenius norm.

A.2.1. Proof of upper bound

Observe that we can always write the Bernoulli observation model in the linear form

$$Y = M^* + W, \tag{10}$$

where $W \in [-1,1]^{n \times n}$ is a random matrix with independent zero-mean entries for every $i \geq j$ given by

$$W_{ij} \sim \begin{cases} 1 - M^*_{ij} & \text{with probability } M^*_{ij} \\ -M^*_{ij} & \text{with probability } 1 - M^*_{ij}, \end{cases} \tag{11}$$

and $W_{ji} = -W_{ij}$ for every $i < j$. This linearized form of the observation model is convenient for subsequent analysis.

Define the difference $\widehat{\Delta} = \widehat{M} - M^*$ between M^* and the optimal solution $\widehat{M}$ to the constrained least-squares problem. Since $\widehat{M}$ is optimal and M^* is feasible, we have $|\!|\!| Y - \widehat{M} |\!|\!|_{\text{F}}^2 \leq |\!|\!| Y - M^* |\!|\!|_{\text{F}}^2$, and hence following some algebra, we arrive at the basic inequality

$$\frac{1}{2} |\!|\!| \widehat{\Delta} |\!|\!|_{\text{F}}^2 \leq \langle\!\langle \widehat{\Delta}, \; W \rangle\!\rangle,$$

where $W \in \mathbb{R}^{n \times n}$ is the noise matrix in the observation model (10), and $\langle\!\langle A, \; B \rangle\!\rangle = \text{trace}(A^T B)$ is the trace inner product.

We introduce some additional objects that are useful in our analysis. First, we recall our earlier definition of the class of bivariate isotonic matrices under the identity permutation, $\mathbb{C}_{\text{BISO}} \in [0,1]^{n \times n}$, is the set:

$$\mathbb{C}_{\text{BISO}} := \big\{ M \in [0,1]^{n \times n} \mid M_{k\ell} \geq M_{ij} \text{ whenever } k \leq i \text{ and } \ell \geq j \big\}. \tag{12}$$

For a given permutation π and matrix M, we let $\pi(M)$ denote the matrix obtained by applying π to its rows and columns. Now let $\mathbb{C}_{\text{DIFF}} \in [-1,1]^{n \times n}$ be the set of differences between pairs of matrices in $\mathbb{C}_{\text{BISO}}$, that is,

$$\mathbb{C}_{\text{DIFF}} := \Big\{ \pi_1(M_1) - \pi_2(M_2) \mid \text{for some } M_1, \; M_2 \in \mathbb{C}_{\text{BISO}}, \text{ and perm. } \pi_1 \text{ and } \pi_2 \Big\}. \tag{13}$$

One can verify that for any $M^* \in \mathbb{C}_{\text{SST}}$, we are guaranteed the inclusion

$$\{ M - M^* \mid M \in \mathbb{C}_{\text{SST}}, \; |\!|\!| M - M^* |\!|\!|_{\text{F}} \leq t \} \subset \{ D \in \mathbb{C}_{\text{DIFF}} \mid |\!|\!| D |\!|\!|_{\text{F}} \leq t \}.$$

Now for each choice of radius $t > 0$, define the random variable

$$Z(t) := \sup_{D \in \mathbb{C}_{\text{DIFF}}, |\!|\!| D |\!|\!|_{\text{F}} \leq t} \langle\!\langle D, \; W \rangle\!\rangle. \tag{14}$$

Using our earlier basic inequality, the Frobenius norm error $|\!|\!| \widehat{\Delta} |\!|\!|_{\text{F}}$ then satisfies the bound

$$\frac{1}{2} |\!|\!| \widehat{\Delta} |\!|\!|_{\text{F}}^2 \leq \langle\!\langle \widehat{\Delta}, \; W \rangle\!\rangle \; \leq \; Z\big(|\!|\!| \widehat{\Delta} |\!|\!|_{\text{F}} \big). \tag{15}$$

Thus, in order to obtain a high probability bound, we need to understand the behavior of the random quantity $Z(\delta)$.

One can verify that the set $\mathbb{C}_{\text{DIFF}}$ is star-shaped, meaning that $\alpha D \in \mathbb{C}_{\text{DIFF}}$ for every $\alpha \in [0,1]$ and every $D \in \mathbb{C}_{\text{DIFF}}$. Using this star-shaped property, we are guaranteed that there is a non-empty set of scalars $\delta_n > 0$ satisfying the critical inequality

$$\mathbb{E}[Z(\delta_n)] \leq \frac{\delta_n^2}{2}. \tag{16}$$

Our interest is in the smallest (strictly) positive solution δ_n to the critical inequality (16), and moreover, our goal is to show that for every $t \geq \delta_n$, we have $|\!|\!|\widehat{\Delta}|\!|\!|_{\text{F}} \leq c\sqrt{t\delta_n}$ with probability at least $1 - c_1 e^{-c_2 n t \delta_n}$.

Define a "bad" event $\mathcal{A}_t$ as

$$\mathcal{A}_t = \left\{ \exists \Delta \in \mathbb{C}_{\text{DIFF}} \mid |\!|\!|\Delta|\!|\!|_{\text{F}} \geq \sqrt{t\delta_n} \quad \text{and} \quad \langle\!\langle \Delta,\ W \rangle\!\rangle \geq 2|\!|\!|\Delta|\!|\!|_{\text{F}}\sqrt{t\delta_n} \right\}. \tag{17}$$

Using the star-shaped property of $\mathbb{C}_{\text{DIFF}}$, it follows by a rescaling argument that

$$\mathbb{P}[\mathcal{A}_t] \leq \mathbb{P}[Z(\delta_n) \geq 2\delta_n \sqrt{t\delta_n}] \qquad \text{for all } t \geq \delta_n.$$

The entries of W lie in $[-1, 1]$ and hence are 1-sub-Gaussian, are i.i.d. on and above the diagonal, are zero-mean, and satisfy skew-symmetry. Moreover, the function $W \mapsto Z(t)$ is convex and Lipschitz with parameter t. Consequently, by Ledoux's concentration theorem (Ledoux, 2001, Theorem 5.9), we have

$$\mathbb{P}\big[Z(\delta_n) \geq \mathbb{E}[Z(\delta_n)] + \sqrt{t\delta_n}\delta_n\big] \leq 2e^{-c_1 t \delta_n} \qquad \text{for all } t \geq \delta_n.$$

By the definition of δ_n, we have $\mathbb{E}[Z(\delta_n)] \leq \delta_n^2 \leq \delta_n\sqrt{t\delta_n}$ for any $t \geq \delta_n$, and consequently

$$\mathbb{P}[\mathcal{A}_t] \leq \mathbb{P}[Z(\delta_n) \geq 2\delta_n \sqrt{t\delta_n}] \ \leq \ 2e^{-c_1 t \delta_n} \quad \text{for all } t \geq \delta_n.$$

Consequently, either $|\!|\!|\widehat{\Delta}|\!|\!|_{\text{F}} \leq \sqrt{t\delta_n}$, or we have $|\!|\!|\widehat{\Delta}|\!|\!|_{\text{F}} > \sqrt{t\delta_n}$. In the latter case, conditioning on the complement $\mathcal{A}_t^c$, our basic inequality implies that $\frac{1}{2}|\!|\!|\widehat{\Delta}|\!|\!|_{\text{F}}^2 \leq 2|\!|\!|\widehat{\Delta}|\!|\!|_{\text{F}}\sqrt{t\delta_n}$, and hence $|\!|\!|\widehat{\Delta}|\!|\!|_{\text{F}} \leq 4\sqrt{t\delta_n}$ with probability at least $1 - 2e^{-c_1 t\delta_n}$. Putting together the pieces yields that

$$|\!|\!|\widehat{\Delta}|\!|\!|_{\text{F}} \leq c_0\sqrt{t\delta_n} \tag{18}$$

with probability at least $1 - 2e^{-c_1 t \delta_n}$ for every $t \geq \delta_n$.

In order to determine a feasible δ_n satisfying the critical inequality (16), we need to bound the expectation $\mathbb{E}[Z(\delta_n)]$. We do using Dudley's entropy integral and bounding the metric entropies of certain sub-classes of matrices. In particular, the remainder of this section is devoted to proving the following claim:

Lemma 2. *There is a universal constant C such that for all $t \in [0, 2n]$,*

$$\mathbb{E}[Z(t)] \leq C\left\{n \log^2(n) + t\sqrt{n \log n}\right\}. \tag{19}$$

Taking this lemma as given for the moment, we see that the critical inequality (16) is satisfied for $\delta_n = C'\sqrt{n}\log n$, and hence from our bound (18), we have

$$\frac{|\!|\!|\widehat{\Delta}|\!|\!|_{\text{F}}^2}{n^2} \leq C'' \frac{\log^2(n)}{n},$$

with probability at least $1 - 2e^{-c_1 n (\log n)^2}$, where C'' and c_1 are positive universal constants. This argument completes the proof of the upper bound.

