Near-Optimal Joint Object Matching via Convex Relaxation
— Supplemental Materials —

Abstract

This supplemental document presents details concerning (1) summary of changes in this new submission and response to reviewers' comments in Cycle 1; (2) analytical derivations that support the theorems made in the main text "Near-optimal Joint Object Matching via Convex Relaxation", submitted to the 31th International Conference on Machine Learning (ICML 2014). One can find here the detailed proof of main theorems.

1 Response to Reviewer's Comments in Cycle 1

Dear Meta Reviewer and Reviewers 1, 6, and 7,

We appreciate your constructive comments, several of which have significantly changed the technical contents, methodologies, and positioning of this paper. In this new submission, we have obtained a more effective convex formulation that eventually allows us to substantially improve our performance guarantees. Below we briefly summarize the improvement, and address reviewer's major comments in Cycle 1. Your comments are in *italics* with our response following in plaintext.

1.1 Major Improvements

1. **Algorithms**. We have derived a more effective convex formulation. Denote by m the size of universe (i.e. the number of unique elements to be matched), which is usually unknown when matching partially similar objects. Our new formulation, however, is based on the observation that m can often be pre-estimated by spectral techniques even under dense errors. With this information at hand, we are able to derive a strengthened relaxation as follows:

$$\begin{aligned} \text{(MatchLift)} \quad \underset{\boldsymbol{X}}{\text{minimize}} \quad & -\sum_{(i,j)\in\mathcal{G}} \left\langle \boldsymbol{X}_{ij}^{\text{in}}, \boldsymbol{X}_{ij} \right\rangle + \lambda \left\langle \mathbf{1}\cdot\mathbf{1}^{\top}, \boldsymbol{X} \right\rangle \\ \text{subject to} \quad & \boldsymbol{X} \geq \mathbf{0}, \\ & \begin{bmatrix} m & \mathbf{1}^{\top} \\ \mathbf{1} & \boldsymbol{X} \end{bmatrix} \succeq \mathbf{0}, \qquad (1) \\ & \boldsymbol{X}_{ii} = \boldsymbol{I}, \qquad (2) \end{aligned}$$

   where λ can be set in a parameter-free manner as $\frac{\sqrt{|\mathcal{E}|}}{2n}$. The most crucial changes are the strenghened PSD constraints (1) that depends on m (compared with $\boldsymbol{X} \succeq \mathbf{0}$ in the initial submission), which turns out to be critical in enabling dense error correction. Intuitively, this strengthened constraint presents us one additional degree of freedom for outlier separation, which often serves to "debias" outliers.

2. **Theory (Dense Error Correction)**. With the tightened convex relaxation, our new algorithm (called MatchLift) significantly outperforms the originally proposed method. Most remarkably, MatchLift enables *dense error correction*, i.e. the algorithm is guaranteed to work when the portion p_{true} of correct inputs exceeds

$$p_{\text{true}} = \Omega\left(\frac{1}{\sqrt{n}}\right).$$

That said, in the asymptotic regime, the algorithms works even when almost of the inputs are badly corrupted by random errors. In contrast, none of the existing algorithms can provably separate more than 50% errors.

3. **Experiments**. We have performed substantially more experiments to verify the practical applicability of our algorithm. First, we compare our algorithm with Jalali et al [1] in synthetic data, where we compare 31×36 sets of different parameters for each of the two scenarios with each parameter configuration simulated by 10 Monte Carlo trials. Second, we not only run the algorithms on the chair and building data sets, but also on the CMU Housing and Hotel data sets as suggested by the reviewer. We have also compared our algorithm against Jalali et al 2011 and two other state-of-the-art graph matching algorithms on CMU Housing and Hotel. All of these experiments confirm the superiority of our algorithm in the joint matching problem.

1.2 Response to Meta Reviewer

We thank the meta reviewer for his/her suggestion on how to improve the positioning of our paper. We have addressed your 3 comments as below.

(1) The authors should make the relationship to clustering problems clear.

We have now discussed the similarity and distinction between the matching problem and the graph clustering problem. In particular, we have highlighted in several paragraphs the following arguments:

- The joint matching problem can be treated as a structured graph clustering (GC) problem, where graph nodes represent points on objects and the edge set encodes all correspondences. In this regard, any GC algorithm provides a heuristic to estimate graph matching. Consequently, the size of the universe m in our case corresponds to the number of clusters in the clustering problem, which allows us to immediately see that $\mathrm{rank}(\boldsymbol{X}) = m$.
- Nevertheless, there are several intrinsic structural properties herein that are not explored by any generic GC approach.
  - First, our input takes a block-matrix form, where each block is highly *structured* (i.e. doubly-substochastic), *sparse*, and *inter-dependent*.
  - Second, the points belonging to the same object are *mutually exclusive* to each other.
  - Third, the corruption rate for different entries can be highly non-symmetric – when translated into GC languages, this means that in-cluster edges might suffer from an order-of-magnitude larger error rate than inter-cluster edges.

(2) The authors should explicitly show the benefit of imposing the additional constraint, in terms of any theoretical guarantees they develop (i.e. contrast the guarantees explicitly with unconstrained clustering guarantees and show the benefit), and in terms of the empirical evaluation.

Our new formulation, which incorporates the feature that points within each set are mutually exclusive, generates significantly better guarantees than generic GC approaches, both theoretically and empirically.

- **Theory**. Various approaches for general graph clustering have been proposed with theoretical guarantees under different randomized settings, which typically operate under the assumptions that in-cluster and inter-cluster correspondences are independently corrupted, which differs drastically from our model. Due to the block structure input model, these two types of corruptions are highly correlated and usually experience order-of-magnitude difference in corruption rate. To facilitate comparison, we evaluate the *best-known deterministic guarantees* obtained by Jalali et al [2]. The key metric $D_{\max}$ therein can be easily bounded by $D_{\max} \geq 1 - p_{\text{true}}$ due to significant in-cluster edge errors. The recovery condition in [2, Theorem 1] requires $D_{\max} < \frac{1}{m+1}$, and thereby $p_{\text{true}} > \frac{m}{m+1}$, which is at least $\frac{1}{3}$ (in fact, typically far worse than $\frac{1}{3}$) even when $m = 2$. In comparison, our results allow the non-corruption rate p_{true} to be ***vanishingly small*** (i.e. $p_{\text{true}} = \Omega\left(\frac{\log^2 n}{\sqrt{n}}\right)$), which is significantly better than [2]. This indicates that generic GC algorithms do not deliver informative guarantees when tailored to our problem.

- **Experiment**. We have compared our algorithm with [1] on both synthetic data (with extensive experiments characterizing the phase transition diagrams) and real-world images, as reported in Section 5 of the main body of the paper. One can see that the GC approach [1] is limited to a small constant barrier, while MatchLift enables dense error correction as n grows. The performance of MatchLift on real-world images also significantly outperforms GC approach [1].

(3) They should demonstrate the relevance and importance of the formulation discussed.

Except for its distinction and relevance to the clustering problem that we detail above, we would like to highlight the most important feature of our new algorithm, which is dense error correction. None of the prior algorithm have demonstrated performance guarantees when there are more than 50% errors present, regardless of how many objects are available. Encouragingly, our algorithm, with the novel constraint we include, significantly improves our algorithm to enable dense outlier separation. This is very important in practice and highlights the importance of joint object matching – that said, no matter how bad /noisy the input sources are, as long as we have sufficiently many instances, we will be able to obtain perfect matching over all pairs of instances.

1.3 Response to Reviewer 1

We appreciate the reviewer's helpful comments, especially for singling out our lack of comparison with graph clustering (GC). We have now substantially changed our algorithm and completely rederived the performance guarantees (which significantly outperform those derived for the original formulation). Consequently, several of the reviewer's comments are no longer relevant to the current formulation. Below we address those comments that apply to the current formulation.

1. Problem formulation: The formulation is the standard SDP relaxation of Correlation Clustering, where the idea of dropping the rank-1 and binary constraints dates back to Goemans-Williamson. It is also well-known that one can improve performance by adding to the convex program any convex constraints that represent prior structural knowledge, such as must-links (Yi-Zhang-Jin-Qian-Jain, ICML 2013) and cluster sizes (Ames 2011); in the present case it is the sub-stochastic constraints arising from the known structures encoded in the S_i's. More specifically for the matching problem, similar convex relaxation approach has been proposed in the literature (Huang-Guibas 2013). Given this long line of prior work, the contribution in this paper seems incremental.

Thanks much for pointing our our lack of comparison with clustering problem. We have now derived a new theorem based on a new conic constraint

$$\begin{bmatrix} m & \mathbf{1}^\top \\ \mathbf{1} & \boldsymbol{X} \end{bmatrix} \succeq \mathbf{0},$$

where the knowledge of m is pre-estimated by spectral methods. It turns out that this constraint, together with the mutually exclusive constraint $\boldsymbol{X}_{ii} = \boldsymbol{I}$, allows us to significantly improve the error-correction ability, both theoretically and empirically. In the presence of this tightened conic constraint, the improvement incurred by sub-stochastic constraints becomes marginal, provided that m can be accurately estimated. However, if m cannot be reliably estimated, then the sub-stochastic constraints will be critical in guaranteeing perfect recovery with a constant portion (e.g. 50%) of input errors, although dense error correction might not be guaranteed.

2. Algorithm: It is also standard practice in graph clustering/partitioning to solve the convex relaxation via first order methods and do a heuristic rounding of the optimal solution. This paper uses ADMM and a particular rounding heuristic, with no further justification for their choices. There are many other possible options, some of which come with rigorous theoretical guarantees (e.g. Mathieu-Schudy).

ADMM formulation has been shown to converge for semidefinite program, which has now been very popular. Most importantly, ADMM usually converges to modest accuracy within a reasonable amount of time, and produces desired results with the assistance of appropriate rounding procedures. This is particularly appealing in our case, whereby the ground truth is a 0-1 matrix. Empirically, ADMM generates desired results in practice and converges for all our experiments. In terms of being rigorous, applying provably accurate interior point methods like MOSEK for our formulation leads to worst case complexity $(nm)^6$, while that of Mathieu *et al* exceeds $(nm)^9$. Practically, we have not been able to let Mathieu *et al* work except for very small dimentionality. This computational limitation was also reported in Jalali *et al* [1].

3.1 It is somewhat difficult to interpret their results.

Thanks for pointing out the confusing part of our previous theorem. We have derived completely new guarantees, which we hope are sufficiently clean to interpret.

3.2 The authors do not discuss if the additional complexities in their formulation lead to any improvements over standard graph clustering algorithms. For example, if we assume p_set, m and p_{false} are all O(1), then their results guarantee success when $p_{\text{obs}} > 1/n$. Modulo some mild differences in the generative models, the same performance can be achieved via existing graph clustering methods (e.g. Jalali et al 2011) without using any additional structures of the matching problem. While this does not exclude the possibility that the proposed method have better performance in other settings, there is no discussion of this issue in the current paper.

Thanks for bringing to our attention the graph clustering algorithms. We note that prior work on randomized model does not deliver guarantee for our problem, as our inputs are highly structured and the errors are mutually dependent. Most importantly, the error rate for in-cluster edges (which can be as worse as $1-\frac{1}{\sqrt{n}}$) and for inter-cluster edges (which is $\frac{1}{m}$) are drastically different. If we apply the most recent deterministic guarantees like [2, Theorem 1], the recovery condition therein requires $D_{\max} < \frac{1}{m+1}$, and thereby $p_{\text{true}} > \frac{m}{m+1}$, which is at best $\frac{1}{3}$ (usually far worse than $\frac{1}{3}$) even when $m = 2$. In contrast, our results allow the *non-corruption rate* p_{true} to be ***vanishingly small*** (i.e. $p_{\text{true}} = \Omega\left(\frac{\log^2 n}{\sqrt{n}}\right)$), which significantly outperforms [2]. This indicates that generic GC algorithms do not deliver informative guarantees when tailored to our problem.

4. Implementation: Convex optimization approaches often suffer from scalability issues. The proposed approach has many additional constraints, which might further degrade scalability. Their experiments use small datasets with 24 elements and 32 sets, which is not quite convincing.

In our new examples, we perform extensive experiments (in order to plot phase transition diagram) when the matrix is as large as 2000×2000. In practice, our ADMM code can solve problems of size 6000×6000 on a regular computer. A more scalable solution is indeed very interesting and important future work.

5(1) Any deterministic guarantee for GC (e.g. Jalali-Srebro 2012) would imply a guarantee that can be compared with Theorem 3. It would be helpful to show that incorporating the additional sub-stochastic constraints indeed leads to stronger guarantees than general GC methods without these constraints.

See the response to comment 3.2.

5(2) There are also some theories in Huang-Guibas's paper. The authors may want to discuss how hard to extend those results to the partial matching setting.

We have now substantially changed our paper, which now demonstrates an ability to correct arbitrarily dense errors. In comparison, Huang-Guibas only reported theoretical support when there are fewer than 50% errors present. Both the results and analyses are drastically different from Huang-Guibas.

6. The paper has "matrix completion" in its title, while their method is only marginally related to matrix completion (for which the standard approach is nuclear norm minimization) but is actually more in line with standard SDP relaxations for graph partitioning.

Indeed, our original title on "matrix completion" is not precise enough. We have now removed "matrix completion" from our title – thanks for the suggestion.

1.4 Response to Reviewer 6

The approach is neat and elegant. My only concern is the experimental evaluation. The dataset considered does not seem to me a challenging benchmark, and more extensive evaluation would be valuable. For instance, *the authors should benchmark the approach on the CMU House and Hotel dataset. Also, the evaluation measure is questionable as it is sensitive to denser matches. The experiments should include comparison using more standard evaluation measures; see eg [3] for evaluation measures to be considered.*

Thanks much for helping us improve our empirical experiments. In additional to several tens of thousands more experiments on synthetic data, we have applied our new method to more benchmark datasets including CMU House and Hotel datasets (as well as the Graf and Bikes datasets from http://www.robots.ox.ac.uk/~vgg/research/affine) as suggested by the reviewers. The detailed comparison and empirical performance are detailed in Section 6. Also, thanks for the suggestion on more standard evaluation measures. We have now applied the metric described in HaCohen et al [3], which evaluates the deviations of manual feature correspondences.

1.5 Response to Reviewer 7

1. It is not clear to me that Fig.2 really corresponds to the numbers stated in the text next to the figure.

We appreciate your comment in pointing out the inconsistency – these were indeed typos. In the current submission, we have derived new (and significantly better) algorithms, and hence these plots for the old methodologies are no longer included in the current version.

2. The experiments on real data are fine, but only two images sets seem quite few. In the context of local descriptor evaluation, there exists at least one dataset that provides images of the same structure seen under different viewing conditions: http://www.robots.ox.ac.uk/~vgg/research/affine. It would be nice if some additional experiments could be performed on this data.

Thanks for pointing out the lack of experimental evaluation in our initial evaluation. According to the reviewers' suggestion, we have added four benchmark data sets: (the Graf and Bikes datasets from http://www.robots.ox.ac.uk/~vgg/research/affine, as well as CMU Housing and Hotel), and compared our algorithm against the best-known graph clustering and graph matching algorithms on these real-world data sets as well as synthetic data. All of them confirm the practical ability of MatchLift, which outperforms all other algorithms. Details can be found in Section 6.

3. The idea follows the work of Huang and Guibas, but is novel enough.

We have now derived completely new formulation, which is significantly different from Huang and Guibas. Our theoretical guarantees, which allow dense error correction, also significantly outperforms those reported in Huang and Guibas (which requires that no more than 50% errors are present).

2 Notation and Convention

Let $\mathcal{I}_i$ ($1 \leq i \leq m$) denote the index set of the shapes containing points i, and let $n_i := |\mathcal{I}_i|$ ($1 \leq i \leq m$) represent the number of shapes containing point i. Define $N := n_1 + \cdots + n_m$. We also set $[m] := \{1, 2, \cdots, m\}$.

Some useful notation is summarized in Table 1.

3 Alternating Direction Method of Multipliers (ADMM)

Recall that our algorithm is given by

$$\begin{aligned}
\text{(MatchLift)} \quad \underset{\boldsymbol{X}}{\text{minimize}} \quad & -\sum_{(i,j)\in\mathcal{G}} \left\langle \boldsymbol{X}_{ij}^{\text{in}}, \boldsymbol{X}_{ij} \right\rangle + \lambda \left\langle \mathbf{1}\cdot\mathbf{1}^{\top}, \boldsymbol{X} \right\rangle \\
\text{subject to} \quad & \boldsymbol{X} \geq \mathbf{0}, \\
& \begin{bmatrix} m & \mathbf{1}^{\top} \\ \mathbf{1} & \boldsymbol{X} \end{bmatrix} \succeq \mathbf{0}, && (3) \\
& \boldsymbol{X}_{ii} = \boldsymbol{I}, && (4)
\end{aligned}$$

where $m_i := |\mathcal{S}_i|$.

Symbol	Description
$\mathbf{1}$	ones vector: a vector with all entries one
$\boldsymbol{X}_{ij}$	(i,j)-th block of a block matrix $\boldsymbol{X}$.
$\langle \boldsymbol{A}, \boldsymbol{B} \rangle$	matrix inner product, i.e. $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \mathrm{tr}\left(\boldsymbol{A}^\top \boldsymbol{B}\right)$.
$\mathrm{diag}(\boldsymbol{X})$	a column vector formed from the diagonal of a square matrix $\boldsymbol{X}$
$\mathrm{Diag}(\boldsymbol{x})$	a diagonal matrix that puts $\boldsymbol{x}$ on the main diagonal
$\boldsymbol{e}_i$	ith unit vector, whose ith component is 1 and all others 0
$\otimes$	tensor product, i.e. $\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{1,1}\boldsymbol{B} & a_{1,2}\boldsymbol{B} & \cdots & a_{1,n_2}\boldsymbol{B} \\ a_{2,1}\boldsymbol{B} & a_{2,2}\boldsymbol{B} & \cdots & a_{2,n_2}\boldsymbol{B} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n_1,1}\boldsymbol{B} & a_{n_1,2}\boldsymbol{B} & \cdots & a_{n_1,n_2}\boldsymbol{B} \end{bmatrix}$
$\Omega_{\mathrm{gt}}, \Omega_{\mathrm{gt}}^{\perp}$	support of $\boldsymbol{X}^{\mathrm{gt}}$, its complement support
$T_{\mathrm{gt}}, T_{\mathrm{gt}}^{\perp}$	tangent space at $\boldsymbol{X}^{\mathrm{gt}}$, its orthogonal complement
$\mathcal{P}_{\Omega_{\mathrm{gt}}}, \mathcal{P}_{\Omega_{\mathrm{gt}}^{\perp}}$	projection onto the space of matrices supported on Ω_{gt} and $\Omega_{\mathrm{gt}}^{\perp}$, respectively
$\mathcal{P}_{T_{\mathrm{gt}}}, \mathcal{P}_{T_{\mathrm{gt}}^{\perp}}$	projection onto T_{gt} and $T_{\mathrm{gt}}^{\perp}$, respectively

Table 1: Summary of Notation and Parameters

For notational simplicity, we represent the convex program as follows:

$$\begin{array}{rlr} \text{minimize} & \langle \boldsymbol{W}, \boldsymbol{X} \rangle & \text{dual variable} \\ \text{subject to} & \mathcal{A}(\boldsymbol{X}) = \boldsymbol{b}, & \boldsymbol{y}_{\mathcal{A}} \\ & \mathcal{B}(\boldsymbol{X}) + \boldsymbol{t} = \boldsymbol{c}, & \boldsymbol{y}_{\mathcal{B}} \\ & \boldsymbol{t} \geq \mathbf{0}, & \boldsymbol{z} \geq \mathbf{0} \\ & \boldsymbol{X} \geq \mathbf{0}, & \boldsymbol{Z} \geq \mathbf{0} \\ & \boldsymbol{X} \succeq \mathbf{0}, & \boldsymbol{S} \succeq \mathbf{0} \end{array}$$

where the matrices and operators are defined as follows

(i) $\boldsymbol{W}$ encapsulate all block coefficient matrices $\boldsymbol{W}_{ij}$ for all $(i,j) \in \mathcal{G}$;

(ii) $\mathcal{A}(\boldsymbol{X}) = \boldsymbol{b}$ represents the constraint that $\boldsymbol{X}_{ii} = \boldsymbol{I}_{m_i}$ $(1 \leq i \leq n)$;

(iii) $\mathcal{B}(\boldsymbol{X}) + \boldsymbol{t} = \boldsymbol{c}$ and $\boldsymbol{t} \geq \mathbf{0}$ reformulate all constraints $\boldsymbol{X}_{ij}\mathbf{1} \leq \mathbf{1}$ and $\boldsymbol{X}_{ij}^T\mathbf{1} \leq \mathbf{1}$;

(iv) The variables on the right hand, i.e., $\boldsymbol{y}_{\mathcal{A}}, \boldsymbol{y}_{\mathcal{B}}, \boldsymbol{z}, \boldsymbol{Z}$ and $\boldsymbol{S}$, represent dual variables associated with respective constraints.

The Lagrangian associated with the convex program can be given as follows

$$\begin{aligned} \mathcal{L} &= \langle \boldsymbol{W}, \boldsymbol{X} \rangle + \langle \boldsymbol{y}_{\mathcal{A}}, \mathcal{A}(\boldsymbol{X}) - \boldsymbol{b} \rangle + \langle \boldsymbol{y}_{\mathcal{B}}, \mathcal{B}(\boldsymbol{X}) + \boldsymbol{t} - \boldsymbol{c} \rangle - \langle \boldsymbol{z}, \boldsymbol{t} \rangle - \langle \boldsymbol{Z}, \boldsymbol{X} \rangle - \langle \boldsymbol{S}, \boldsymbol{X} \rangle \\ &= \langle \boldsymbol{W} + \mathcal{A}^*(\boldsymbol{y}_{\mathcal{A}}) + \mathcal{B}^*(\boldsymbol{y}_{\mathcal{B}}) - \boldsymbol{Z} - \boldsymbol{S}, \boldsymbol{X} \rangle - \langle \boldsymbol{b}, \boldsymbol{y}_{\mathcal{A}} \rangle - \langle \boldsymbol{c}, \boldsymbol{y}_{\mathcal{B}} \rangle - \langle \boldsymbol{z} - \boldsymbol{y}_{\mathcal{B}}, \boldsymbol{t} \rangle. \end{aligned}$$

where $\mathcal{A}^*$ denotes the conjugate operator w.r.t. an operator $\mathcal{A}$. The augmented Lagrangian for the convex program can now be written as

$$\begin{aligned} \mathcal{L}_{1/\mu} =& \langle \boldsymbol{b}, \boldsymbol{y}_{\mathcal{A}} \rangle + \langle \boldsymbol{c}, \boldsymbol{y}_{\mathcal{B}} \rangle + \langle \boldsymbol{z} - \boldsymbol{y}_{\mathcal{B}}, \boldsymbol{t} \rangle + \langle \boldsymbol{Z} + \boldsymbol{S} - \boldsymbol{W} - \mathcal{A}^*(\mathrm{y}_{\mathcal{A}}) - \mathcal{B}^*(\mathrm{y}_{\mathcal{B}}), \boldsymbol{X} \rangle \\ &+ \frac{1}{2\mu}\|\boldsymbol{z} - \boldsymbol{y}_{\mathcal{B}}\|_{\mathrm{F}}^2 + \frac{1}{2\mu}\|\boldsymbol{Z} + \boldsymbol{S} - \boldsymbol{W} - \mathcal{A}^*(\mathrm{y}_{\mathcal{A}}) - \mathcal{B}^*(\mathrm{y}_{\mathcal{B}})\|_{\mathrm{F}}^2. \end{aligned}$$