It remains to prove Lemma 2, and we do so by using Dudley's entropy integral, as well as some auxiliary results on metric entropy. We use the notation $\log N(\epsilon, \mathbb{C}, \rho)$ to denote the ϵ metric entropy of the class $\mathbb{C}$ in the metric ρ. Our proof requires the following auxiliary lemma:

Lemma 3. *For every $\epsilon > 0$, we have the metric entropy bound*

$$\log N(\epsilon, \mathbb{C}_{\textit{DIFF}}, |\!|\!|.|\!|\!|_{F}) \leq 8\frac{n^2}{\epsilon^2}\big(\log\frac{n}{\epsilon}\big)^2 + 8n\log n.$$

See the end of this section for the proof of this claim. Taking it as given for the moment, let us now prove Lemma 2. Letting $\mathbb{B}_F(t)$ denote the Frobenius norm ball of radius t, the truncated form of Dudley's entropy integral inequality yields

$$\begin{aligned}\mathbb{E}[Z(t)]] &\leq c \inf_{\delta \in [0,n]} \Big\{ n\delta + \int_{\frac{\delta}{2}}^{t} \sqrt{\log N(\epsilon, \mathbb{C}_{\text{DIFF}} \cap \mathbb{B}_F(t), \|\!|.\|\!|_{\text{F}})} d\epsilon \Big\} \\ &\leq c \Big\{ n^{-8} + \int_{\frac{1}{2}n^{-9}}^{t} \sqrt{\log N(\epsilon, \mathbb{C}_{\text{DIFF}}, \|\!|.\|\!|_{\text{F}})} d\epsilon \Big\}, \qquad (20)\end{aligned}$$

where the second step follows by setting $\delta = n^{-9}$, and making use of the set inclusion $(\mathbb{C}_{\text{DIFF}} \cap \mathbb{B}_F(t)) \subseteq \mathbb{C}_{\text{DIFF}}$. For any $\epsilon \geq \frac{1}{2}n^{-9}$, applying Lemma 3 yields the upper bound

$$\sqrt{\log N(\epsilon, \mathbb{C}_{\text{DIFF}}, \|\!|.\|\!|_{\text{F}})} \leq c \Big\{ \frac{n}{\epsilon} \log \frac{n}{\epsilon} + \sqrt{n \log n} \Big\}.$$

Over the range $\epsilon \geq n^{-9}/2$, we have $\log \frac{n}{\epsilon} \leq c \log n$, and hence

$$\sqrt{\log N(\epsilon, \mathbb{C}_{\text{DIFF}}, \|\!|.\|\!|_{\text{F}})} \leq c \Big\{ \frac{n}{\epsilon} \log n + \sqrt{n \log n} \Big\}.$$

Substituting this bound into our earlier inequality (20) yields

$$\begin{aligned}\mathbb{E}[Z(t)] &\leq c \Big\{ n^{-8} + \big(n \log n\big) \log(nt) + t\sqrt{n \log n} \Big\} \\ &\overset{(i)}{\leq} c \Big\{ \big(n \log n\big) \log(n^2) + t\sqrt{n \log n} \Big\} \\ &\leq c \Big\{ n \log^2(n) + t\sqrt{n \log n} \Big\},\end{aligned}$$

where step (i) uses the upper bound $t \leq 2n$.

The only remaining detail is to prove Lemma 3.

Proof of Lemma 3: We first derive an upper bound on the metric entropy of the class $\mathbb{C}_{\text{BISO}}$ (defined in (12)). Let $\mathcal{F}$ denote the set of all bivariate monotonic functions $[0,1]^2 \to [0,1]^2$. For any matrix $M \in \mathbb{C}_{\text{BISO}}$, choose $g_M \in \mathcal{F}$ as

$$g_M(x,y) = M_{\lceil n(1-x) \rceil, \lceil ny \rceil},$$

where to handle corner conditions we define $M_{0,i} = M_{1,i}$ and $M_{i,0} = M_{i,1}$ for all i. Then we have

$$\|g_M\|_2^2 = \int_{x=0}^{1} \int_{y=0}^{1} (g_M(x,y))^2 dx dy = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} M_{i,j}^2 = \frac{1}{n^2} \|\!|M\|\!|_{\text{F}}^2.$$

Thus, we have

$$\log N(\epsilon, \mathbb{C}_{\text{BISO}}, \|\!|.\|\!|_{\text{F}}) \leq \log N\left(\frac{\epsilon}{n}, \mathcal{F}, \|.\|_2\right) \;\leq\; \frac{n^2}{\epsilon^2} \big(\log \frac{n}{\epsilon}\big)^2, \qquad (21)$$

where the last inequality (21) follows from Theorem 1.1. of Gao & Wellner (2007).

We now bound the metric entropy of $\mathbb{C}_{\text{DIFF}}$ in terms of the metric entropy of $\mathbb{C}_{\text{BISO}}$. For any $\epsilon > 0$, let $\mathbb{C}_{\text{BISO}}^{\epsilon}$ denote an ϵ-covering set in $\mathbb{C}_{\text{BISO}}$ that satisfies the inequality (21). Consider the set

$$\mathbb{C}_{\text{DIFF}}^{\epsilon} := \{\pi_1(M_1) - \pi_2(M_2) \mid \text{for some permutations } \pi_1, \pi_2 \text{ and some } M_1, M_2 \in \mathbb{C}_{\text{BISO}}^{\epsilon/\sqrt{2}}\}.$$

For any $D \in \mathbb{C}_{\text{DIFF}}$, we can write $D = \pi_1(M_1') - \pi_2(M_2')$ for some permutations π_1 and π_2 and some matrices M_1' and $M_2' \in \mathbb{C}_{\text{BISO}}$. We know there exist some $M_1, M_2 \in \mathbb{C}_{\text{BISO}}^{\epsilon/\sqrt{2}}$ such that $\|\!|M_1' - M_1\|\!|_{\text{F}} \leq \epsilon/\sqrt{2}$ and $\|\!|M_2' - M_2\|\!|_{\text{F}} \leq \epsilon/\sqrt{2}$. Then we have $\pi_1(M_1) - \pi_2(M_2) \in \mathbb{C}_{\text{DIFF}}^{\epsilon}$, and moreover

$$\begin{aligned}\|\!|D - (\pi_1(M_1) - \pi_2(M_2))\|\!|_{\text{F}}^2 &\leq \|\!|\pi_1(M_1) - \pi_1(M_1')\|\!|_{\text{F}}^2 + \|\!|\pi_2(M_2) - \pi_1(M_2')\|\!|_{\text{F}}^2 \\ &\leq \epsilon^2.\end{aligned}$$

Thus the set $\mathbb{C}^{\epsilon}_{\text{DIFF}}$ forms an ϵ-covering set for the class $\mathbb{C}_{\text{DIFF}}$. One can now count the number of elements in this set to get

$$N(\epsilon, \mathbb{C}_{\text{DIFF}}, \|\!|.\|\!|_{\text{F}}) \leq \big(n! N(\epsilon/\sqrt{2}, \mathbb{C}_{\text{BISO}}, \|\!|.\|\!|_{\text{F}})\big)^2.$$

Some straightforward algebraic manipulations yield the claimed result.

A.3. Proof of Theorem 2

Recall from (10) that we can write our observation model as $Y = M^* + W$, where $W \in \mathbb{R}^{n \times n}$ is a zero-mean matrix with entries that are drawn independently (except for the skew-symmetry condition) from the interval $[-1, 1]$.

Our proof of the upper bound hinges upon the following two lemmas.

Lemma 4. *If $\lambda_n \geq 1.01 \|\!|W\|\!|_{op}$, then*

$$\|\!|T_{\lambda_n}(Y) - M^*\|\!|_F^2 \leq c \sum_{j=1}^{n} \min\left\{\lambda_n^2, \sigma_j^2(M^*)\right\}$$

with probability at least $1 - c_1 e^{-c' n}$, where c, c_1 and c' are positive universal constants.

Our second lemma is an approximation-theoretic result:

Lemma 5. *For any matrix $M^* \in \mathbb{C}_{SST}$ and any $s \in \{1, 2, \ldots, n-1\}$, we have*

$$\frac{1}{n^2} \sum_{j=s+1}^{n} \sigma_j^2(M^*) \leq \frac{1}{s}.$$

See the end of this section for the proofs of these two auxiliary results.

Based on these two lemmas, it is easy to complete the proof of the theorem. The entries of W lie in $[-1, 1]$ and hence are 1-sub-Gaussian, are i.i.d. on and above the diagonal, are zero-mean, and satisfy skew-symmetry. Consequently, we may apply Theorem 3.4 of Chatterjee (2014), which guarantees that

$$\mathbb{P}\Big[\|\!|W\|\!|_{\text{op}} > (2+t)\sqrt{n}\Big] \leq c e^{-f(t) n},$$

where c is a universal constant, and the quantity $f(t)$ is strictly positive for each $t > 0$. Thus, the choice $\lambda_n = 2.1\sqrt{n}$ guarantees that $\lambda_n \geq 1.01 \|\!|W\|\!|_{\text{op}}$ with probability at least $1 - ce^{-cn}$, as is required for applying Lemma 4. Applying this lemma then yields the upper bound

$$\|\!|T_{\lambda_n}(Y) - M^*\|\!|_{\text{F}}^2 \leq c \sum_{j=1}^{n} \min\left\{n, \sigma_j^2(M^*)\right\}$$

with probability at least $1 - c_1 e^{-c_2 n}$. Applying Lemma 5 yields that for any $s \in \{1, \ldots, n\}$,

$$\frac{1}{n^2} \|\!|T_{\lambda_n}(Y) - M^*\|\!|_{\text{F}}^2 \leq c\Big\{\frac{s}{n} + \frac{1}{s}\Big\},$$

with probability at least $1 - c_1 e^{-c_2 n}$. Setting $s = \lceil \sqrt{n} \rceil$ and performing some algebra shows that

$$\mathbb{P}\Big[\frac{1}{n^2} \|\!|T_{\lambda_n}(Y) - M^*\|\!|_{\text{F}}^2 > \frac{c_u}{\sqrt{n}}\Big] \leq c_1 e^{-c_2 n},$$

as claimed. Since $\frac{1}{n^2} \|\!|T_{\lambda_n}(Y) - M^*\|\!|_{\text{F}}^2 \leq 1$, we are also guaranteed that

$$\frac{1}{n^2} \mathbb{E}[\|\!|T_{\lambda_n}(Y) - M^*\|\!|_{\text{F}}^2] \leq \frac{c_u}{\sqrt{n}} + c_1 e^{-c_2 n} \; \leq \frac{c_u'}{\sqrt{n}}.$$

Proof of Lemma 4 Fix $\delta = 0.01$. Let b be the number of singular values of M^* above $\frac{\delta}{1+\delta}\lambda_n$, and let M_b^* be the version of M^* truncated to its top b singular values. We then have

$$\|T_{\lambda_n}(Y) - M^*\|_{\mathrm{F}}^2 \le 2\|T_{\lambda_n}(Y) - M_b^*\|_{\mathrm{F}}^2 + 2\|M_b^* - M^*\|_{\mathrm{F}}^2$$
$$\le 2\,\mathrm{rank}(T_{\lambda_n}(Y) - M_b^*)\|T_{\lambda_n}(Y) - M_b^*\|_{\mathrm{op}}^2 + 2\sum_{j=b+1}^{n} \sigma_j^2(M^*).$$

We claim that $T_{\lambda_n}(Y)$ has rank at most b. Indeed, for any $j \ge b+1$, we have

$$\sigma_j(Y) \le \sigma_j(M^*) + \|W\|_{\mathrm{op}} \;\le\; \lambda_n,$$

where we have used the facts that $\sigma_j(M^*) \le \frac{\delta}{1+\delta}\lambda_n$ for every $j \ge b+1$ and $\lambda_n \ge (1+\delta)\|W\|_{\mathrm{op}}$. As a consequence we have $\sigma_j(T_{\lambda_n}(Y)) = 0$, and hence $\mathrm{rank}(T_{\lambda_n}(Y) - M_b^*) \le 2b$. Moreover, we have

$$\|T_{\lambda_n}(Y) - M_b^*\|_{\mathrm{op}} \le \|T_{\lambda_n}(Y) - Y\|_{\mathrm{op}} + \|Y - M^*\|_{\mathrm{op}} + \|M^* - M_b^*\|_{\mathrm{op}}$$
$$\le \lambda_n + \|W\|_{\mathrm{op}} + \frac{\delta}{1+\delta}\lambda_n$$
$$\le 2\lambda_n.$$

Putting together the pieces, we conclude that

$$\|T_{\lambda_n}(Y) - M^*\|_{\mathrm{F}}^2 \le 16b\lambda_n^2 + 2\sum_{j=b+1}^{n} \sigma_j^2(M^*) \overset{(i)}{\le} C\sum_{j=1}^{n} \min\{\sigma_j^2(M^*), \lambda_n^2\},$$

for some constant[2] C. Here inequality (i) follows since $\sigma_j(M^*) \le \frac{\delta}{1+\delta}\lambda_n$ whenever $j \ge b+1$ and $\sigma_j(M^*) > \frac{\delta}{1+\delta}\lambda_n$ whenever $j \le b$.

Proof of Lemma 5 In this proof, we make use of a construction due to Chatterjee (2014). For a given matrix M^*, we can define the vector $t \in \mathbb{R}^n$ of row sums—namely, with entries $t_i = \sum_{j=1}^n M_{ij}^*$ for $i \in [n]$. Using this vector, we can define a rank s approximation M to the original matrix M^* by grouping the rows according to the vector t according to the following procedure:

- Observing that each $t_i \in [0, n]$, let us divide the full interval $[0, n]$ into s groups—say of the form $[0, n/s), [n/s, 2n/s), \ldots [(s-1)n/s, n]$. If t_i falls into the interval α for some $\alpha \in [s]$, we then map row i to the group G_α of indices.
- For each group G_α, we choose a particular row index $k = k(\alpha) \in G_\alpha$ in an arbitrary fashion. For every other row index $i \in G_\alpha$, we set $M_{ij} = M_{kj}$ for all $j \in [n]$.

By construction, the matrix M has at most s distinct rows, and hence rank at most s. Let us now bound the Frobenius norm error in this rank s approximation. Fixing an arbitrary group index $\alpha \in [s]$ and an arbitrary row in $i \in G_\alpha$, we then have

$$\sum_{j=1}^{n}(M_{ij}^* - M_{ij})^2 \le \sum_{j=1}^{n} |M_{ij}^* - M_{ij}|.$$

By construction, we either have $M_{ij}^* \ge M_{ij}$ for every $j \in [n]$, or $M_{ij}^* \le M_{ij}$ for every $j \in [n]$. Thus, letting $k \in G_\alpha$ denote the chosen row, we are guaranteed that

$$\sum_{j=1}^{n} |M_{ij}^* - M_{ij}| \le |t_i - t_k| \le \frac{n}{s},$$

where we have used the fact the pair (t_i, t_k) must lie in an interval of length at most n/s. Putting together the pieces yields the claim.

[2]To be clear, the precise value of the constant C is determined by δ, which has been fixed as $\delta = 0.01$.

A.3.1. Proof of lower bound

We now turn to the proof of the lower bound in Theorem 2. We split our analysis into two cases, depending on the magnitude of λ_n.

Case 1: First suppose that $\lambda_n \leq \frac{\sqrt{n}}{3}$. In this case, we consider the matrix $M^* := \frac{1}{2} 1 1^T$ in which all items are equally good, so any comparison is simply a fair coin flip. Let the observation matrix $Y \in \{0,1\}^{n \times n}$ be arbitrary. By definition of the singular value thresholding operation, we have $|\!|\!| Y - T_{\lambda_n}(Y) |\!|\!|_{\text{op}} \leq \lambda_n$, and hence the SVT estimator $\widehat{M}_{\lambda_n} = T_{\lambda_n}(Y)$ has Frobenius norm at most

$$|\!|\!| Y - \widehat{M}_{\lambda_n} |\!|\!|_{\text{F}}^2 \leq n \lambda_n^2 \leq \frac{n^2}{9}.$$

Since $M^* \in \{\frac{1}{2}\}^{n \times n}$ and $Y \in \{0,1\}^{n \times n}$, we are guaranteed that $|\!|\!| M^* - Y |\!|\!|_{\text{F}} = \frac{n}{2}$. Applying the triangle inequality yields the lower bound

$$|\!|\!| \widehat{M}_{\lambda_n} - M^* |\!|\!|_{\text{F}} \;\geq\; |\!|\!| M^* - Y |\!|\!|_{\text{F}} - |\!|\!| \widehat{M}_{\lambda_n} - Y |\!|\!|_{\text{F}} \;\geq \frac{n}{2} - \frac{n}{3} \;=\; \frac{n}{6}.$$

Case 2: Otherwise, we may assume that $\lambda_n > \frac{\sqrt{n}}{3}$. Consider the matrix $M^* \in \mathbb{R}^{n \times n}$ with entries

$$[M^*]_{ij} = \begin{cases} 1 & \text{if } i > j \\ \frac{1}{2} & \text{if } i = j \\ 0 & \text{if } i < j. \end{cases} \tag{22}$$

By construction, the matrix M^* corresponds to the degenerate case of noiseless comparisons.

Consider the matrix $Y \in \mathbb{R}^{n \times n}$ generated according to the observation model (10). (To be clear, all of its off-diagonal entries are deterministic, whereas the diagonal is population with i.i.d. Bernoulli variates.) Our proof requires the following auxiliary result regarding the singular values of Y:

Lemma 6. *The singular values of the observation matrix $Y \in \mathbb{R}^{n \times n}$ generated by the noiseless comparison matrix M^* satisfy the bounds*

$$\frac{n}{4\pi(i+1)} - \frac{1}{2} \leq \sigma_{n-i-1}(Y) \leq \frac{n}{\pi(i-1)} + \frac{1}{2} \qquad \text{for all integers } i \in [1, \tfrac{n}{6} - 1].$$

We prove this lemma at the end of this section.

Taking it as given, we get that $\sigma_{n-i-1}(Y) \leq \frac{\sqrt{n}}{3}$ for every integer $i \geq 2\sqrt{n}$, and $\sigma_{n-i}(Y) \geq \frac{n}{50i}$ for every integer $i \in [1, \frac{n}{25}]$. It follows that

$$\sum_{i=1}^{n} (\sigma_i(Y))^2 \mathbf{1}\{\sigma_i(Y) \leq \frac{\sqrt{n}}{3}\} \geq \frac{n^2}{2500} \sum_{i=2\sqrt{n}}^{\frac{n}{25}} \frac{1}{i^2} \;\geq\; c n^{\frac{3}{2}},$$

for some universal constant $c > 0$. Recalling that $\lambda_n \geq \frac{\sqrt{n}}{3}$, we have the lower bound $|\!|\!| Y - \widehat{M}_{\lambda_n} |\!|\!|_{\text{F}}^2 \geq c n^{\frac{3}{2}}$. Furthermore, since the observations (apart from the diagonal entries) are noiseless, we have $|\!|\!| Y - M^* |\!|\!|_{\text{F}}^2 \leq \frac{n}{4}$. Putting the pieces together yields the lower bound

$$|\!|\!| \widehat{M}_{\lambda_n} - M^* |\!|\!|_{\text{F}} \;\geq |\!|\!| \widehat{M}_{\lambda_n} - Y |\!|\!|_{\text{F}} - |\!|\!| M^* - Y |\!|\!|_{\text{F}} \geq \; c n^{\frac{3}{4}} - \frac{\sqrt{n}}{2} \;\geq\; c' n^{\frac{3}{4}},$$

where the final step holds when n is large enough (i.e., larger than a universal constant).

Proof of Lemma 6: Instead of working with the original observation matrix Y, it is convenient to work with a transformed version. Define the matrix $\bar{Y} = Y - \text{diag}(Y) + I_n$, so that the matrix $\bar{Y}$ is identical to Y except that all its diagonal entries are set to 1. Using this intermediate object, define the $(n \times n)$ matrix

$$\widetilde{Y} := (\bar{Y}(\bar{Y})^T)^{-1} - e_n e_n^T, \tag{23}$$

where e_n denotes the n^{th} standard basis vector. One can verify that this matrix has entries

$$[\widetilde{Y}]_{ij} = \begin{cases} 1 & \text{if } i = j = 1 \text{ or } i = j = n \\ 2 & \text{if } 1 < i = j < n \\ -1 & \text{if } i = j + 1 \text{ or } i = j - 1 \\ 0 & \text{otherwise.} \end{cases}$$

Consequently, it is equal to the graph Laplacian[3] of an undirected chain graph on n nodes. Consequently, from standard results in spectral graph theory (Brouwer & Haemers, 2011), the eigenvalues of $\widetilde{Y}$ are given by $\{4\sin^2(\frac{\pi i}{n})\}_{i=0}^{n-1}$. Recall the elementary sandwich relationship $\frac{x}{2} \leq \sin x \leq x$, valid for every $x \in [0, \frac{\pi}{6}]$. Using this fact, we are guaranteed that

$$\frac{\pi^2 i^2}{n^2} \leq \lambda_{i+1}(\widetilde{Y}) \leq \frac{4\pi^2 i^2}{n^2} \quad \text{for all integers } i \in [1, \tfrac{n}{6}]. \tag{24}$$

We now use this intermediate result to establish the claimed bounds on the singular values of Y. Observe that the matrices $\widetilde{Y}$ and $(\bar{Y}(\bar{Y})^T)^{-1}$ differ only by the rank one matrix $e_n e_n^T$. Standard results in matrix perturbation theory (Thompson, 1976) guarantee that a rank-one perturbation can shift the position (in the large-to-small ordering) of any eigenvalue by at most one. Consequently, the eigenvalues of the matrix $(\bar{Y}(\bar{Y})^T)^{-1}$ must be sandwiched as

$$\frac{\pi^2 (i-1)^2}{n^2} \leq \lambda_{i+1}((\bar{Y}(\bar{Y})^T)^{-1}) \leq \frac{4\pi^2 (i+1)^2}{n^2} \qquad \text{for all integers } i \in [1, \tfrac{n}{6} - 1].$$

It follows that the singular values of $\bar{Y}$ are sandwiched as

$$\frac{n}{4\pi(i+1)} \leq \sigma_{n-i-1}(\bar{Y}) \leq \frac{n}{\pi(i-1)} \qquad \text{for all integers } i \in [1, \tfrac{n}{6} - 1].$$

Observe that $\bar{Y} - Y$ is a $\{0, \frac{1}{2}\}$-valued diagonal matrix, and hence $|\!|\!|\bar{Y} - Y|\!|\!|_{\text{op}} \leq \frac{1}{2}$, which implies that $\max_{i=1,\ldots,n} |\sigma_i(Y) - \sigma_i(\bar{Y})| \leq \frac{1}{2}$, and hence

$$\frac{n}{4\pi(i+1)} - \frac{1}{2} \leq \sigma_{n-i-1}(Y) \leq \frac{n}{\pi(i-1)} + \frac{1}{2} \qquad \text{for all integers } i \in [1, \tfrac{n}{6} - 1],$$

as claimed.