Here, the linear terms above represent the negative standard Lagrangian, whereas the quadratic parts represent the augmenting terms. μ is the penalty parameter that balances the standard Lagrangian and the augmenting terms. The ADMM then proceeds by alternately optimizing each primal and dual variable with others fixed, which results in closed-form solution for each subproblem. Denote by superscript k the iteration number, then we can present the ADMM iterative update procedures as follows

$$
\begin{aligned}
\boldsymbol{y}_{\mathcal{A}}^{(k+1)} &= (\mathcal{A}\mathcal{A}^*)^{-1}\left\{\mathcal{A}\left(-\boldsymbol{W}+\boldsymbol{S}^{(k)}+\mu\boldsymbol{X}^{(k)}+\boldsymbol{Z}^{(k)}-\mathcal{B}^*\left(\boldsymbol{y}_{\mathcal{B}}^{(k)}\right)\right)-\mu\boldsymbol{b}\right\}, \\
\boldsymbol{y}_{\mathcal{B}}^{(k+1)} &= (\mathcal{B}\mathcal{B}^*+\boldsymbol{I})^{-1}\left\{\mathcal{B}\left(-\boldsymbol{W}+\boldsymbol{S}^{(k)}+\mu\boldsymbol{X}^{(k)}+\boldsymbol{Z}^{(k)}-\mathcal{A}^*\left(\boldsymbol{y}_{\mathcal{A}}^{(k+1)}\right)\right)+\boldsymbol{z}^{(k)}+\mu\boldsymbol{t}^{(k)}-\mu\boldsymbol{c}\right\}, \\
\boldsymbol{z}^{(k+1)} &= \left(\boldsymbol{y}_{\mathcal{B}}^{(k+1)}-\mu\boldsymbol{t}^{(k)}\right)_+, \qquad (5)\\
\boldsymbol{Z}^{(k+1)} &= \left(\boldsymbol{W}+\mathcal{A}^*\left(\boldsymbol{y}_{\mathcal{A}}^{(k+1)}\right)+\mathcal{B}^*\left(\boldsymbol{y}_{\mathcal{B}}^{(k+1)}\right)-\boldsymbol{S}^{(k)}-\mu\boldsymbol{X}^{(k)}\right)_+, \\
\boldsymbol{S}^{(k+1)} &= \mathcal{P}_{\mathrm{psd}}\left(\boldsymbol{W}+\mathcal{A}^*\left(\mathbf{y}_{\mathcal{A}}^{(k+1)}\right)+\mathcal{B}^*\left(\mathbf{y}_{\mathcal{B}}^{(k+1)}\right)-\boldsymbol{Z}^{(k+1)}-\mu\boldsymbol{X}^{(k)}\right), \qquad (6)\\
\boldsymbol{X}^{(k+1)} &= \boldsymbol{X}^{k}+\frac{1}{\mu}\left(\boldsymbol{Z}^{(k+1)}+\boldsymbol{S}^{(k+1)}-\boldsymbol{W}-\mathcal{A}^*\left(\mathbf{y}_{\mathcal{A}}^{(k+1)}\right)-\mathcal{B}^*\left(\mathbf{y}_{\mathcal{B}}^{(k+1)}\right)\right) \qquad (7)\\
&= -\frac{1}{\mu}\mathcal{P}_{\mathrm{nsd}}\left(\boldsymbol{W}+\mathcal{A}^*\left(\mathbf{y}_{\mathcal{A}}^{(k+1)}\right)+\mathcal{B}^*\left(\mathbf{y}_{\mathcal{B}}^{(k+1)}\right)-\boldsymbol{Z}^{(k+1)}-\mu\boldsymbol{X}^{(k)}\right), \qquad (8)\\
\boldsymbol{t}^{(k+1)} &= \boldsymbol{t}^{(k)}+\frac{\boldsymbol{z}^{(k+1)}-\boldsymbol{y}_{\mathcal{B}}^{(k+1)}}{\mu}. \qquad (9)
\end{aligned}
$$

Here, the operator $\mathcal{P}_{\mathrm{psd}}$ (resp. $\mathcal{P}_{\mathrm{nsd}}$) denotes the projection onto the positive (resp. negative) semidefinite cone, and $(\cdot)_+$ operator projects all entries of a vector / matrix to non-negative values. Within a reasonable amount of time, ADMM typically returns moderately acceptable results.

4 Proof of Theorem 1

Theorem 1 (Exact Recovery). *Consider the randomized model described in the main body of the paper. There exists a universal constant $c > 0$ such that if*

$$
p_{\mathrm{true}} > c_0 \frac{\log^2(mn)}{\sqrt{np_{\mathrm{obs}}}p_{\mathrm{set}}^2},
$$

then $\boldsymbol{X}^{\mathrm{gt}}$ is the unique solution to MatchLift with probability exceeding $1-\frac{1}{(mn)^3}$.

To prove Theorem 1, we first analyze the Karush–Kuhn–Tucker (KKT) condition for exact recovery, which provides a sufficient and almost necessary condition for uniqueness and optimality. Valid dual certificates are then constructed to guarantee exact recovery.

4.1 Preliminaries and Notations

Without loss of generality, we can treat $\boldsymbol{X}^{\mathrm{gt}}$ as a sub-matrix of an augmented square matrix $\boldsymbol{X}_{\mathrm{sup}}^{\mathrm{gt}}$ such that

$$
\boldsymbol{X}_{\mathrm{sup}}^{\mathrm{gt}} := \mathbf{1}\cdot\mathbf{1}^{\top}\otimes\boldsymbol{I}_n, \qquad (10)
$$

and

$$
\boldsymbol{X}^{\mathrm{gt}} := \begin{bmatrix} \boldsymbol{\Pi}_1 & & & \\ & \boldsymbol{\Pi}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{\Pi}_n \end{bmatrix} \boldsymbol{X}_{\mathrm{sup}}^{\mathrm{gt}} \begin{bmatrix} \boldsymbol{\Pi}_1^{\top} & & & \\ & \boldsymbol{\Pi}_2^{\top} & & \\ & & \ddots & \\ & & & \boldsymbol{\Pi}_n^{\top} \end{bmatrix}, \qquad (11)
$$

where the matrices $\boldsymbol{\Pi}_i \in \mathbb{R}^{|\mathcal{S}_i|\times m}$ are defined such that $\boldsymbol{\Pi}_i$ denotes the submatrix of $\boldsymbol{I}_m$ coming from its rows at indices from $\mathcal{S}_i$. For instance, if $\mathcal{S}_i = \{2,3\}$, then one has

$$
\boldsymbol{\Pi}_i = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \end{bmatrix}.
$$

With this notation, $\boldsymbol{\Pi}_i \boldsymbol{M} \boldsymbol{\Pi}_j^\top$ represents a submatrix of $\boldsymbol{M} \in \mathbb{R}^{m \times m}$ coming from the rows at indices from $\mathcal{S}_i$ and columns at indices from $\mathcal{S}_j$. Conversely, for any matrix $\tilde{\boldsymbol{M}} \in \mathbb{R}^{|\mathcal{S}_i| \times |\mathcal{S}_j|}$, the matrix $\boldsymbol{\Pi}_i^\top \tilde{\boldsymbol{M}} \boldsymbol{\Pi}_j$ converts $\tilde{\boldsymbol{M}}$ to an $m \times m$ matrix space via zero padding.

With this notation, we can represent $\boldsymbol{X}^{\text{in}}$ as a submatrix of $\boldsymbol{X}_{\text{sup}}^{\text{in}}$, which is a corrupted version of $\boldsymbol{X}_{\text{sup}}^{\text{gt}}$ and obeys

$$\boldsymbol{X}_{ij}^{\text{in}} := \boldsymbol{\Pi}_i \left(\boldsymbol{X}_{\text{sup}}^{\text{in}} \right)_{ij} \boldsymbol{\Pi}_j^\top. \tag{12}$$

For notational simplicity, we set

$$\boldsymbol{W}_{ij} := \begin{cases} -\boldsymbol{X}_{ij}^{\text{in}} + \frac{\sqrt{p_{\text{obs}}}}{2} \mathbf{1} \cdot \mathbf{1}^\top, & \text{if } (i,j) \in \mathcal{G}, \\ \frac{\sqrt{p_{\text{obs}}}}{2} \mathbf{1} \cdot \mathbf{1}^\top, & \text{else}. \end{cases} \tag{13}$$

Before continuing to the proof, it is convenient to introduce some notations that will be used throughout. Denote by Ω_{gt} and $\Omega_{\text{gt}}^\perp$ the support of $\boldsymbol{X}^{\text{gt}}$ and its complement support, respectively, and let $\mathcal{P}_{\Omega_{\text{gt}}}$ and $\mathcal{P}_{\Omega_{\text{gt}}^\perp}$ represent the orthogonal projection onto the linear space of matrices supported on Ω_{gt} and its complement support $\Omega_{\text{gt}}^\perp$, respectively. Define T_{gt} to be the tangent space at $\boldsymbol{X}^{\text{gt}}$ w.r.t. all symmetric matrices of rank at most m, i.e. the space of symmetric matrices of the form

$$T_{\text{gt}} := \left\{ \begin{bmatrix} \boldsymbol{\Pi}_1 \\ \boldsymbol{\Pi}_2 \\ \vdots \\ \boldsymbol{\Pi}_n \end{bmatrix} \boldsymbol{M} + \boldsymbol{M}^\top \begin{bmatrix} \boldsymbol{\Pi}_1^\top & \boldsymbol{\Pi}_2^\top & \cdots & \boldsymbol{\Pi}_n^\top \end{bmatrix} : \ \boldsymbol{M} \in \mathbb{R}^{m \times N} \right\}, \tag{14}$$

and denote by $T_{\text{gt}}^\perp$ its orthogonal complement. We then denote by $\mathcal{P}_{T_{\text{gt}}}$ (resp. $\mathcal{P}_{T_{\text{gt}}^\perp}$) the orthogonal projection onto T_{gt} (resp. $T_{\text{gt}}^\perp$). In passing, if we define

$$\boldsymbol{\Sigma} := \text{Diag} \left\{ \left[\frac{n}{n_1}, \cdots, \frac{n}{n_m} \right] \right\}, \tag{15}$$

then the columns of

$$\boldsymbol{U} := \frac{1}{\sqrt{n}} \begin{bmatrix} \boldsymbol{\Pi}_1 \\ \boldsymbol{\Pi}_2 \\ \vdots \\ \boldsymbol{\Pi}_n \end{bmatrix} \boldsymbol{\Sigma}^{\frac{1}{2}} \tag{16}$$

form the set of eigenvectors of $\boldsymbol{X}^{\text{gt}}$, and for any symmetric matrix $\boldsymbol{M}$,

$$\mathcal{P}_{T_{\text{gt}}^\perp} (\boldsymbol{M}) = \left(\boldsymbol{I} - \boldsymbol{U}\boldsymbol{U}^\top \right) \boldsymbol{M} \left(\boldsymbol{I} - \boldsymbol{U}\boldsymbol{U}^\top \right). \tag{17}$$

Furthermore, we define a vector $\boldsymbol{d}$ to be

$$\boldsymbol{d} := \begin{bmatrix} \boldsymbol{\Pi}_1 \\ \boldsymbol{\Pi}_2 \\ \vdots \\ \boldsymbol{\Pi}_n \end{bmatrix} \boldsymbol{\Sigma} \mathbf{1}_m. \tag{18}$$

Put another way, if any row index j of $\boldsymbol{X}_{\text{gt}}$ is associated with the element $s \in [m]$, then $\boldsymbol{d}_j = \frac{n}{n_s}$. One can then easily verify that

$$\left\langle \boldsymbol{d} \cdot \boldsymbol{d}^\top, \boldsymbol{X}_{\text{gt}} - \frac{1}{m} \mathbf{1} \cdot \mathbf{1}^\top \right\rangle = \left\langle \boldsymbol{d} \cdot \boldsymbol{d}^\top, \boldsymbol{X}_{\text{gt}} \right\rangle - \frac{1}{m} \left(\mathbf{1}^\top \cdot \boldsymbol{d} \right)^2 = 0. \tag{19}$$

In fact, when n_i's are sufficiently close to each other, $\boldsymbol{d} \cdot \boldsymbol{d}^\top$ is a good approximation of $\mathbf{1} \cdot \mathbf{1}^\top$, as claimed in the following lemma.

Lemma 1. *Consider a set of Bernoulli random variables $\nu_i \sim$ Bernoulli(p) $(1 \le i \le n)$, and set $s := \sum_{i=1}^{n} \nu_i$. Let n_i $(1 \le i \le m)$ be independent copies of s, and denote $N = n_1 + \cdots + n_m$. If $p > \frac{c_7 \log^2(mn)}{n}$, then the matrix*

$$\boldsymbol{A} := (np)^2 \begin{bmatrix} \frac{1}{n_1}\mathbf{1}_{n_1} \\ \frac{1}{n_2}\mathbf{1}_{n_2} \\ \vdots \\ \frac{1}{n_m}\mathbf{1}_{n_m} \end{bmatrix} \begin{bmatrix} \frac{1}{n_1}\mathbf{1}_{n_1}^\top & \frac{1}{n_2}\mathbf{1}_{n_2}^\top & \cdots & \frac{1}{n_m}\mathbf{1}_{n_m}^\top \end{bmatrix} \tag{20}$$

satisfies

$$\left\| \frac{1}{m}\boldsymbol{A} - \frac{1}{m}\mathbf{1}_N \cdot \mathbf{1}_N^\top \right\| \le c_8 \sqrt{np\log(mn)} \tag{21}$$

and

$$\left\| \boldsymbol{A} - \mathbf{1}_N \cdot \mathbf{1}_N^\top \right\|_\infty \le c_9 \sqrt{\frac{\log(mn)}{np}} \tag{22}$$

with probability exceeding $1 - \frac{1}{m^5 n^5}$, where c_7, c_8, c_9 are some universal constants.

Proof. See Appendix A.1. □

Since $p^2 \boldsymbol{d} \cdot \boldsymbol{d}^\top$ is equivalent to $\boldsymbol{A}$ defined in (20) up to row / column permutation, Lemma 1 reveals that

$$\left\| \frac{p^2}{m}\boldsymbol{d} \cdot \boldsymbol{d}^\top - \frac{1}{m}\mathbf{1}_N \cdot \mathbf{1}_N^\top \right\| \le c_8 \sqrt{np\log(mn)}$$

with high probability.

The following bound on the operator norm of a random block matrix is useful for deriving our main results.

Lemma 2. *Let $\boldsymbol{M} = [\boldsymbol{M}_{ij}]_{1 \le i,j \le n}$ be a symmetric block matrix, where $\boldsymbol{M}_{ij}$'s are jointly independent $m_i \times m_j$ matrices satisfying*

$$\mathbb{E}\boldsymbol{M}_{ij} = \mathbf{0}, \quad \mathbb{E}\left\|\boldsymbol{M}_{ij}\right\|^2 \le 1, \quad \textit{and} \quad \left\|\boldsymbol{M}_{ij}\right\| \le \sqrt{n}, \quad (1 \le i, j \le n). \tag{23}$$

Besides, $m_i \le m$ holds for all $1 \le i \le n$. Then there exists an absolute constant $c_0 > 0$ such that

$$\|\boldsymbol{M}\| \le c_0 \sqrt{n} \log(mn)$$

holds with probability exceeding $1 - \frac{1}{m^5 n^5}$.

Proof. See Appendix A.2. □

Additionally, the second smallest eigenvalue of the Laplacian matrix of a random Erdős–Rényi graph can be bounded below by the following lemma.

Lemma 3. *Consider an Erdős–Rényi graph $\mathcal{G} \sim \mathcal{G}(n, p)$ and any positive integer m, and let $\boldsymbol{L} \in \mathbb{R}^{n \times n}$ represent its (unnormalized) Laplacian matrix. There exist absolute constants $c_3, c_4 > 0$ such that if $p > c_3 \log^2(mn)/n$, then the algebraic connectivity $a(\mathcal{G})$ of $\mathcal{G}$ (i.e. the second smallest eigenvalue of $\boldsymbol{L}$) satisfies*

$$a(\mathcal{G}) \ge np - c_4\sqrt{np}\log(mn) \tag{24}$$

with probability exceeding $1 - \frac{2}{(mn)^5}$.

Proof. See Appendix A.3. □

Finally, if we denote by n_s (resp. $n_{s,t}$) the number of sets $\mathcal{S}_i$ $(1 \le i \le n)$ containing the element s (resp. containing s and t simultaneously), then these quantities sharply concentrate around their mean values, as stated in the following lemma.

Lemma 4. *There are some universal constants $c_8, c_9 > 0$ such that if $p_{\mathrm{set}}^2 > \frac{\log(mn)}{n}$, then*

$$\left|n_s - np_{\mathrm{set}}\right| \leq \sqrt{c_8 np_{\mathrm{set}} \log(mn)}, \quad \forall 1 \leq s \leq m,$$
$$\left|n_{s,t} - np_{\mathrm{set}}^2\right| \leq \sqrt{c_8 np_{\mathrm{set}}^2 \log(mn)}, \quad \forall 1 \leq s < t \leq m,$$

hold with probability exceeding $1 - \frac{1}{(mn)^{10}}$.

Proof. In passing, the claim follows immediately from the Bernstein inequality that

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} \nu_i - np\right| > t\right) \leq 2\exp\left(-\frac{\frac{1}{2}t^2}{np(1-p) + \frac{1}{3}t}\right)$$

where $\nu_i \sim \text{Bernoulli}(p)$ are i.i.d. random variables. Interested readers are referred to [4] for a tutorial. $\square$

4.2 Optimality and Uniqueness Condition

Recall that $n_i := |\mathcal{I}_i|$. The convex relaxation is exact if one can construct valid dual certificates, as summarized in the following lemma.

Lemma 5. *Suppose that there exist dual certificates $\alpha > 0$, $\boldsymbol{Z} = [\boldsymbol{Z}_{ij}]_{1\leq i,j\leq n} \in \mathbb{S}^{N\times N}$ and $\boldsymbol{Y} = [\boldsymbol{Y}_{ij}]_{1\leq i,j\leq n} \in \mathbb{S}^{N\times N}$ obeying*

$$\boldsymbol{Y} - \alpha \boldsymbol{d}\boldsymbol{d}^\top \succeq \boldsymbol{0}, \tag{25}$$
$$\mathcal{P}_{\Omega_{\mathrm{gt}}}(\boldsymbol{Z}) = \boldsymbol{0}, \quad \mathcal{P}_{\Omega_{\mathrm{gt}}^\perp}(\boldsymbol{Z}) \geq \boldsymbol{0}, \tag{26}$$
$$\boldsymbol{Y}_{ij} = \boldsymbol{W}_{ij} - \boldsymbol{Z}_{ij}, \quad 1 \leq i < j \leq n, \tag{27}$$
$$\boldsymbol{Y} - \alpha \boldsymbol{d}\boldsymbol{d}^\top \in T_{\mathrm{gt}}^\perp. \tag{28}$$

Then $\boldsymbol{X}^{\mathrm{gt}}$ is the unique solution to MatchLift if either of the following two conditions is satisfied:
i) All entries of $\boldsymbol{Z}_{ij}$ ($\forall i \neq j$) within the support $\Omega_{\mathrm{gt}}^\perp$ are strictly positive;
ii) For all $\boldsymbol{M}$ satisfying $\mathcal{P}_{T_{\mathrm{gt}}^\perp}(\boldsymbol{M}) \succ \boldsymbol{0}$,

$$\left\langle \boldsymbol{Y} - \alpha \boldsymbol{d}\boldsymbol{d}^\top, \mathcal{P}_{T_{\mathrm{gt}}^\perp}(\boldsymbol{M}) \right\rangle > 0, \tag{29}$$

and, additionally,

$$\frac{n}{n_i} + \frac{n}{n_j} \neq \frac{n^2}{n_i n_j}, \quad 1 \leq i, j \leq m. \tag{30}$$

Proof. See Appendix A.4. $\square$

That said, to prove Theorem 1, it is sufficient (under the hypotheses of Theorem 1) to generate, with high probability, valid dual certificates $\boldsymbol{Y}$, $\boldsymbol{Z}$ and $\alpha > 0$ obeying the optimality conditions of Lemma 5. This is the objective of the next subsection.

4.3 Construction of Dual Certificates

Decompose the input $\boldsymbol{X}^{\mathrm{in}}$ into two components $\boldsymbol{X}^{\mathrm{in}} = \boldsymbol{X}^{\mathrm{false}} + \boldsymbol{X}^{\mathrm{true}}$, where

$$\boldsymbol{X}^{\mathrm{true}} = \mathcal{P}_{\Omega_{\mathrm{gt}}}\left(\boldsymbol{X}^{\mathrm{in}}\right), \quad \text{and} \quad \boldsymbol{X}^{\mathrm{false}} = \mathcal{P}_{\Omega_{\mathrm{gt}}^\perp}\left(\boldsymbol{X}^{\mathrm{in}}\right). \tag{31}$$

That said, $\boldsymbol{X}^{\mathrm{true}}$ (resp. $\boldsymbol{X}^{\mathrm{false}}$) consists of all correct (resp. incorrect) correspondences (i.e. non-zero entries) encoded in $\boldsymbol{X}^{\mathrm{in}}$. This allows us to write

$$\boldsymbol{W}_{ij} = \begin{cases} -\boldsymbol{X}_{ij}^{\mathrm{false}} + \frac{\sqrt{p_{\mathrm{obs}}}}{2}\boldsymbol{E}_{ij} - \boldsymbol{X}_{ij}^{\mathrm{true}} + \frac{\sqrt{p_{\mathrm{obs}}}}{2}\boldsymbol{E}_{ij}^\perp, & \text{if } (i,j) \in \mathcal{G}, \\ \frac{\sqrt{p_{\mathrm{obs}}}}{2}\boldsymbol{E}_{ij} + \frac{\sqrt{p_{\mathrm{obs}}}}{2}\boldsymbol{E}_{ij}^\perp, & \text{else}, \end{cases} \tag{32}$$

where $\boldsymbol{E}$ and $\boldsymbol{E}^{\perp}$ are defined to be

$$\boldsymbol{E} := \mathcal{P}_{\Omega_{\mathrm{gt}}}\left(\mathbf{1}\cdot\mathbf{1}^{\top}\right), \quad \text{and} \quad \boldsymbol{E}^{\perp} := \mathbf{1}\cdot\mathbf{1}^{\top} - \boldsymbol{E}. \tag{33}$$

We propose constructing the dual certificate $\boldsymbol{Y}$ by producing three symmetric matrix components $\boldsymbol{Y}^{\mathrm{true},1}$, $\boldsymbol{Y}^{\mathrm{true},2}$, and $\boldsymbol{Y}^{\mathrm{L}}$ separately, as follows.

1. *Construction of* $\boldsymbol{Z}^{\mathrm{m}}$ *and* $\boldsymbol{R}^{\mathrm{m}}$. For any $\beta \geq 0$, define α_{β} as

$$\alpha_{\beta} := \arg \min_{\tau : \beta\mathbf{1}\cdot\mathbf{1}^{\top} - \tau\boldsymbol{d}\cdot\boldsymbol{d}^{\top} \geq \mathbf{0}} \left\| \beta\mathbf{1}\cdot\mathbf{1}^{\top} - \tau\boldsymbol{d}\cdot\boldsymbol{d}^{\top} \right\|_{\infty}. \tag{34}$$

By setting $\beta_0 := \frac{\sqrt{p_{\mathrm{obs}}}}{2} - \frac{p_{\mathrm{obs}}}{m} - \sqrt{\frac{c_{10} p_{\mathrm{obs}} \log(mn)}{n p_{\mathrm{set}}^3}}$, we produce $\boldsymbol{Z}^{\mathrm{m}}$ and $\boldsymbol{R}^{\mathrm{m}}$ as follows

$$\boldsymbol{Z}^{\mathrm{m}} = \mathcal{P}_{\Omega_{\mathrm{gt}}^{\perp}}\left(\left(\frac{\sqrt{p_{\mathrm{obs}}}}{2} - \frac{p_{\mathrm{obs}}}{m}\right)\mathbf{1}\cdot\mathbf{1}^{\top} - \alpha_{\beta_0}\boldsymbol{d}\cdot\boldsymbol{d}^{\top}\right) \tag{35}$$

and

$$\boldsymbol{R}^{\mathrm{m}} = \mathcal{P}_{\Omega_{\mathrm{gt}}}\left(\left(\frac{\sqrt{p_{\mathrm{obs}}}}{2} - \frac{p_{\mathrm{obs}}}{m}\right)\mathbf{1}\cdot\mathbf{1}^{\top} - \alpha_{\beta_0}\boldsymbol{d}\cdot\boldsymbol{d}^{\top}\right) \tag{36}$$

for some sufficiently large constant $c_{10} > 0$.