A.4. Proof of Theorem 3

We now prove our results on the high SNR subclass of $\mathbb{C}_{\text{SST}}$, in particular establishing a lower bound and then analyzing the two-stage estimator described in Section 3.3 so as to obtain the upper bound.

A.4.1. Proof of lower bound

In order to prove the lower bound, we follow the proof of the lower bound of Theorem 1, with the only difference being that the vector $q \in \mathbb{R}^{n-1}$ is restricted to lie in the interval $[\frac{1}{2} + \gamma, 1]^{n-1}$.

A.4.2. Proof of upper bound

Without loss of generality, assume that the true matrix M^* is associated to the identity permutation. Recall that the second step of our procedure involves performing constrained regression over the set $\mathbb{C}_{\text{BISO}}(\widehat{\pi}_{\text{FAS}})$. The error in such an estimate is necessarily of two types: the usual estimation error induced by the noise in our samples, and in addition, some form of approximation error that is induced by the difference between $\widehat{\pi}_{\text{FAS}}$ and the correct identity permutation.

In order to formalize this notion, for any fixed permutation π, consider the constrained least-squares estimator

$$\widehat{M}_{\pi} \in \arg\min_{M \in \mathbb{C}_{\text{BISO}}(\pi)} |\!|\!|Y - M|\!|\!|_{\text{F}}^2. \tag{25}$$

[3]In particular, the Laplacian of a graph is given by $L = D - A$, where A is the graph adjacency matrix, and D is the diagonal degree matrix.

Our first result provides an upper bound on the error matrix $\widehat{M}_\pi - M^*$ that involves both approximation and estimation error terms.

Lemma 7. *There is a universal constant $c_0 > 0$ such that error in the constrained LS estimate (25) satisfies the upper bound*

$$\frac{\|\widehat{M}_\pi - M^*\|_F^2}{c_0} \leq \underbrace{\|M^* - \pi(M^*)\|_F^2}_{\text{Approx. error}} + \underbrace{n\log^2(n)}_{\text{Estimation error}} \tag{26}$$

with probability at least $1 - c_1 e^{-c_2 n}$.

There are two remaining challenges in the proof. Since the second step of our estimator involves the FAS-minimizing permutation $\widehat{\pi}_{\text{FAS}}$, we cannot simply apply Lemma 7 to it directly. (The permutation $\widehat{\pi}_{\text{FAS}}$ is random, whereas this lemma applies to any fixed permutation). Consequently, we first need to extend the bound (26) to one that is uniform over a set that includes $\widehat{\pi}_{\text{FAS}}$ with high probability. Our second challenge is to upper bound the approximation error term $\|M^* - \widehat{\pi}_{\text{FAS}}(M^*)\|_{\text{F}}^2$ that is induced by using the permutation $\widehat{\pi}_{\text{FAS}}$ instead of the correct identity permutation.

In order to address these challenges, for any constant $c > 0$, define the set

$$\widehat{\Pi}(c) := \{\pi \mid \max_{i\in[n]} |i - \pi(i)| \leq c\log n\}.$$

This set corresponds to permutations that are relatively close to the identity permutation in the sup-norm sense. Our second lemma shows that any permutation in $\widehat{\Pi}(c)$ is "good enough" in the sense that the approximation error term in the upper bound (26) is well-controlled:

Lemma 8. *For any $M^* \in \mathbb{C}_{\textit{BISO}}$ and any permutation $\pi \in \widehat{\Pi}(c)$, we have*

$$\|M^* - \pi(M^*)\|_F^2 \leq 2c'' n\log n, \tag{27}$$

where c'' is a positive constant that may depend only on c.

Taking these two lemmas as given, let us now complete the proof of Theorem 3. (We return to prove these lemmas at the end of this section.) The work of Braverman & Mossel (2008) showed that for the class $\mathbb{C}_{\text{HIGH}}(\gamma)$, there exists a positive constant c—depending on γ but independent of n—such that

$$\mathbb{P}\Big[\widehat{\pi}_{\text{FAS}} \in \widehat{\Pi}(c)\Big] \geq 1 - \frac{c_3}{n^2}. \tag{28}$$

From the definition of class $\widehat{\Pi}(c)$, there is a positive constant c' (whose value may depend only on c) such that its cardinality is upper bounded as $\text{card}(\widehat{\Pi}(c)) \leq n^{2c'\log n}$. Consequently, by combining the union bound with Lemma 7 we conclude that, with probability at least $1 - c_1' e^{-c_2' n} - \frac{c_3}{n^2}$, the error matrix $\widehat{\Delta}_{\text{FAS}} := \widehat{M}_{\widehat{\pi}_{\text{FAS}}} - M^*$ satisfies the upper bound (26). Combined with the approximation-theoretic guarantee from Lemma 8, we find that

$$\begin{aligned}\frac{\|\widehat{\Delta}_{\text{FAS}}\|_{\text{F}}^2}{c_0} &\leq \|M^* - \widehat{\pi}_{\text{FAS}}(M^*)\|_{\text{F}}^2 + n\log^2(n)\\ &\leq c'' n\log n + +n\log^2(n),\end{aligned}$$

from which the claim follows.

It remains to prove the two auxiliary lemmas, and we do so in the following subsections.

Proof of Lemma 7: The proof of this lemma involves a slight generalization of the proof of the upper bound in Theorem 1 (see Section A.2.1 for this proof). From the optimality of $\widehat{M}_\pi$ and feasibility of $\pi(M^*)$ for the constrained least-squares program (25), we are guaranteed that $\|Y - \widehat{M}_\pi\|_{\text{F}}^2 \leq \|Y - \pi(M^*)\|_{\text{F}}^2$. Introducing the error matrix $\widehat{\Delta}_\pi := \widehat{M}_\pi - M^*$, some algebraic manipulations yield the modified basic inequality

$$\|\widehat{\Delta}_\pi\|_{\text{F}}^2 \leq \|M^* - \pi(M^*)\|_{\text{F}}^2 + 2\langle\!\langle W,\ \widehat{M}_\pi - \pi(M^*)\rangle\!\rangle.$$

Let us define $\widehat{\Delta} := \widehat{M}_\pi - \pi(M^*)$. Further, for each choice of radius $t > 0$, recall the definitions of the random variable $Z(t)$ and event $\mathcal{A}_t$ from equations (14) and (17), respectively. With these definitions, we have the upper bound

$$\|\widehat{\Delta}_\pi\|_{\text{F}}^2 \leq \|M^* - \pi(M^*)\|_{\text{F}}^2 + 2Z\big(\|\widehat{\Delta}\|_{\text{F}}\big). \tag{29}$$

Lemma 3 proved earlier shows that the inequality $\mathbb{E}[Z(\delta_n)] \leq \frac{\delta_n^2}{2}$ is satisfied by $\delta_n = c\sqrt{n}\log n$. In a manner identical to the proof in Section A.2.1, one can show that

$$\mathbb{P}[\mathcal{A}_t] \leq \mathbb{P}[Z(\delta_n) \geq 2\delta_n\sqrt{t\delta_n}] \;\leq\; 2e^{-c_1 t\delta_n} \quad \text{for all } t \geq \delta_n.$$

Given these results, we break the next step into two cases depending on the magnitude of $\widehat{\Delta}$. <u>Case I:</u> Suppose $\|\widehat{\Delta}\|_{\text{F}} \leq \sqrt{t\delta_n}$. In this case, we have

$$\begin{aligned}\|\widehat{\Delta}_\pi\|_{\text{F}}^2 &\leq 2\|M^* - \pi(M^*)\|_{\text{F}}^2 + 2\|\widehat{\Delta}\|_{\text{F}}^2 \\ &\leq 2\|M^* - \pi(M^*)\|_{\text{F}}^2 + t\delta_n.\end{aligned}$$

<u>Case II:</u> Or otherwise we will have $\|\widehat{\Delta}\|_{\text{F}} > \sqrt{t\delta_n}$. Conditioning on the complement $\mathcal{A}_t^c$, our basic inequality (29) implies that

$$\begin{aligned}\|\widehat{\Delta}_\pi\|_{\text{F}}^2 &\leq \|M^* - \pi(M^*)\|_{\text{F}}^2 + 4\|\widehat{\Delta}\|_{\text{F}}\sqrt{t\delta_n} \\ &\leq \|M^* - \pi(M^*)\|_{\text{F}}^2 + \frac{\|\widehat{\Delta}\|_{\text{F}}^2}{8} + 32t\delta_n, \\ &\leq \|M^* - \pi(M^*)\|_{\text{F}}^2 + \frac{2\|\widehat{\Delta}_\pi\|_{\text{F}}^2 + 2\|M^* - \pi(M^*)\|_{\text{F}}^2}{8} + 32t\delta_n,\end{aligned}$$

with probability at least $1 - 2e^{-c_1 t\delta_n}$.

Finally, setting $t = \delta_n = c\sqrt{n}\log(n)$ in either case and re-arranging yields the bound (26).