2. *Construction of* $\boldsymbol{Y}^{\mathrm{true},1}$ *and* $\boldsymbol{Y}^{\mathrm{true},2}$. We set

$$\boldsymbol{Y}_{ij}^{\mathrm{true},1} = \begin{cases} -\boldsymbol{X}_{ij}^{\mathrm{true}} + \frac{p_{\mathrm{obs}}}{m}\boldsymbol{E}_{ij}, & \text{if } i < j, \\ \sum_{j=1}^{n} \boldsymbol{\Pi}_i\boldsymbol{\Pi}_i^{\top}\left(\boldsymbol{X}_{ij}^{\mathrm{true}} - \frac{p_{\mathrm{obs}}}{m}\boldsymbol{E}_{ij}\right)\boldsymbol{\Pi}_j\boldsymbol{\Pi}_i^{\top}, & \text{if } i = j, \end{cases}$$

and

$$\boldsymbol{Y}_{ij}^{\mathrm{true},2} = \begin{cases} \boldsymbol{R}_{ij}^{\mathrm{m}}, & \text{if } i < j, \\ -\sum_{j=1}^{n} \boldsymbol{\Pi}_i\boldsymbol{\Pi}_i^{\top}\boldsymbol{R}_{ij}^{\mathrm{m}}\boldsymbol{\Pi}_j\boldsymbol{\Pi}_i^{\top}, & \text{if } i = j. \end{cases}$$

3. *Construction of* $\boldsymbol{Y}^{\mathrm{L}}$ *and* $\boldsymbol{Z}^{\mathrm{L}}$ *via an iterative procedure.* Next, we generate $\boldsymbol{Y}^{\mathrm{L}}$ via the following iterative procedure. Here, for any matrix $\boldsymbol{M}$, we let $\boldsymbol{M}_{ij}(s, s')$ represent the entry in the $(i,j)^{\mathrm{th}}$ block $\boldsymbol{M}_{ij}$ that encodes the correspondence from s to s'.

Construction of a dual certificate $\boldsymbol{Y}^{\mathrm{L}}$.

---

1. **initialize**: Set the symmetric matrix $\boldsymbol{Y}^{\mathrm{L},0}$ such that

$$\boldsymbol{Y}_{ij}^{\mathrm{L},0} = \begin{cases} -\boldsymbol{X}_{ij}^{\mathrm{false}} + \frac{p_{\mathrm{obs}}}{m}\boldsymbol{E}_{ij}^{\perp}, & \text{if } i < j, \\ \mathbf{0}, & \text{if } i = j, \end{cases}$$

and start with $\boldsymbol{Z}^{\mathrm{L}} = \mathbf{0}$.

2. **for** each *non-zero* entry $\boldsymbol{Y}_{ij}^{\mathrm{L},0}(s, s')$:
3. Set $a = \boldsymbol{Y}_{ij}^{\mathrm{L},0}(s, s')$, $B_{i,j,s,s'} = \{l \notin \{i,j\} \mid (s,s') \in \mathcal{S}_l\}$ and $n_{i,j}^{s,s'} = |B_{i,j,s,s'}|$.
4. **for** each set $l \in B_{i,j,s,s'}$: perform

$$\begin{cases} \boldsymbol{Z}_{il}^{\mathrm{L}}(s,s') \leftarrow \boldsymbol{Z}_{il}^{\mathrm{L}}(s,s') - \frac{a}{n_{i,j}^{s,s'}}, & \boldsymbol{Z}_{li}^{\mathrm{L}}(s',s) \leftarrow \boldsymbol{Z}_{li}^{\mathrm{L}}(s',s) - \frac{a}{n_{i,j}^{s,s'}}, \\ \boldsymbol{Z}_{lj}^{\mathrm{L}}(s,s') \leftarrow \boldsymbol{Z}_{lj}^{\mathrm{L}}(s,s') - \frac{a}{n_{i,j}^{s,s'}}, & \boldsymbol{Z}_{jl}^{\mathrm{L}}(s',s) \leftarrow \boldsymbol{Z}_{jl}^{\mathrm{L}}(s',s) - \frac{a}{n_{i,j}^{s,s'}}, \\ \boldsymbol{Z}_{ll}^{\mathrm{L}}(s,s') \leftarrow \boldsymbol{Z}_{ll}^{\mathrm{L}}(s,s') + \frac{a}{n_{i,j}^{s,s'}}, & \boldsymbol{Z}_{ll}^{\mathrm{L}}(s',s) \leftarrow \boldsymbol{Z}_{ll}^{\mathrm{L}}(s',s) + \frac{a}{n_{i,j}^{s,s'}}. \end{cases}$$

5. **output**: $\boldsymbol{Y}^{\mathrm{L}} = \boldsymbol{Y}^{\mathrm{L},0} + \boldsymbol{Z}^{\mathrm{L}}$.

$$
\begin{array}{c|cc|cc|ccc|cc}
 & 1 & 2 & 1 & 3 & 2 & 3 & 4 & 2 & 3 \\
\hline
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
\hline
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}
$$

(a) Input $\boldsymbol{Y}^0$

$$
\begin{array}{c|cc|cc|ccc|cc}
 & 1 & 2 & 1 & 3 & 2 & 3 & 4 & 2 & 3 \\
\hline
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5 \\
\hline
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5 & 0 \\
\hline
2 & 0 & 0 & 0 & 0.5 & 0 & -0.5 & 0 & 0 & 0 \\
3 & 0 & 0.5 & 0 & 0 & -0.5 & 0 & 0 & 0 & 0 \\
4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
2 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & -0.5 \\
3 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & -0.5 & 0
\end{array}
$$

(b) $\boldsymbol{Z}^{\mathrm{L}}$

Figure 1: A toy example for constructing $\boldsymbol{Z}^{\mathrm{L}}$, where 4 shapes $\mathcal{S}_1 = \{1,2\}$, $\mathcal{S}_2 = \{1,3\}$, $\mathcal{S}_3 = \{2,3,4\}$, and $\mathcal{S}_4 = \{2,3\}$ are considered. The input incorrectly maps point 1 to 3 between $\mathcal{S}_1$ and $\mathcal{S}_2$, and both points are contained in $\mathcal{S}_3$ and $\mathcal{S}_4$.

4. *Construction of $\boldsymbol{Y}$ and $\boldsymbol{Z}$:* define $\boldsymbol{Y}$ and $\boldsymbol{Z}$ such that

$$
\boldsymbol{Y} = \boldsymbol{Y}^{\mathrm{true},1} + \boldsymbol{Y}^{\mathrm{true},2} + \boldsymbol{Y}^{\mathrm{L}} + \alpha_{\beta_0} \boldsymbol{d} \cdot \boldsymbol{d}^{\top}, \tag{37}
$$

$$
\boldsymbol{Z}_{ij} = \begin{cases} \boldsymbol{Z}_{ij}^{\mathrm{m}} - \boldsymbol{Z}_{ij}^{\mathrm{L}}, & \text{if } i \neq j, \\ \boldsymbol{0}, & \text{if } i = j. \end{cases} \tag{38}
$$

Remark 1. Below is a toy example to illustrate the proposed procedure for constructing $\boldsymbol{Z}^{\mathrm{L}}$. Consider three sets $\mathcal{S}_1 = \{1,2\}$, $\mathcal{S}_2 = \{1,3\}$, $\mathcal{S}_3 = \{2,3,4\}$, and $\mathcal{S}_4 = \{1,3\}$. Suppose that $\boldsymbol{Y}^{\mathrm{L},0}$ only contains two non-zero entries that incorrectly maps elements 1 to 3 in $\boldsymbol{Y}_{12}^{\mathrm{L},0}$, as illustrated in Fig. 1(a). The resulting $\boldsymbol{Z}^{\mathrm{L}}$ is shown in Fig. 1(b). Clearly, $\boldsymbol{Y}^{\mathrm{L},0} + \boldsymbol{Z}^{\mathrm{L}}$ obeys $\boldsymbol{Y}^{\mathrm{L},0} + \boldsymbol{Z}^{\mathrm{L}} \in T_{\mathrm{gt}}^{\perp}$.

With the above construction procedure, one can easily verify that:
(1) $\boldsymbol{Y}^{\mathrm{true},1}$, $\boldsymbol{Y}^{\mathrm{true},2}$ and $\boldsymbol{Y}^{\mathrm{L}}$ are all contained in the space $T_{\mathrm{gt}}^{\perp}$;
(2) $\mathcal{P}_{\Omega^{\mathrm{gt}}}(\boldsymbol{Z}) = \boldsymbol{0}$;
(3) For any $i \neq j$, if we set $\boldsymbol{M}^{\mathrm{m}} := \alpha_{\beta_0} \boldsymbol{d} \cdot \boldsymbol{d}^{\top}$, then

$$
\begin{aligned}
\boldsymbol{Y}_{ij} =& \boldsymbol{Y}_{ij}^{\mathrm{true},1} + \boldsymbol{Y}_{ij}^{\mathrm{true},2} + \boldsymbol{Y}_{ij}^{\mathrm{L}} + \boldsymbol{M}_{ij}^{\mathrm{m}} \\
=& -\boldsymbol{X}_{ij}^{\mathrm{true}} + \frac{p_{\mathrm{obs}}}{m} \boldsymbol{E}_{ij} + \boldsymbol{R}_{ij}^{\mathrm{m}} - \boldsymbol{X}_{ij}^{\mathrm{false}} + \frac{p_{\mathrm{obs}}}{m} \boldsymbol{E}_{ij}^{\perp} + \boldsymbol{Z}_{ij}^{\mathrm{L}} + \boldsymbol{M}_{ij}^{\mathrm{m}} \\
=& -\boldsymbol{X}_{ij}^{\mathrm{true}} - \boldsymbol{X}_{ij}^{\mathrm{false}} + \frac{\sqrt{p_{\mathrm{obs}}}}{2} \boldsymbol{1} \cdot \boldsymbol{1}^{\top} - \left(\left(\frac{\sqrt{p_{\mathrm{obs}}}}{2} - \frac{p_{\mathrm{obs}}}{m} \right) \boldsymbol{1} \cdot \boldsymbol{1}^{\top} - \boldsymbol{R}_{ij}^{\mathrm{m}} \right) + \boldsymbol{Z}_{ij}^{\mathrm{L}} + \boldsymbol{M}_{ij}^{\mathrm{m}} \\
=& \boldsymbol{W}_{ij} - \left(\boldsymbol{Z}_{ij}^{\mathrm{m}} - \boldsymbol{Z}_{ij}^{\mathrm{L}} \right).
\end{aligned} \tag{39}
$$

Furthermore, from Lemma 1 one can obtain

$$
\left\| \boldsymbol{d} \cdot \boldsymbol{d}^{\top} - \boldsymbol{1} \cdot \boldsymbol{1}^{\top} \right\|_{\infty} = O\left(\sqrt{\frac{\log(mn)}{n p_{\mathrm{set}}}} \right).
$$

This taken collectively with (34) ensures that

$$
\alpha_{\beta_0} = \frac{\sqrt{p_{\mathrm{obs}}}}{2} - \frac{p_{\mathrm{obs}}}{m} - O\left(\sqrt{\frac{c_{10} p_{\mathrm{obs}} \log(mn)}{n p_{\mathrm{set}}^3}} \right) > 0 \tag{40}
$$

as long as $p_{\mathrm{set}}^3 > \frac{c_{15} \log(mn)}{n}$ for some constant $c_{15} > 0$.

Consequently, we will establish that $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are valid dual certificates if they satisfy

$$\begin{cases} \text{all entries of } \boldsymbol{Z}_{ij}^{\mathrm{m}} - \boldsymbol{Z}_{ij}^{\mathrm{L}} \ (\forall i \neq j) \text{ within } \Omega_{\mathrm{gt}}^{\perp} \text{ are strictly positive;} \\ \boldsymbol{Y}^{\mathrm{true},1} + \boldsymbol{Y}^{\mathrm{true},2} + \boldsymbol{Y}^{\mathrm{L}} \succeq \boldsymbol{0}. \end{cases} \tag{41}$$

Lemma 6. *There are some universal constants $c_0, c_1 > 0$ such that*

$$\left\| \boldsymbol{Y}^{\mathrm{L}} \right\| \leq c_0 \sqrt{\frac{n p_{\mathrm{obs}} \log(mn)}{p_{\mathrm{set}}^2}}$$

and

$$\left\| \boldsymbol{Z}_{ij}^{\mathrm{L}} \right\|_{\infty} \leq \sqrt{\frac{c_1 p_{\mathrm{obs}} \log(mn)}{n p_{\mathrm{set}}^3}}, \quad \forall i \neq j$$

with probability exceeding $1 - \frac{1}{(mn)^4}$.

Proof. See Appendix A.5. □

Lemma 7. *There are some universal constants $c_5, c_6, c_7 > 0$ such that if $p_{\mathrm{true}} p_{\mathrm{obs}} p_{\mathrm{set}} > \frac{c_7 \log^2(mn)}{n}$, then with probability exceeding $1 - \frac{1}{(mn)^{10}}$, one has*

$$\left\| \boldsymbol{Y}^{\mathrm{true},2} \right\| \leq c_5 \sqrt{\frac{n p_{\mathrm{obs}} \log(mn)}{p_{\mathrm{set}}}},$$

and

$$\left\langle \boldsymbol{v}\boldsymbol{v}^{\top}, \boldsymbol{Y}^{\mathrm{true},1} \right\rangle \geq \frac{1}{2} n p_{\mathrm{set}} p_{\mathrm{true}} p_{\mathrm{obs}} - c_6 \sqrt{n p_{\mathrm{set}} p_{\mathrm{obs}}} \log(mn)$$

for all unit vector $\boldsymbol{v}$ satisfying $\boldsymbol{v}\boldsymbol{v}^{\top} \in T_{\mathrm{gt}}^{\perp}$.

Proof. See Appendix A.6. □

Combining Lemmas 6 and 7 yields that there exists an absolute constant $c_0 > 0$ such that if

$$p_{\mathrm{true}} > c_0 \frac{\log^2(mn)}{\sqrt{n p_{\mathrm{obs}} p_{\mathrm{set}}^4}},$$

then

$$\boldsymbol{Y} = \boldsymbol{Y}^{\mathrm{true},1} + \boldsymbol{Y}^{\mathrm{true},2} + \boldsymbol{Y}^{\mathrm{L}} \succeq \boldsymbol{0}.$$

On the other hand, observe that all entries of the non-negative matrix $\boldsymbol{Z}^{\mathrm{m}}$ lying in the index set $\Omega_{\mathrm{gt}}^{\perp}$ are *bounded below* in magnitude by $\sqrt{\frac{c_{10} p_{\mathrm{obs}} \log(mn)}{n p_{\mathrm{set}}^3}}$. For sufficiently large c_{10}, one can conclude that all entries of $\boldsymbol{Z}_{il}^{\mathrm{m}} - \boldsymbol{Z}_{il}^{\mathrm{L}}$ outside Ω_{gt} are strictly positive.

So far we have justified that $\boldsymbol{Y}$ and $\boldsymbol{Z}$ satisfy (41), thereby certifying that the proposed algorithm correctly recovers the ground-truth matching.

A Proofs of Auxiliary Lemmas

A.1 Proof of Lemma 1

Denote by $\overline{\boldsymbol{A}} := \boldsymbol{1}_N \cdot \boldsymbol{1}_N^T$. From Bernstein inequality, n_i sharply concentrates around np such that if $p > \frac{c_6 \log^2(mn)}{n}$

$$|n_i - np| \leq c_5 \sqrt{np \log(mn)}, \qquad \forall 1 \leq i \leq m \tag{42}$$

with probability exceeding $1-(mn)^{-10}$, where $c_5, c_6 > 0$ are some absolute constants.

The bound (42) also implies that

$$\left\| \boldsymbol{I} - \begin{bmatrix} \frac{np}{n_1} & & & \\ & \frac{np}{n_2} & & \\ & & \ddots & \\ & & & \frac{np}{n_m} \end{bmatrix} \right\| \leq \max_{1\leq i\leq m} \frac{|n_i - np|}{n_i} \leq \frac{c_5\sqrt{np\log(mn)}}{np - c_5\sqrt{np\log(mn)}}$$
$$\leq 2c_5\sqrt{\frac{\log(mn)}{np}}.$$

Similarly, one has

$$|N - nmp| \leq c_5\sqrt{pmn\log(mn)}$$

with probability exceeding $1-(mn)^{-10}$, which implies that

$$\left\|\overline{\boldsymbol{A}}\right\| = N \leq nmp + c_5\sqrt{pmn\log(mn)} < 2nmp.$$

Rewrite $\boldsymbol{A}$ as

$$\boldsymbol{A} := \begin{bmatrix} \frac{np}{n_1}\mathrm{Diag}\left(\mathbf{1}_{n_1}\right) & & \\ & \ddots & \\ & & \frac{np}{n_m}\mathrm{Diag}\left(\mathbf{1}_{n_m}\right) \end{bmatrix} \cdot \overline{\boldsymbol{A}} \cdot \begin{bmatrix} \frac{np}{n_1}\mathrm{Diag}\left(\mathbf{1}_{n_1}\right) & & \\ & \ddots & \\ & & \frac{np}{n_m}\mathrm{Diag}\left(\mathbf{1}_{n_m}\right) \end{bmatrix}.$$

This allows us to bound the deviation of $\boldsymbol{A}$ from $\overline{\boldsymbol{A}}$ as follows

$$\begin{aligned}
\left\|\boldsymbol{A} - \overline{\boldsymbol{A}}\right\| &\leq \left\| \boldsymbol{A} - \begin{bmatrix} \frac{np}{n_1}\mathrm{Diag}\left(\mathbf{1}_{n_1}\right) & & \\ & \ddots & \\ & & \frac{np}{n_m}\mathrm{Diag}\left(\mathbf{1}_{n_m}\right) \end{bmatrix} \overline{\boldsymbol{A}} \right\| + \left\| \begin{bmatrix} \frac{np}{n_1}\mathrm{Diag}\left(\mathbf{1}_{n_1}\right) & & \\ & \ddots & \\ & & \frac{np}{n_m}\mathrm{Diag}\left(\mathbf{1}_{n_m}\right) \end{bmatrix} \overline{\boldsymbol{A}} - \overline{\boldsymbol{A}} \right\| \\
&\leq \left(\left\| \begin{bmatrix} \frac{np}{n_1}\mathrm{Diag}\left(\mathbf{1}_{n_1}\right) & & \\ & \ddots & \\ & & \frac{np}{n_m}\mathrm{Diag}\left(\mathbf{1}_{n_m}\right) \end{bmatrix} \right\| + 1 \right) \left\|\overline{\boldsymbol{A}}\right\| \left\| \boldsymbol{I} - \begin{bmatrix} \frac{np}{n_1}\mathrm{Diag}\left(\mathbf{1}_{n_1}\right) & & \\ & \ddots & \\ & & \frac{np}{n_m}\mathrm{Diag}\left(\mathbf{1}_{n_m}\right) \end{bmatrix} \right\| \\
&\leq \left(1 + c_5\sqrt{\frac{\log(mn)}{np}} + 1\right) 2nmp \cdot 2c_5\sqrt{\frac{\log(mn)}{np}} \\
&\leq c_6 m\sqrt{np\log(mn)}
\end{aligned}$$

for some universal constant $c_6 > 0$.

On the other hand, it follows immediately from (42) that

$$\begin{aligned}
\left\|\boldsymbol{A} - \mathbf{1}\cdot\mathbf{1}^\top\right\|_\infty = \max_{1\leq i,j\leq m}\left|\frac{(np)^2}{n_i n_j} - 1\right| &= \max_{1\leq i,j\leq m}\left|\frac{pn\left(pn - n_j\right) + \left(pn - n_i\right)n_j}{n_i n_j}\right| \\
&\leq \max_{1\leq i,j\leq m} \frac{\left|pn + c_5\sqrt{np\log(mn)}\right|}{\left(pn - c_5\sqrt{np\log(mn)}\right)^2} c_5\sqrt{np\log(mn)} \\
&\leq c_9\sqrt{\frac{\log(mn)}{np}}
\end{aligned}$$

for some absolute constant $c_9 > 0$.

A.2 Proof of Lemma 2

The norm of $\boldsymbol{M}$ can be bounded via the moment method, which attempts to control $\operatorname{tr}(\boldsymbol{M}^k)$ for some even integer k. See [5, Section 2.3.4] for a nice introduction.

Specifically, observe that $\mathbb{E}\operatorname{tr}(\boldsymbol{M}^k)$ can be expanded as follows

$$\mathbb{E}\operatorname{tr}\left(\boldsymbol{M}^k\right) = \sum_{1\le i_1,\cdots,i_k\le n} \mathbb{E}\operatorname{tr}\left(\boldsymbol{M}_{i_1i_2}\boldsymbol{M}_{i_2i_3}\cdots\boldsymbol{M}_{i_ki_1}\right),$$

a trace sum over all k-cycles in the vertex set $\{1,\cdots,n\}$. Note that (i,i) are also treated as valid edges. For each term $\mathbb{E}\operatorname{tr}(\boldsymbol{M}_{i_1i_2}\boldsymbol{M}_{i_2i_3}\cdots\boldsymbol{M}_{i_ki_1})$, if there exists an edge occurring exactly once, then the term vanishes due to the independence assumption. Thus, it suffices to examine the terms in which each edge is repeated at least twice. Consequently, there are at most $k/2$ relevant edges, which span at most $k/2+1$ distinct vertices. We also need to assign vertices to $k/2$ edges, which adds up to no more than $(k/2)^k$ different choices.

By following the same procedure and notation as adopted in [5, Page 119], we divide all non-vanishing k-cycles into $(k/2)^k$ classes based on the above labeling order; each class is associated with j $(1\le j\le k/2)$ edges $e_1,\cdots,e_j$ with multiplicities $a_1,\cdots,a_j$, where $(e_1,\cdots,a_1,\cdots,a_j)$ determines the class of cycles and $a_1+\cdots+a_j=k$. Since there are at most n^{j+1} distinct vertices, one can see that no more than n^{j+1} cycles falling within this particular class. For notational simplicity, set $K=\sqrt{n}$, and hence $\|\boldsymbol{M}_{ij}\|\le K$. By assumption (23), one has

$$\begin{aligned}\mathbb{E}\operatorname{tr}\left(\boldsymbol{M}_{i_1i_2}\boldsymbol{M}_{i_2i_3}\cdots\boldsymbol{M}_{i_ki_1}\right) &\le m\mathbb{E}\left(\|\boldsymbol{M}_{e_1}\|^{a_1}\cdots\left\|\boldsymbol{M}_{e_j}\right\|^{a_j}\right)\\ &\le m\mathbb{E}\left\|\boldsymbol{M}_{e_1}\right\|^2\cdots\mathbb{E}\left\|\boldsymbol{M}_{e_j}\right\|^2 K^{a_1-2}\cdots K^{a_j-2}\\ &\le mK^{k-2j}.\end{aligned}$$

Thus, the total contribution of this class does not exceed

$$mn^{j+1}K^{k-2j} = mn^{\frac{k}{2}+1}.$$

By summing over all classes one obtains the crude bound

$$\mathbb{E}\operatorname{tr}\left(\boldsymbol{M}^k\right) \le m\left(\frac{k}{2}\right)^k n^{\frac{k}{2}+1},$$

which follows that

$$\mathbb{E}\left\|\boldsymbol{M}\right\|^k \le \mathbb{E}\operatorname{tr}\left(\boldsymbol{M}^k\right) \le m\left(\frac{k}{2}\right)^k n^{\frac{k}{2}+1}.$$

If we set $k=\log(mn)$, then from Markov's inequality we have

$$\mathbb{P}\left(\|\boldsymbol{M}\| \ge \frac{k}{2}n^{\frac{1}{2}+\frac{1}{k}}\left(mn\right)^{\frac{5}{k}}m^{\frac{1}{k}}\right) \le \frac{\mathbb{E}\left\|\boldsymbol{M}\right\|^k}{\left(\frac{k}{2}n^{\frac{1}{2}+\frac{1}{k}}\left(mn\right)^{\frac{5}{k}}m^{\frac{1}{k}}\right)^k} \le \frac{m\left(\frac{k}{2}\right)^k n^{\frac{k}{2}+1}}{m\left(\frac{k}{2}\right)^k n^{\frac{k}{2}+1}\left(mn\right)^5} \le \frac{1}{\left(mn\right)^5}.$$

Since $n^{\frac{1}{\log n}}=O(1)$, there exists a constant $c_0>0$ such that

$$\mathbb{P}\left(\|\boldsymbol{M}\| \ge c_0 n^{\frac{1}{2}}\log(mn)\right) \le \frac{1}{m^5n^5},$$

which completes the proof.

A.3 Proof of Lemma 3

When $\mathcal{G}\sim\mathcal{G}(n,p)$, the adjacency matrix $\boldsymbol{A}$ consists of independent Bernoulli components (except for diagonal entries), each with mean p and variance $p(1-p)$. Lemma 2 immediately implies that if $p>\frac{2\log(mn)}{n}$, then

$$\frac{1}{\sqrt{p(1-p)}}\left\|\boldsymbol{A}-p\boldsymbol{1}_n\cdot\boldsymbol{1}_n^\top\right\| \le c_0\sqrt{n}\log(mn)+1 \tag{43}$$

with probability at least $1-(mn)^{-5}$. That said, there exists an absolute constant $c_1>0$ such that

$$\left\|\boldsymbol{A}-p\boldsymbol{1}_n\cdot\boldsymbol{1}_n^\top\right\|\leq c_1\sqrt{pn}\log\left(mn\right) \tag{44}$$

with probability exceeding $1-(mn)^{-5}$.