Proof of Lemma 8: For any matrix M and any value i, let M_i denote its i^{th} row. Also define the clipping function $b : \mathbb{Z} \to [n]$ via $b(x) = \min\{\max\{1, x\}, n\}$. Using this notation, we have

$$\begin{aligned}\|M^* - \pi(M^*)\|_{\text{F}}^2 &= \sum_{i=1}^n \|M_i^* - M_{\pi^{-1}(i)}^*\|_2^2 \\ &\leq \sum_{i=1}^n \max_{0 \leq j \leq c\log n}\{\|M_i^* - M_{b(i-j)}^*\|_2^2, \|M_i^* - M_{b(i+j)}^*\|_2^2\},\end{aligned}$$

where we have used the definition of the set $\widehat{\Pi}(c)$ to obtain the final inequality. Since M^* corresponds to the identity permutation, we have $M_1^* \geq M_2^* \geq \cdots \geq M_n^*$, where the inequalities are in the pointwise sense. Consequently, we have

$$\begin{aligned}\|M^* - \pi(M^*)\|_{\text{F}}^2 &\leq \sum_{i=1}^n \max\Big\{\|M_i^* - M_{b(i-c\log n)}^*\|_2^2, \|M_i^* - M_{b(i+c\log n)}^*\|_2^2\Big\} \\ &\leq 2\sum_{i=1}^{n-c\log n} \|M_i^* - M_{i+c\log n}^*\|_2^2.\end{aligned}$$

One can verify that the inequality $\sum_{i=1}^{k-1}(a_i - a_{i+1})^2 \leq (a_1 - a_k)^2$ holds for all ordered sequences of real numbers $a_1 \geq a_2 \geq \cdots \geq a_k$. As stated earlier, the rows of M^* dominate each other pointwise, and hence we conclude that

$$\|M^* - \pi(M^*)\|_{\text{F}}^2 \leq 2c\log n\|M_1^* - M_n^*\|_2^2 \;\leq\; 2cn\log n,$$

which establishes the claim (27).

A.5. Proof of Theorem 4

We now turn to our theorem giving upper and lower bounds on estimating pairwise probability matrices for parametric models. Let us begin with a proof of the claimed lower bound.

A.5.1. LOWER BOUND

We prove our lower bound by constructing a set of matrices that are well-separated in Frobenius norm. Using this set, we then use an argument based on Fano's inequality to lower bound the minimax risk. Underlying our construction of the matrix collection is a collection of Boolean vectors. For any two Boolean vectors $b, b' \in \{0,1\}^n$, let $d_{\mathrm{H}}(b, b') = \sum_{j=1}^n \mathbf{1}[b_j \neq b'_j]$ denote the Hamming distance between them.

Lemma 9. *For any fixed $\alpha \in (0, 1/4)$, there is a collection of Boolean vectors $\{b^1, \ldots, b^T\}$ such that*

$$\min\left\{d_H(b^j, b^k), d_H(b^j, 0)\right\} \geq \lceil \alpha n \rceil \qquad \text{for all distinct } j \neq k \in \{1, \ldots, T\}, \text{ and} \tag{30a}$$

$$T \equiv T(\alpha) \geq \exp\left\{(n-1)\, D_{\mathrm{KL}}(2\alpha \| \frac{1}{2})\right\} - 1. \tag{30b}$$

Given the collection $\{b^j, j \in [T(\alpha)]\}$ guaranteed by this lemma, we then define the collection of real vectors $\{w^j, j \in [T(\alpha)]\}$ via

$$w^j = \delta\Big(I - \frac{1}{n} 1 1^T\Big) b^j \qquad \text{for each } j \in [T(\alpha)],$$

where $\delta \in (0,1)$ is a parameter to be specified later in the proof. By construction, for each index $j \in [T(\alpha)]$, we have $\langle 1,\, w^j\rangle = 0$ and $\|w^j\|_\infty \leq \delta$. Based on these vectors, we then define the collection of matrices $\left\{M^j, j \in [T(\alpha)]\right\}$ via

$$[M^k]_{ij} := F([w^k]_i - [w^k]_j).$$

By construction, this collection of matrices is contained within our parametric family. We also claim that they are well-separated in Frobenius norm:

Lemma 10. *For any distinct pair $j, k \in [T(\alpha)]$, we have*

$$\frac{|\!|\!|M^j - M^k|\!|\!|_F^2}{n^2} \geq \frac{\alpha^2}{4}(F(\delta) - F(0))^2. \tag{31}$$

In order to apply Fano's inequality, our second requirement is an upper bound on the mutual information $I(Y; J)$, where J is a random index uniformly distributed over the index set $[T] = \{1, \ldots, T\}$. By Jensen's inequality, we have $I(Y; J) \leq \frac{1}{\binom{T}{2}} \sum_{j \neq k} D_{\mathrm{KL}}(\mathbb{P}^j \| \mathbb{P}^k)$, where $\mathbb{P}^j$ denotes the distribution of Y when the true underlying matrix is M^j. Let us upper bound these KL divergences.

For any pair of distinct indices $u, v \in [n]^2$, let x_{uv} be a differencing vector—that is, a vector in components u and v are set equal are 1 and -1, respectively, with all remaining components equal to 0. We are then guaranteed that

$$\langle x_{uv},\, w^j\rangle = \delta \langle x_{uv},\, b^j\rangle, \quad \text{and} \quad F(\langle x_{uv},\, w^j\rangle) \in \left\{F(-\delta), F(0), F(\delta)\right\},$$

where $F(\delta) \geq F(0) \geq F(-\delta)$ by construction. Using these facts, we have

$$\begin{aligned} D_{\mathrm{KL}}(\mathbb{P}^j \| \mathbb{P}^k) &\leq \frac{1}{2} \sum_{u,v \in [n]} \frac{\big(F(\langle x_{uv},\, w^j\rangle) - F(\langle x_{uv},\, w^k\rangle)\big)^2}{\min\{F(\langle x_{uv},\, w^k\rangle), 1 - F(\langle x_{uv},\, w^k\rangle)\}} \\ &\leq \frac{1}{2} n^2 \frac{(F(\delta) - F(-\delta))^2}{F(-\delta)} \\ &\leq 2n^2 \frac{(F(\delta) - F(0))^2}{F(-\delta)}. \end{aligned} \tag{32}$$

This upper bound on the KL divergence (32) and lower bound on the Frobenius norm (31), when combined with Fano's inequality, imply that any estimator $\widehat{M}$ has its worst-case risk over our family lower bounded as

$$\sup_{j \in [T(\alpha)]} \frac{1}{n^2} \mathbb{E}\big[\|\widehat{M} - M(w^j)\|_{\mathrm{F}}^2\big] \geq \frac{1}{8}\alpha^2 (F(\delta) - F(0))^2 \Big(1 - \frac{\frac{2}{F(-\delta)} n^2 (F(\delta) - F(0))^2 + \log 2}{n}\Big).$$

Choosing a value of $\delta > 0$ such that $(F(\delta) - F(0))^2 = \frac{F(-\delta)}{20n}$ gives the claimed result. (Such a value of δ is guaranteed to exist with $F(-\delta) \in [\frac{1}{4}, \frac{1}{2}]$ given our assumption that F is continuous and strictly increasing.)

The only remaining details are to prove Lemmas 9 and 10.

Proof of Lemma 9: The Gilbert-Varshamov bound (Gilbert, 1952; Varshamov, 1957) guarantees the existence of a collection of vectors $\{b^0, \ldots, b^{\bar{T}-1}\}$ contained with the Boolean hypercube $\{0,1\}^n$ such that $\bar{T} \geq 2^{n-1} \left(\sum_{\ell=0}^{\lceil \alpha n \rceil - 1} \binom{n-1}{\ell}\right)^{-1}$, and

$$d_{\mathrm{H}}(b^j, b^k) \geq \lceil \alpha n \rceil \qquad \text{for all } j \neq k,\, j, k \in [\bar{T} - 1].$$

Moreover, their construction allows loss of generality that the all-zeros vector is a member of the set—say $b^0 = 0$. We are thus guaranteed that $d_{\mathrm{H}}(b^j, 0) \geq \lceil \alpha n \rceil$ for all $j \in \{1, \ldots, \bar{T} - 1\}$.

Since $n \geq 2$ and $\alpha \in (0, \frac{1}{4})$, we have $\frac{\lceil \alpha n \rceil - 1}{n-1} \leq 2\alpha \leq \frac{1}{2}$. Applying standard bounds on the tail of the binomial distribution yields

$$\frac{1}{2^{n-1}} \sum_{\ell=0}^{\lceil \alpha n \rceil - 1} \binom{n-1}{\ell} \leq \exp\Big(-(n-1) D_{\mathrm{KL}}(\frac{\lceil \alpha n \rceil - 1}{n-1} \| \frac{1}{2})\Big) \leq \exp\Big(-(n-1) D_{\mathrm{KL}}(2\alpha \| \frac{1}{2})\Big).$$

Consequently, the number of non-zero code words $T := \bar{T} - 1$ is at least

$$T(\alpha) := \exp\Big((n-1) D_{\mathrm{KL}}(2\alpha \| \frac{1}{2})\Big) - 1.$$

Thus, the collection $\{b^1, \ldots, b^T\}$ has the desired properties.

Proof of Lemma 10: By definition of the matrix ensemble, we have

$$\|M(w^j) - M(w^k)\|_{\mathrm{F}}^2 = \sum_{u,v \in [n]} (F(\langle x_{uv},\, w^j \rangle) - F(\langle x_{uv},\, w^k \rangle))^2. \tag{33}$$

By construction, the Hamming distances between the triplet of vectors $\{w^j, w^k, 0\}$ are lower bounded $d_{\mathrm{H}}(w^j, 0) \geq \alpha n$, $d_{\mathrm{H}}(w^k, 0) \geq \alpha n$ and $d_{\mathrm{H}}(w^j, w^k) \geq \alpha n$. We claim that this implies that

$$\mathrm{card}\Big\{u \neq v \in [n]^2 \ \mid \langle x_{uv},\, w^j \rangle \neq \langle x_{uv},\, w^k \rangle\Big\} \geq \frac{\alpha^2}{4} n^2. \tag{34}$$

Taking this auxiliary claim as given for the moment, applying it to Equation (33) yields the lower bound $\|M(w^1) - M(w^2)\|_{\mathrm{F}}^2 \geq \frac{1}{4}\alpha^2 n^2 (F(\delta) - F(0))^2$, as claimed.