On the other hand, from Bernstein inequality, the degree of each vertex exceeds

$$d_{\min}:=pn-c_2\sqrt{pn\log\left(mn\right)} \tag{45}$$

with probability at least $1-(mn)^{-10}$, where c_2 is some constant. When $p>\frac{2\log(mn)}{n}$, $\mathcal{G}$ is connected, and hence the least eigenvalue of $\boldsymbol{L}$ is zero with the eigenvector $\boldsymbol{1}_n$. This taken collectively with (44) and (45) suggests that when $p>\frac{c_3^2\log^2(mn)}{n}$, one has

$$a\left(\mathcal{G}\right)\geq d_{\min}-\left\|\boldsymbol{A}-p\boldsymbol{1}_n\cdot\boldsymbol{1}_n^\top\right\|\geq pn-c_3\sqrt{pn}\log\left(mn\right)$$

with high probability.

A.4 Proof of Lemma 5

Suppose that $\boldsymbol{X}^{\mathrm{gt}}+\boldsymbol{H}$ is the solution to MatchLift for some perturbation $\boldsymbol{H}\neq\boldsymbol{0}$. By Schur complement condition for positive definiteness, the feasibility constraint $\begin{bmatrix} m & \boldsymbol{1}^\top \\ \boldsymbol{1} & \boldsymbol{X}^{\mathrm{gt}}+\boldsymbol{H}\end{bmatrix}\succeq\boldsymbol{0}$ is equivalent to

$$\begin{cases}\boldsymbol{X}^{\mathrm{gt}}+\boldsymbol{H} & \succeq\boldsymbol{0},\\ \boldsymbol{X}^{\mathrm{gt}}+\boldsymbol{H}-\frac{1}{m}\boldsymbol{1}\cdot\boldsymbol{1}^\top & \succeq\boldsymbol{0},\end{cases}$$

which immediately yields

$$\mathcal{P}_{T_{\mathrm{gt}}^\perp}\left(\boldsymbol{H}\right)=\mathcal{P}_{T_{\mathrm{gt}}^\perp}\left(\boldsymbol{X}^{\mathrm{gt}}+\boldsymbol{H}\right)\succeq\boldsymbol{0}, \tag{46}$$

and

$$\left\langle\boldsymbol{d}\cdot\boldsymbol{d}^\top,\boldsymbol{H}\right\rangle=\left\langle\boldsymbol{d}\cdot\boldsymbol{d}^\top,\boldsymbol{X}^{\mathrm{gt}}-\frac{1}{m}\boldsymbol{1}\cdot\boldsymbol{1}^\top+\boldsymbol{H}\right\rangle\geq 0. \tag{47}$$

The above inequalities follow from the facts $\mathcal{P}_{T_{\mathrm{gt}}^\perp}\left(\boldsymbol{X}^{\mathrm{gt}}\right)=\boldsymbol{0}$ and $\left\langle\boldsymbol{d}\cdot\boldsymbol{d}^\top,\boldsymbol{X}^{\mathrm{gt}}-\frac{1}{m}\boldsymbol{1}\cdot\boldsymbol{1}^\top\right\rangle=0$.

From Assumption (28), one can derive

$$\begin{aligned}\left\langle\boldsymbol{Y}-\alpha\boldsymbol{d}\cdot\boldsymbol{d}^\top,\mathcal{P}_{T_{\mathrm{gt}}^\perp}\left(\boldsymbol{H}\right)\right\rangle+\left\langle\alpha\boldsymbol{d}\cdot\boldsymbol{d}^\top,\boldsymbol{H}\right\rangle&=\left\langle\boldsymbol{Y}-\alpha\boldsymbol{d}\cdot\boldsymbol{d}^\top,\boldsymbol{H}\right\rangle+\left\langle\alpha\boldsymbol{d}\cdot\boldsymbol{d}^\top,\boldsymbol{H}\right\rangle\\ &=\left\langle\boldsymbol{Y},\boldsymbol{H}\right\rangle=\sum_{i\neq j}\left\langle\boldsymbol{Y}_{ij},\boldsymbol{H}_{ij}\right\rangle. \end{aligned}\tag{48}$$

This allows us to bound

$$\begin{aligned}&\left\langle\boldsymbol{Y}-\alpha\boldsymbol{d}\cdot\boldsymbol{d}^\top,\mathcal{P}_{T_{\mathrm{gt}}^\perp}\left(\boldsymbol{H}\right)\right\rangle+\sum_{i\neq j}\left\langle\boldsymbol{Z}_{ij},\boldsymbol{H}_{ij}\right\rangle\\ \leq&\left\langle\boldsymbol{Y}-\alpha\boldsymbol{d}\cdot\boldsymbol{d}^\top,\mathcal{P}_{T_{\mathrm{gt}}^\perp}\left(\boldsymbol{H}\right)\right\rangle+\left\langle\alpha\boldsymbol{d}\cdot\boldsymbol{d}^\top,\boldsymbol{H}\right\rangle+\sum_{i\neq j}\left\langle\boldsymbol{Z}_{ij},\boldsymbol{H}_{ij}\right\rangle &(49)\\ =&\sum_{i\neq j}\left\langle\boldsymbol{Y}_{ij},\boldsymbol{H}_{ij}\right\rangle+\sum_{i\neq j}\left\langle\boldsymbol{Z}_{ij},\boldsymbol{H}_{ij}\right\rangle &(50)\\ =&\sum_{i\neq j}\left\langle\boldsymbol{W}_{ij},\boldsymbol{H}_{ij}\right\rangle, &(51)\end{aligned}$$

where the first inequality follows from (47), and the last equality follows from Assumption (27).

In order to preclude the possibility that $\boldsymbol{X}^{\mathrm{gt}}+\boldsymbol{H}$ is the solution to MatchLift, we need to show that $\sum_{i\neq j}\langle \boldsymbol{W}_{ij},\boldsymbol{H}_{ij}\rangle>0$. From (51) it suffices to establish that

$$\left\langle \boldsymbol{Y}-\alpha\boldsymbol{d}\cdot\boldsymbol{d}^{\top},\mathcal{P}_{T_{\mathrm{gt}}^{\perp}}\left(\boldsymbol{H}\right)\right\rangle+\sum_{i\neq j}\langle \boldsymbol{Z}_{ij},\boldsymbol{H}_{ij}\rangle>0 \tag{52}$$

for any feasible $\boldsymbol{H}\neq\boldsymbol{0}$. In fact, since $\boldsymbol{Y}-\alpha\boldsymbol{d}\cdot\boldsymbol{d}^{\top}$ and $\mathcal{P}_{T_{\mathrm{gt}}^{\perp}}\left(\boldsymbol{H}\right)$ are both positive semidefinite, one must have

$$\left\langle \boldsymbol{Y}-\alpha\boldsymbol{d}\cdot\boldsymbol{d}^{\top},\mathcal{P}_{T_{\mathrm{gt}}^{\perp}}\left(\boldsymbol{H}\right)\right\rangle\geq 0. \tag{53}$$

On the other hand, the constraints

$$\mathrm{supp}\left(\boldsymbol{Z}\right)\subseteq\Omega_{\mathrm{gt}}^{\perp},\quad \mathcal{P}_{\Omega_{\mathrm{gt}}^{\perp}}\left(\boldsymbol{Z}\right)\geq\boldsymbol{0},\ \text{and}\ \mathcal{P}_{\Omega_{\mathrm{gt}}^{\perp}}\left(\boldsymbol{H}\right)\geq\boldsymbol{0}$$

imply that

$$\sum_{i\neq j}\langle \boldsymbol{Z}_{ij},\boldsymbol{H}_{ij}\rangle\geq 0. \tag{54}$$

Putting (53) and (54) together gives

$$\left\langle \boldsymbol{Y}-\alpha\boldsymbol{d}\cdot\boldsymbol{d}^{\top},\mathcal{P}_{T_{\mathrm{gt}}^{\perp}}\left(\boldsymbol{H}\right)\right\rangle+\sum_{i\neq j}\langle \boldsymbol{Z}_{ij},\boldsymbol{H}_{ij}\rangle\geq 0.$$

Comparing this with (52), we only need to establish either $\left\langle \boldsymbol{Y}-\alpha\boldsymbol{d}\cdot\boldsymbol{d}^{\top},\mathcal{P}_{T_{\mathrm{gt}}^{\perp}}\left(\boldsymbol{H}\right)\right\rangle>0$ or $\sum_{i\neq j}\langle \boldsymbol{Z}_{ij},\boldsymbol{H}_{ij}\rangle>0$.

i) Suppose first that all entries of $\boldsymbol{Z}_{ij}$ ($\forall i\neq j$) in the support $\Omega_{\mathrm{gt}}^{\perp}$ are strictly positive. If the identity $\sum_{i\neq j}\langle \boldsymbol{Z}_{ij},\boldsymbol{H}_{ij}\rangle=0$ holds, then the strict positivity assumption of $\boldsymbol{Z}_{ij}$ on $\Omega_{\mathrm{gt}}^{\perp}$ as well as the constraint $\mathcal{P}_{\Omega_{\mathrm{gt}}^{\perp}}\left(\boldsymbol{H}\right)\geq\boldsymbol{0}$ immediately leads to

$$\mathcal{P}_{\Omega_{\mathrm{gt}}^{\perp}}\left(\boldsymbol{H}\right)=\boldsymbol{0}.$$

Besides, the feasibility constraint requires that $\mathcal{P}_{\Omega_{\mathrm{gt}}}\left(\boldsymbol{H}_{ij}\right)\leq\boldsymbol{0}$. If $\mathcal{P}_{\Omega_{\mathrm{gt}}}\left(\boldsymbol{H}_{ij}\right)\neq\boldsymbol{0}$, then all non-zero entries of $\boldsymbol{H}_{ij}$ are *negative*, and hence

$$\left\langle \boldsymbol{d}\cdot\boldsymbol{d}^{\top},\boldsymbol{H}\right\rangle=\left\langle \boldsymbol{d}\cdot\boldsymbol{d}^{\top},\mathcal{P}_{\Omega_{\mathrm{gt}}}\left(\boldsymbol{H}\right)\right\rangle<0,$$

which follows since all entries of $\boldsymbol{d}$ are strictly positive. This contradicts with (47). Consequently, we must either have $\boldsymbol{H}=\boldsymbol{0}$ or $\sum_{i\neq j}\langle \boldsymbol{Z}_{ij},\boldsymbol{H}_{ij}\rangle>0$. This together with (52) establishes the claim.

ii) Next, we prove the claim under Assumptions (29) and (30). In fact, Assumption (29) together with (46) asserts that $\left\langle \boldsymbol{Y},\mathcal{P}_{T_{\mathrm{gt}}^{\perp}}\left(\boldsymbol{H}\right)\right\rangle\leq 0$ can only occur if $\mathcal{P}_{T_{\mathrm{gt}}^{\perp}}\left(\boldsymbol{H}\right)=\boldsymbol{0}$. This necessarily leads to $\boldsymbol{H}=\boldsymbol{0}$, as claimed by Lemma 8.

Lemma 8. *Suppose that $\boldsymbol{X}^{\mathrm{gt}}+\boldsymbol{H}$ is feasible for MatchLift, and assume that*

$$\frac{n}{n_i}+\frac{n}{n_j}\neq\frac{n^2}{n_in_j},\quad \forall 1\leq i,j\leq m. \tag{55}$$

If $\mathcal{P}_{T_{\mathrm{gt}}^{\perp}}\left(\boldsymbol{H}\right)=\boldsymbol{0}$, then one has $\boldsymbol{H}=\boldsymbol{0}$.

Proof. See Appendix A.7. □

In summary, we can conclude that $\boldsymbol{X}^{\mathrm{gt}}$ is the unique optimizer in both cases.