It remains to prove the auxiliary claim (34). We relabel $j = 1$ and $k = 2$ for simplicity in notation. For $(y, z) \in \{0,1\} \times \{0,1\}$, let set $\mathcal{I}_{yz} \subseteq [n]$ denote the set of indices on which w^1 takes value y and w^2 takes value z. We then split the proof into two cases:

Case 1: Suppose $|\ \mathcal{I}_{00} \cup \mathcal{I}_{11}\ | \geq \frac{\alpha n}{2}$. The minimum distance condition $d_{\mathrm{H}}(w^1, w^2) \geq \alpha n$ implies that $|\ \mathcal{I}_{01} \cup \mathcal{I}_{10}\ | \geq \alpha n$. For any $i \in \mathcal{I}_{00} \cup \mathcal{I}_{11}$ and any $j \in \mathcal{I}_{01} \cup \mathcal{I}_{10}$, it must be that $\langle x_{uv},\, w^1 \rangle \neq \langle x_{uv},\, w^2 \rangle$. Thus there are at least $\frac{\alpha^2}{2} n^2$ such pairs of indices.

Case 2: Otherwise, we may assume that $| \mathcal{I}_{00} \cup \mathcal{I}_{11} | < \frac{\alpha n}{2}$. This condition, along with the minimum Hamming weight conditions $d_{\text{H}}(w^1, 0) \geq \alpha n$ and $d_{\text{H}}(w^2, 0) \geq \alpha n$, gives $\mathcal{I}_{10} \geq \frac{\alpha n}{2}$ and $\mathcal{I}_{01} \geq \frac{\alpha n}{2}$. For any $i \in \mathcal{I}_{01}$ and any $j \in \mathcal{I}_{10}$, it must be that $\langle x_{uv}, w^1 \rangle \neq \langle x_{uv}, w^2 \rangle$. Thus there are at least $\frac{\alpha^2}{4} n^2$ such pairs of indices.

A.5.2. Upper bound

In our earlier work (Shah et al., 2016a, Theorem 2b) we prove that when F is strongly log-concave and twice differentiable, then there is a universal constant c_u such that the maximum likelihood estimator $\widehat{w}_{\text{ML}}$ has mean squared error at most

$$\sup_{w^* \in [-1,1]^n, \langle w^*, 1 \rangle = 0} \mathbb{E}[\|\widehat{w}_{\text{ML}} - w^*\|_2^2] \leq c_u. \tag{35}$$

Moreover, given the log-concavity assumption, the MLE is computable in polynomial-time. Let $M(\widehat{w}_{\text{ML}})$ and $M(w^*)$ denote the pairwise comparison matrices induced, via Equation (2), by $\widehat{w}_{\text{ML}}$ and w^*. It suffices to bound the Frobenius norm $|\!|\!| M(\widehat{w}_{\text{ML}}) - M(w^*) |\!|\!|_{\text{F}}$.

Consider any pair of vectors w^1 and w^2 that lie in the hypercube $[-1, 1]^n$. For any pair of indices $(i, j) \in [n]^2$, we have

$$((M(w^1))_{ij} - (M(w^2))_{ij})^2 = (F(w^1_i - w^1_j) - F(w^2_i - w^2_j))^2 \;\leq\; \zeta^2((w^1_i - w^1_j) - (w^2_i - w^2_j))^2,$$

where we have defined $\zeta := \max_{z \in [-1,1]} F'(z)$. Putting together the pieces yields

$$|\!|\!| M(w^1) - M(w^2) |\!|\!|_{\text{F}}^2 \leq \zeta^2 (w^1 - w^2)^T (nI - 11^T)(w^1 - w^2) \;=\; n\zeta^2 \|w^1 - w^2\|_2^2. \tag{36}$$

Applying this bound with $w^1 = \widehat{w}_{\text{ML}}$ and $w^2 = w^*$ and combining with the bound (35) yields the claim.

A.6. Proof of Theorem 5

We now turn to the proof of Theorem 5, which characterizes the behavior of different estimators for the partially observed case.

A.6.1. Proof of part (a)

In this section, we prove the lower and upper bounds stated in part (a).

Proof of lower bound: We begin by proving the lower bound. The Gilbert-Varshamov bound (Gilbert, 1952; Varshamov, 1957) guarantees the existence of a set of vectors $\{b^1, \ldots, b^T\}$ in the Boolean cube $\{0, 1\}^{\frac{n}{2}}$ with cardinality at least $T := 2^{cn}$ such that

$$d_{\text{H}}(b^j, b^k) \geq \lceil 0.1n \rceil \qquad \text{for all distinct pairs } j, k \in [T] := \{1, \ldots, T\}.$$

Fixing some $\delta \in (0, \frac{1}{4})$ whose value will be specified later, for each $k \in [T]$, we define a matrix $M^k \in \mathbb{C}_{\text{SST}}$ with entries

$$[M^k]_{uv} = \begin{cases} \frac{1}{2} + \delta & \text{if } u \leq \frac{n}{2}, [b^k]_u = 1 \text{ and } v \geq \frac{n}{2} \\ \frac{1}{2} & \text{otherwise,} \end{cases}$$

for every pair of indices $u \leq v$. We complete the matrix by setting $[M^k]_{vu} = 1 - [M^k]_{uv}$ for all indices $u > v$.

By construction, for each distinct pair $j, k \in [T]$, we have the lower bound

$$|\!|\!| M^j - M^k |\!|\!|_{\text{F}}^2 = n\delta^2 \|b^j - b^k\|_2^2 \;\geq\; c_0 n^2 \delta^2.$$

Let $\mathbb{P}^j$ and $\mathbb{P}^j_{uv}$ denote (respectively) the distributions of the matrix Y and entry Y_{uv} when the underlying matrix is M^j. Since the entries of Y are generated independently, we have $D_{\text{KL}}(\mathbb{P}^j \| \mathbb{P}^k) = \sum_{1 \leq u < v \leq n} D_{\text{KL}}(\mathbb{P}^j_{uv} \| \mathbb{P}^k_{uv})$. The matrix entry

Y_{uv} is generated according to the model

$$Y_{uv} = \begin{cases} 1 & \text{w.p. } p_{\rm obs} M^*_{uv} \\ 0 & \text{w.p. } p_{\rm obs}(1 - M^*_{uv}) \\ \text{not observed} & \text{w.p. } 1 - p_{\rm obs}. \end{cases}$$

Consequently, the KL divergence can be upper bounded as

$$\begin{aligned} D_{\rm KL}(\mathbb{P}^j_{uv} \| \mathbb{P}^k_{uv}) &= p_{\rm obs}\Big(M^j_{uv} \log \frac{M^j_{uv}}{M^k_{uv}} + (1 - M^j_{uv}) \log \frac{(1 - M^j_{uv})}{(1 - M^k_{uv})} \Big) \\ &\leq p_{\rm obs} \bigg\{ M^j_{uv} \big(\frac{M^j_{uv} - M^k_{uv}}{M^k_{uv}} \big) + (1 - M^j_{uv}) \big(\frac{M^k_{uv} - M^j_{uv}}{1 - M^k_{uv}} \big) \bigg\} \qquad (37) \\ &= p_{\rm obs} \ \frac{(M^j_{uv} - M^k_{uv})^2}{M^k_{uv}(1 - M^k_{uv})} \qquad (38) \\ &\leq 16 p_{\rm obs} \ (M^j_{uv} - M^k_{uv})^2, \qquad (39) \end{aligned}$$

where inequality (37) follows from the fact that $\log(t) \leq t - 1$ for all $t > 0$; and inequality (39) follows since the numbers $\{M^j_{uv}, M^k_{uv}\}$ both lie in the interval $[\frac{1}{4}, \frac{3}{4}]$. Putting together the pieces, we conclude that

$$D_{\rm KL}(\mathbb{P}^j \| \mathbb{P}^k) \leq c_1 p_{\rm obs} |\!|\!| M^j - M^k |\!|\!|^2_{\rm F} \ \leq \ c'_1 p_{\rm obs} n^2 \delta^2.$$

Thus, applying Fano's inequality to the packing set $\{M^1, \ldots, M^T\}$ yields that any estimator $\widehat{M}$ has mean squared error lower bounded by

$$\sup_{k \in [T]} \frac{1}{n^2} \mathbb{E}[|\!|\!| \widehat{M} - M^k |\!|\!|^2_{\rm F}] \geq c_0 \delta^2 \Big(1 - \frac{c'_1 p_{\rm obs} n^2 \delta^2 + \log 2}{cn} \Big).$$

Finally, choosing $\delta^2 = \frac{c_2}{2 c_1 p_{\rm obs} n}$ yields the lower bound $\sup_{k \in [T]} \frac{1}{n^2} \mathbb{E}[|\!|\!| \widehat{M} - M^k |\!|\!|^2_{\rm F}] \geq c_3 \frac{1}{n p_{\rm obs}}$. Note that in order to satisfy the condition $\delta \leq \frac{1}{4}$, we must have $p_{\rm obs} \geq \frac{16 c_2}{2 c_1 n}$.

Proof of upper bound: For this proof, recall the definition of matrix Y' from (9a). Observe that the matrix Y' can equivalently be written in a linearized form as

$$Y' = M^* + \frac{1}{p_{\rm obs}} W', \qquad (40a)$$

where W' has entries that are independent on and above the diagonal, satisfy skew-symmetry, and are distributed as

$$[W']_{ij} = \begin{cases} p_{\rm obs}(\frac{1}{2} - [M^*]_{ij}) + \frac{1}{2} & \text{with probability } p_{\rm obs}[M^*]_{ij} \\ p_{\rm obs}(\frac{1}{2} - [M^*]_{ij}) - \frac{1}{2} & \text{with probability } p_{\rm obs}(1 - [M^*]_{ij}) \\ p_{\rm obs}(\frac{1}{2} - [M^*]_{ij}) & \text{with probability } 1 - p_{\rm obs}. \end{cases} \qquad (40b)$$

We begin by introducing some additional notation. Letting Π denote the set of all permutations of n items. For each $\pi \in \Pi$, we define the set

$$\pi(\mathbb{C}_{\rm BISO}) := \big\{ M \in [0,1]^{n \times n} \mid M_{k\ell} \geq M_{ij} \text{ whenever } \pi(k) \leq \pi(i) \text{ and } \pi(\ell) \geq \pi(j) \big\},$$

corresponding to the subset of SST matrices that are faithful to the permutation π. We then define the estimator $M_\pi \in \arg\min_{M \in \pi(\mathbb{C}_{\rm BISO})} |\!|\!| Y' - M |\!|\!|^2_{\rm F}$, in terms of which the least squares estimator (9b) can be rewritten as

$$\widehat{M} \in \arg\min_{\pi \in \Pi} |\!|\!| Y' - M_\pi |\!|\!|^2_{\rm F}.$$

Define a set of permutations $\Pi' \subseteq \Pi$ as

$$\Pi' := \{\pi \in \Pi \mid |\!|\!|Y' - M_\pi|\!|\!|_{\mathrm{F}}^2 \le |\!|\!|Y' - M^*|\!|\!|_{\mathrm{F}}^2\}.$$

Note that the set Π' is guaranteed to be non-empty since the permutation corresponding to $\widehat{M}$ always lies in Π'. We claim that for any $\pi \in \Pi'$, we have

$$\mathbb{P}\big(|\!|\!|M_\pi - M^*|\!|\!|_{\mathrm{F}}^2 \le c_u \frac{n}{p_{\mathrm{obs}}} \log^2 n\big) \ge 1 - e^{-3n\log n}, \tag{41}$$

for some positive universal constant c_u. Given this bound, since there are at most $e^{n \log n}$ permutations in the set Π', a union bound over all these permutations applied to (41) yields

$$\mathbb{P}\Big(\max_{\pi \in \Pi'} |\!|\!|M_\pi - M^*|\!|\!|_{\mathrm{F}}^2 > c_u \frac{n}{p_{\mathrm{obs}}} \log^2 n\Big) \le e^{-2n\log n}.$$

Since $\widehat{M}$ is equal to M_π for some $\pi \in \Pi'$, this tail bound yields the claimed result.

The remainder of our proof is devoted to proving the bound (41). By definition, any permutation $\pi \in \Pi'$ must satisfy the inequality

$$|\!|\!|Y - M_\pi|\!|\!|_{\mathrm{F}}^2 \le |\!|\!|Y - M^*|\!|\!|_{\mathrm{F}}^2.$$

Denote the error in the estimate as $\widehat{\Delta}_\pi := M_\pi - M^*$. Starting with the least-squares objective (9a), the optimality of M_π and feasibility M^* and some algebraic manipulations then lead to the basic inequality

$$\frac{1}{2}|\!|\!|\widehat{\Delta}_\pi|\!|\!|_{\mathrm{F}}^2 \le \frac{1}{p_{\mathrm{obs}}}\langle\!\langle W', \ \widehat{\Delta}_\pi\rangle\!\rangle. \tag{42}$$

Now consider the set of matrices

$$\mathbb{C}_{\mathrm{DIFF}}(\pi) := \Big\{\alpha(M - M^*) \mid M \in \pi(\mathbb{C}_{\mathrm{BISO}}), \ \alpha \in [0,1]\Big\}, \tag{43}$$

and note that $\mathbb{C}_{\mathrm{DIFF}}(\pi) \subseteq [-1,1]^{n\times n}$. (To be clear, the set $\mathbb{C}_{\mathrm{DIFF}}(\pi)$ also depends on the value of M^*, but considering M^* as fixed, we omit this dependence from the notation for brevity.) For each choice of radius $t > 0$, define the random variable

$$Z_\pi(t) := \sup_{\substack{D \in \mathbb{C}_{\mathrm{DIFF}}(\pi), \\ |\!|\!|D|\!|\!|_{\mathrm{F}} \le t}} \frac{1}{p_{\mathrm{obs}}}\langle\!\langle D, \ W'\rangle\!\rangle. \tag{44}$$

Using the basic inequality (42), the Frobenius norm error $|\!|\!|\widehat{\Delta}_\pi|\!|\!|_{\mathrm{F}}$ then satisfies the bound

$$\frac{1}{2}|\!|\!|\widehat{\Delta}_\pi|\!|\!|_{\mathrm{F}}^2 \le \frac{1}{p_{\mathrm{obs}}}\langle\!\langle W', \ \widehat{\Delta}_\pi\rangle\!\rangle \ \le \ Z_\pi\big(|\!|\!|\widehat{\Delta}_\pi|\!|\!|_{\mathrm{F}}\big). \tag{45}$$

Thus, in order to obtain a high probability bound, we need to understand the behavior of the random quantity $Z_\pi(t)$.

One can verify that the set $\mathbb{C}_{\mathrm{DIFF}}(\pi)$ is star-shaped, meaning that $\alpha D \in \mathbb{C}_{\mathrm{DIFF}}(\pi)$ for every $\alpha \in [0,1]$ and every $D \in \mathbb{C}_{\mathrm{DIFF}}(\pi)$. Using this star-shaped property, we are guaranteed that there is a non-empty set of scalars $\delta_n > 0$ satisfying the critical inequality

$$\mathbb{E}[Z_\pi(\delta_n)] \le \frac{\delta_n^2}{2}. \tag{46}$$

Our interest is in an upper bound to the smallest (strictly) positive solution δ_n to the critical inequality (46), and moreover, our goal is to show that for every $t \ge \delta_n$, we have $|\!|\!|\widehat{\Delta}|\!|\!|_{\mathrm{F}} \le c\sqrt{t\delta_n}$ with high probability.

Define a "bad" event $\mathcal{A}_t$ as

$$\mathcal{A}_t = \{\exists \Delta \in \mathbb{C}_{\mathrm{DIFF}}(\pi) \mid |\!|\!|\Delta|\!|\!|_{\mathrm{F}} \ge \sqrt{t\delta_n} \quad \text{and} \quad \frac{1}{p_{\mathrm{obs}}}\langle\!\langle \Delta, \ W'\rangle\!\rangle \ge 2|\!|\!|\Delta|\!|\!|_{\mathrm{F}}\sqrt{t\delta_n}\}. \tag{47}$$

Using the star-shaped property of $\mathbb{C}_{\mathrm{DIFF}}(\pi)$, it follows by a rescaling argument that

$$\mathbb{P}[\mathcal{A}_t] \le \mathbb{P}[Z_\pi(\delta_n) \ge 2\delta_n\sqrt{t\delta_n}] \qquad \text{for all } t \ge \delta_n.$$

The following lemma helps control the behavior of the random variable $Z_\pi(\delta_n)$.

Lemma 11. *For any $\delta > 0$, the mean of $Z_\pi(\delta)$ is bounded as*

$$\mathbb{E}[Z_\pi(\delta)] \leq c_u \frac{n}{p_{\mathrm{obs}}} \log^2 n,$$

and for every $u > 0$, its tail probability is bounded as

$$\mathbb{P}\Big(Z_\pi(\delta) > \mathbb{E}[Z_\pi(\delta)] + u\Big) \leq \exp\Big(\frac{-cu^2 p_{\mathrm{obs}}}{\delta^2 + \mathbb{E}[Z_\pi(\delta)] + u}\Big),$$

where c_u and c are positive universal constants.

From this lemma, we have the tail bound

$$\mathbb{P}\Big(Z_\pi(\delta_n) > \mathbb{E}[Z_\pi(\delta_n)] + \delta_n\sqrt{t\delta_n}\Big) \leq \exp\Big(\frac{-c(\delta_n\sqrt{t\delta_n})^2 p_{\mathrm{obs}}}{\delta_n^2 + \mathbb{E}[Z_\pi(\delta_n)] + (\delta_n\sqrt{t\delta_n})}\Big), \quad \text{for all } t \geq \delta_n.$$

By the definition of δ_n in (46), we have $\mathbb{E}[Z(\delta_n)] \leq \delta_n^2 \leq \delta_n\sqrt{t\delta_n}$ for any $t \geq \delta_n$, and consequently

$$\mathbb{P}[\mathcal{A}_t] \leq \mathbb{P}[Z(\delta_n) \geq 2\delta_n\sqrt{t\delta_n}] \;\leq\; \exp\Big(\frac{-c(\delta_n\sqrt{t\delta_n})^2 p_{\mathrm{obs}}}{3\delta_n\sqrt{t\delta_n}}\Big), \quad \text{for all } t \geq \delta_n.$$

Consequently, either $|\!|\!|\widehat{\Delta}_\pi|\!|\!|_{\mathrm{F}} \leq \sqrt{t\delta_n}$, or we have $|\!|\!|\widehat{\Delta}_\pi|\!|\!|_{\mathrm{F}} > \sqrt{t\delta_n}$. In the latter case, conditioning on the complement $\mathcal{A}_t^c$, our basic inequality implies that $\frac{1}{2}|\!|\!|\widehat{\Delta}_\pi|\!|\!|_{\mathrm{F}}^2 \leq 2|\!|\!|\widehat{\Delta}_\pi|\!|\!|_{\mathrm{F}}\sqrt{t\delta_n}$ and hence $|\!|\!|\widehat{\Delta}_\pi|\!|\!|_{\mathrm{F}} \leq 4\sqrt{t\delta_n}$. Putting together the pieces yields that

$$\mathbb{P}\big(|\!|\!|\widehat{\Delta}_\pi|\!|\!|_{\mathrm{F}} \leq 4\sqrt{t\delta_n}\big) \geq 1 - \exp\big(-c'\delta_n\sqrt{t\delta_n}p_{\mathrm{obs}}\big), \quad \text{for all } t \geq \delta_n. \tag{48}$$

Finally, from the bound on the expected value of $Z_\pi(t)$ in Lemma 11, we see that the critical inequality (46) is satisfied for $\delta_n = \sqrt{\frac{c_u n}{p_{\mathrm{obs}}}}\log n$. Setting $t = \delta_n = \sqrt{\frac{c_u n}{p_{\mathrm{obs}}}}\log n$ in (48) yields

$$\mathbb{P}\Big(|\!|\!|\widehat{\Delta}_\pi|\!|\!|_{\mathrm{F}} \leq 4\frac{c_u n}{p_{\mathrm{obs}}}\log^2 n\Big) \geq 1 - \exp\Big(-3n\log n\Big), \tag{49}$$

for some universal constant $c_u > 0$, thus proving the bound (41).

It remains to prove Lemma 11.

Proof of Lemma 11 Bounding $\mathbb{E}[Z_\pi(\delta)]$: We establish an upper bound on $\mathbb{E}[Z_\pi(\delta)]$ by using Dudley's entropy integral, as well as some auxiliary results on metric entropy. We use the notation $\log N(\epsilon, \mathbb{C}, \rho)$ to denote the ϵ metric entropy of the class $\mathbb{C}$ in the metric ρ. Introducing the random variable $\widetilde{Z}_\pi := \sup_{D \in \mathbb{C}_{\text{DIFF}}(\pi)} \langle\!\langle D,\; W'\rangle\!\rangle$, note that we have $\mathbb{E}[Z_\pi(\delta)] \leq \frac{1}{p_{\mathrm{obs}}}\mathbb{E}[\widetilde{Z}_\pi]$. The truncated form of Dudley's entropy integral inequality yields

$$\mathbb{E}[\widetilde{Z}_\pi] \leq c\Big\{n^{-8} + \int_{\frac{1}{2}n^{-9}}^{2n} \sqrt{\log N(\epsilon, \mathbb{C}_{\text{DIFF}}(\pi), |\!|\!|.|\!|\!|_{\mathrm{F}})}d\epsilon\Big\}, \tag{50}$$

where we have used the fact that the diameter of the set $\mathbb{C}_{\text{DIFF}}(\pi)$ is at most $2n$ in the Frobenius norm.

From our earlier bound (21) earlier, we are guaranteed that for each $\epsilon > 0$, the metric entropy is upper bounded as

$$\log N\Big(\epsilon, \{\alpha M \mid M \in \mathbb{C}_{\text{BISO}}, \alpha \in [0,1]\}, |\!|\!|\cdot|\!|\!|_{\mathrm{F}}\Big) \leq 8\frac{n^2}{\epsilon^2}\big(\log\frac{n}{\epsilon}\big)^2.$$

Consequently, we have $\log N(\epsilon, \mathbb{C}_{\text{DIFF}}(\pi), |\!|\!|.|\!|\!|_{\mathrm{F}}) \leq 16\frac{n^2}{\epsilon^2}\big(\log\frac{n}{\epsilon}\big)^2$. Substituting this bound on the metric entropy of $\mathbb{C}_{\text{DIFF}}(\pi)$ and the inequality $\epsilon \geq \frac{1}{2}n^{-9}$ into the Dudley bound (50) yields

$$\mathbb{E}[\widetilde{Z}_\pi] \leq cn(\log n)^2.$$

The inequality $\mathbb{E}[Z_\pi(\delta)] \leq \frac{1}{p_{\text{obs}}}\mathbb{E}[\widetilde{Z}_\pi]$ then yields the claimed result.

Bounding the tail probability of $Z_\pi(\delta)$: In order to establish the claimed tail bound, we use a Bernstein-type bound on the supremum of empirical processes due to Klein et al. (2005, Theorem 1.1c), which we state in a simplified form here.

Lemma 12. *Let $X := (X_1, \ldots, X_m)$ be any sequence of zero-mean, independent random variables, each taking values in $[-1, 1]$. Let $\mathcal{V} \subset [-1, 1]^m$ be any measurable set of m-length vectors. Then for any $u > 0$, the supremum $X^\dagger = \sup_{v \in \mathcal{V}} \langle X,\, v\rangle$ satisfies the upper tail bound*

$$\mathbb{P}\big(X^\dagger > \mathbb{E}[X^\dagger] + u\big) \leq \exp\Big(\frac{-u^2}{2\sup_{v \in \mathcal{V}} \mathbb{E}[\langle v,\, X\rangle^2] + 4\mathbb{E}[X^\dagger] + 3u}\Big).$$

We now call upon Lemma 12 setting $\mathcal{V} = \mathbb{C}_{\text{DIFF}}(\pi) \cap \mathbb{B}(\delta)$, $m = (n \times n)$, $X = W'$, and $X^\dagger = p_{\text{obs}} Z_\pi(\delta)$. The entries of the matrix W' have a mean of zero, are bounded by 1 in absolute value, and are independent on and above the diagonal (with the entries below the diagonal being deterministic functions of their counterparts above their diagonal). Then we have $\mathbb{E}[X^\dagger] \leq p_{\text{obs}}\mathbb{E}[Z_\pi(\delta)]$ and $\mathbb{E}[\langle\!\langle D,\ W'\rangle\!\rangle^2] \leq 4p_{\text{obs}}|\!|\!|D|\!|\!|_{\text{F}}^2 \leq 4p_{\text{obs}}\delta^2$ for every $D \in \mathcal{V}$. With these assignments, and some algebraic manipulations, we obtain that for every $u > 0$,

$$\mathbb{P}\big(Z_\pi(\delta) > \mathbb{E}[Z_\pi(\delta)] + u\big) \leq \exp\Big(\frac{-u^2 p_{\text{obs}}}{8\delta^2 + 4\mathbb{E}[Z_\pi(\delta)] + 3u}\Big),$$

as claimed.

A.6.2. Proof of part (b)

In order to prove the claimed bound, we analyze the SVT estimator $T_{\lambda_{n,p_{\text{obs}}}}(Y')$ with the threshold $\lambda_{n,p_{\text{obs}}} = \frac{2.1\sqrt{n}}{p_{\text{obs}}}$. Naturally then, our analysis is similar to that of complete observations case from Section A.3. Recall our formulation of the problem in terms of the observation matrix Y' along with the noise matrix W' from equations (9a), (40a) and (40b). The result of Lemma 4 continues to hold in this case of partial observations, translated for this setting as: If $\lambda_{n,p_{\text{obs}}} \geq \frac{1.01}{p_{\text{obs}}}|\!|\!|W|\!|\!|_{\text{op}}$, then

$$|\!|\!|T_{\lambda_{n,p_{\text{obs}}}}(Y') - M^*|\!|\!|_{\text{F}}^2 \leq c\sum_{j=1}^{n} \min\big\{\lambda_{n,p_{\text{obs}}}^2, \sigma_j^2(M^*)\big\}$$

with probability at least $1 - c_1 e^{-c'n}$, where c, c_1 and c' are positive universal constants.

The entries of W' are bounded by 1 in absolute value, are independent on and above the diagonal, are zero-mean, and satisfy skew-symmetry. As before, Theorem 3.4 of Chatterjee (2014) then implies that

$$\mathbb{P}\Big[|\!|\!|W'|\!|\!|_{\text{op}} > (2+t)\sqrt{n}\Big] \leq c_1 e^{-f(t)n}$$

where c_1 is a universal constant and $f(t)$ is strictly positive for each $t > 0$. With our choice $\lambda_{n,p_{\text{obs}}} = \frac{2.1\sqrt{n}}{p_{\text{obs}}}$, the event $\{\lambda_{n,p_{\text{obs}}} \geq \frac{1.01}{p_{\text{obs}}}|\!|\!|W'|\!|\!|_{\text{op}}\}$ holds with probability at least $1 - c_1 e^{-c_2 n}$. Conditioned on this event, the approximation-theoretic result from Lemma 5 gives

$$\frac{1}{n^2}|\!|\!|T_{\lambda_{n,p_{\text{obs}}}}(Y') - M^*|\!|\!|_{\text{F}}^2 \leq c\Big(\frac{s\lambda_{n,p_{\text{obs}}}^2}{n^2} + \frac{1}{s}\Big)$$

with probability at least $1 - c_1 e^{-c_2 n}$. Substituting $\lambda_{n,p_{\text{obs}}} = \frac{2.1\sqrt{n}}{p_{\text{obs}}}$ in this bound and setting $s = p_{\text{obs}}\sqrt{n}$ gives the claimed result.

A.6.3. Proof of part (c)

As in our of proof of the fully observed case from Section A.5.2, we consider the two-stage estimator based on first computing the MLE $\widehat{w}_{\text{ML}}$ of w^* from the observed data, and then constructing the matrix estimate $M(\widehat{w}_{\text{ML}})$ via Equation (2). Let us now upper bound the mean-squared error associated with this estimator.

Our observation model can be (re)described in the following way. Consider an Erdős-Rényi graph on n vertices with each edge drawn independently with a probability $p_{\rm obs}$. For each edge in this graph, we obtain one observation of the pair of vertices at the end-points of that edge. Let L be the (random) Laplacian matrix of this graph, that is, $L = D - A$ where D is an $(n \times n)$ diagonal matrix with $[D]_{ii}$ being the degree of item i in the graph (equivalently, the number of pairwise comparison observations that involve item i) and A is the $(n \times n)$ adjacency matrix of the graph. Let $\lambda_2(L)$ denote the second largest eigenvalue of L. From Theorem 2(b) of our paper (Shah et al., 2016a) on estimating parametric models,[4] for this graph, there is a universal constant c_1 such that the maximum likelihood estimator $\widehat{w}_{\rm ML}$ has mean squared error upper bounded as

$$\mathbb{E}[\|\widehat{w}_{\rm ML} - w^*\|_2^2 \mid L] \leq c_1 \frac{n}{\lambda_2(L)}.$$

The estimator $\widehat{w}_{\rm ML}$ is computable in a time polynomial in n.

Since $p_{\rm obs} \geq c_0 \frac{(\log n)^2}{n}$, known results on the eigenvalues of random graphs (Oliveira, 2009; Chung & Radcliffe, 2011; Kolokolnikov et al., 2014) imply that

$$\mathbb{P}\Big[\lambda_2(L) \geq c_2 n p_{\rm obs}\Big] \geq 1 - \frac{1}{n^4} \tag{51}$$

for a universal constant c_2 (that may depend on c_0). As shown earlier in Equation (36), for any valid score vectors w^1, w^2, we have $|\!|\!| M(w^1) - M(w^2) |\!|\!|_{\rm F}^2 \leq n\zeta^2 \|w^1 - w^2\|_2^2$ where $\zeta := \max_{z \in [-1,1]} F'(z)$ is a constant independent of n and $p_{\rm obs}$. Putting these results together and performing some simple algebraic manipulations leads to the upper bound

$$\frac{1}{n^2}\mathbb{E}\Big[|\!|\!| M(\widehat{w}_{\rm ML}) - M^* |\!|\!|_{\rm F}^2\Big] \leq \frac{c_3 \zeta^2}{n p_{\rm obs}},$$

which establishes the claim.

[4]Note that the Laplacian matrix used in the statement of (Shah et al., 2016a, Theorem 2(b)) is a scaled version of the matrix L introduced here, with each entry of L divided by the total number of observations.