A.5 Proof of Lemma 6

First, we would like to bound the operator norm of $\boldsymbol{Y}_{\mathrm{L}}$. Since each observed yet corrupted $\boldsymbol{X}_{ij}^{\mathrm{in}}$ is randomly drawn with mean $\frac{p_{\mathrm{obs}}}{m}\mathbf{1}\cdot\mathbf{1}^{\top}$, it is straightforward to see that

$$\mathbb{E}\boldsymbol{Y}^{\mathrm{L},0}=\mathbb{E}\left(-\boldsymbol{X}^{\mathrm{false}}+\frac{p_{\mathrm{obs}}}{m}\boldsymbol{E}^{\perp}\right)=\mathbf{0}.$$

By observing that $\boldsymbol{Z}^{\mathrm{L}}$ is constructed as a linear transform of $\boldsymbol{Y}^{\mathrm{L},0}$, one can also obtain

$$\mathbb{E}\boldsymbol{Z}^{\mathrm{L}}=\mathbf{0},\quad\Rightarrow\quad\mathbb{E}\boldsymbol{Y}^{\mathrm{L}}=\mathbb{E}\boldsymbol{Z}^{\mathrm{L}}+\mathbb{E}\boldsymbol{Y}^{\mathrm{L},0}=\mathbf{0}.$$

Thus, it suffices to examine the deviation of $\left\|\boldsymbol{Y}^{\mathrm{L}}\right\|$ caused by the uncertainty of $\boldsymbol{X}^{\mathrm{false}}$.

Denote by $\boldsymbol{A}^{i,j}\in\mathbb{R}^{N\times N}$ the component of $\boldsymbol{Z}^{\mathrm{L}}$ generated due to the block $-\boldsymbol{X}_{ij}^{\mathrm{false}}$, which clearly satisfies

$$\boldsymbol{Z}^{\mathrm{L}}=\boldsymbol{A}^{i,j}-\mathbb{E}\boldsymbol{A}^{i,j}.$$

For each non-zero entry of $\boldsymbol{X}_{ij}^{\mathrm{false}}$, if it encodes an incorrect correspondence between elements s and t, then it will affect no more than $8n_{s,t}$ entries in $\boldsymbol{A}^{i,j}$, where each of these entries are affected in magnitude by an amount at most $\frac{1}{n_{\mathrm{s,t}}}$. Recall that $n_{s,t}$ represents the number of sets $\mathcal{S}_i$ $(1\leq i\leq n)$ containing s and t simultaneously, which sharply concentrates within $\left[np_{\mathrm{set}}^{2}\pm O\left(\sqrt{np_{\mathrm{set}}^{2}\log(mn)}\right)\right]$ as asserted in Lemma 4. As a result, the *sum of squares* of these affected entries is bounded by

$$\frac{8n_{s,t}}{n_{s,t}^{2}}=O\left(\frac{1}{n_{s,t}}\right).\tag{56}$$

Moreover, since each row / column of $\boldsymbol{X}_{ij}^{\mathrm{false}}$ can have at most one non-zero entry, we can rearrange $\boldsymbol{A}^{i,j}$ with row / column permutation such that $\boldsymbol{A}^{i,j}$ becomes a block-diagonal matrix, where the components affected by different entries of $\boldsymbol{X}_{ij}^{\mathrm{false}}$ are separated into distinct diagonal blocks. This together with (56) leads to

$$\left\|\boldsymbol{A}^{i,j}\right\|\leq\left\|\boldsymbol{A}^{i,j}\right\|_{\mathrm{F}}\leq\max_{s\neq t}\sqrt{\frac{8}{n_{s,t}}},$$

and hence

$$\left\|\mathbb{E}\boldsymbol{A}^{i,j}\left(\boldsymbol{A}^{i,j}\right)^{\top}\right\|\leq p_{\mathrm{obs}}\left(\max_{s\neq t}\sqrt{\frac{8}{n_{s,t}}}\right)^{2}\leq\frac{c_{16}p_{\mathrm{obs}}}{np_{\mathrm{set}}^{2}}$$

for some absolute constant $c_{16}>0$, where the last inequality follows from Lemma 4.

Observe that $\boldsymbol{A}^{i,j}-\mathbb{E}\boldsymbol{A}^{i,j}$ $(i\neq j)$ are independently generated with mean zero, whose operator norm is bounded above by $2\max_{s\neq t}\sqrt{\frac{8}{n_{s,t}}}$. Applying the matrix Bernstein inequality [6, Theorem 1.4] suggests that there exist universal constants $c_5,c_6>0$ such that for any $t=O\left(\sqrt{n}\mathrm{poly}\log(mn)\right)$,

$$\mathbb{P}\left(\left\|\sum_{(i,j)\in\mathcal{G}}\boldsymbol{A}^{i,j}-\mathbb{E}\boldsymbol{A}^{i,j}\right\|>t\right)\leq n^{2}\exp\left(-\frac{\frac{1}{2}t^{2}}{n^{2}\left(\frac{c_{16}p_{\mathrm{obs}}}{np_{\mathrm{set}}^{2}}\right)+\frac{2\max_{s\neq t}\sqrt{\frac{8}{n_{s,t}}}}{3}}\right).$$

Put in another way, there exists a universal constant $c_6>0$ such that

$$\left\|\boldsymbol{Z}^{\mathrm{L}}\right\|=\left\|\sum_{i\neq j}\boldsymbol{A}^{i,j}-\mathbb{E}\boldsymbol{A}^{i,j}\right\|<c_{6}\sqrt{\frac{np_{\mathrm{obs}}}{p_{\mathrm{set}}^{2}}\log(mn)}\tag{57}$$

holds with probability exceeding $1-\frac{1}{(mn)^{10}}$. This follows from Lemma 4.

Additionally, observe that $\mathbb{E}\boldsymbol{Y}_{ij}^{\mathrm{L},0}=\boldsymbol{0}$ and

$$\left\|\frac{1}{\sqrt{p_{\mathrm{obs}}}}\boldsymbol{Y}_{ij}^{\mathrm{L},0}\right\|\leq\sqrt{n}$$

as long as $p_{\mathrm{obs}}>\frac{1}{n}$. Applying Lemma 2 suggests that

$$\left\|\boldsymbol{Y}^{\mathrm{L},0}\right\|<c_{0}\sqrt{np_{\mathrm{obs}}\log\left(mn\right)}$$

with probability at least $1-\frac{1}{(mn)^{5}}$. This combined with (57) yields

$$\left\|\boldsymbol{Y}^{\mathrm{L}}\right\|\leq\left\|\boldsymbol{Y}^{\mathrm{L},0}\right\|+\left\|\boldsymbol{Z}^{\mathrm{L}}\right\|<c_{11}\sqrt{\frac{np_{\mathrm{obs}}\log\left(mn\right)}{p_{\mathrm{set}}^{2}}}$$

with probability at least $1-\frac{3}{(mn)^{5}}$, where c_{11} is some universal constant.

On the other hand, for each (s,t) entry of $\boldsymbol{Z}_{il}^{\mathrm{L}}$ $(i\neq l)$, it can only be affected by those *observed* blocks $\boldsymbol{X}_{ij}^{\mathrm{false}}$ (or $\boldsymbol{X}_{jl}^{\mathrm{false}}$) satisfying $t\in\mathcal{S}_{j}$ (or $s\in\mathcal{S}_{j}$). Consequently, each entry of $\boldsymbol{Z}_{il}^{\mathrm{L}}$ can be expressed as a sum of $\Theta\left(np_{\mathrm{set}}p_{\mathrm{obs}}\right)$ zero-mean independent variables, each of them being bounded in magnitude by $\frac{1}{(\min_{s\neq t}n_{s,t})}$. From Hoeffding's inequality one can derive

$$\mathbb{P}\left(\left\|\boldsymbol{Z}_{il}^{\mathrm{L}}\right\|_{\infty}>t\right)\leq m^{2}\mathbb{P}\left(-\frac{t^{2}}{c_{7}np_{\mathrm{set}}p_{\mathrm{obs}}\frac{1}{\left(\min\limits_{s\neq t}n_{s,t}\right)^{2}}}\right)\leq m^{2}\mathbb{P}\left(-\frac{t^{2}}{\tilde{c}_{7}p_{\mathrm{obs}}\frac{1}{np_{\mathrm{set}}^{3}}}\right)$$

for some constants $c_{7},\tilde{c}_{7}>0$, indicating that

$$\left\|\boldsymbol{Z}_{il}^{\mathrm{L}}\right\|_{\infty}\leq\sqrt{\frac{c_{8}p_{\mathrm{obs}}\log\left(mn\right)}{np_{\mathrm{set}}^{3}}},\quad\forall i\neq l$$

with probability exceeding $1-\frac{1}{(mn)^{10}}$.

A.6 Proof of Lemma 7

The matrix $\boldsymbol{Y}^{\mathrm{true},1}$ can be decomposed into two parts $\overline{\boldsymbol{Y}}^{\mathrm{true},1}$ and $\tilde{\boldsymbol{Y}}^{\mathrm{true},1}$. Here, $\overline{\boldsymbol{Y}}^{\mathrm{true},1}$ consists of all components satisfying $\boldsymbol{X}_{ij}^{\mathrm{in}}=\boldsymbol{X}_{ij}^{\mathrm{gt}}$, while $\tilde{\boldsymbol{Y}}^{\mathrm{true},1}$ consists of all $\boldsymbol{X}_{ij}^{\mathrm{in}}$ that are random outliers.

By Lemma 3, one can verify that for all unit vector $\boldsymbol{v}$ such that $\boldsymbol{v}\boldsymbol{v}^{\top}\in T_{\mathrm{gt}}^{\perp}$,

$$\begin{aligned}\left\langle\boldsymbol{v}\boldsymbol{v}^{\top},\overline{\boldsymbol{Y}}^{\mathrm{true},1}\right\rangle&\geq\left(1-\frac{p_{\mathrm{obs}}}{m}\right)\min_{1\leq s\leq m}\left(n_{s}p_{\mathrm{true}}p_{\mathrm{obs}}-c_{4}\sqrt{n_{s}p_{\mathrm{obs}}}\log\left(mn\right)\right)\\&\geq\frac{1}{2}np_{\mathrm{set}}p_{\mathrm{true}}p_{\mathrm{obs}}-c_{5}\sqrt{np_{\mathrm{set}}p_{\mathrm{obs}}}\log\left(mn\right)\end{aligned}\tag{58}$$

for some constant $c_{5}>0$. Here, we have made use of the concentration result stated in Lemma 4. In addition, each non-zero entry of $\tilde{\boldsymbol{Y}}^{\mathrm{true},1}$ has mean zero and variance $\frac{p_{\mathrm{obs}}}{m}\left(1-\frac{p_{\mathrm{obs}}}{m}\right)$. Consequently, applying Lemma 2 gives

$$\left\|\tilde{\boldsymbol{Y}}^{\mathrm{true},1}\right\|\leq c_{15}\max_{1\leq s\leq m}\sqrt{p_{\mathrm{obs}}n_{s}}\log\left(nm\right)<\tilde{c}_{15}\sqrt{np_{\mathrm{set}}p_{\mathrm{obs}}}\log\left(nm\right).$$

This taken collectively with (58) yields that

$$\left\langle\boldsymbol{v}\boldsymbol{v}^{\top},\boldsymbol{Y}^{\mathrm{true},1}\right\rangle\geq\frac{1}{2}np_{\mathrm{set}}p_{\mathrm{true}}p_{\mathrm{obs}}-\left(c_{5}+\tilde{c}_{15}\right)\sqrt{np_{\mathrm{set}}p_{\mathrm{obs}}}\log\left(mn\right).\tag{59}$$

On the other hand, we know from the construction procedure and Lemma 1 that

$$\|\boldsymbol{R}^{\mathrm{m}}\|_{\infty} \leq \sqrt{\frac{c_{10} p_{\mathrm{obs}} \log(mn)}{n p_{\mathrm{set}}^{3}}}$$

for some constant $c_{10} > 0$, and a crude upper bound yields

$$\left\|\boldsymbol{Y}^{\mathrm{true},2}\right\| \leq \left\|\boldsymbol{Y}^{\mathrm{true},2}\right\|_{1} \leq c_{11} n p_{\mathrm{set}} \sqrt{\frac{p_{\mathrm{obs}} \log(mn)}{n p_{\mathrm{set}}^{3}}} = c_{11} \sqrt{\frac{n p_{\mathrm{obs}} \log(mn)}{p_{\mathrm{set}}}}$$

for some universal constant $c_{11} > 0$.

A.7 Proof of Lemma 8

Define an augmented matrix $\boldsymbol{H}^{\mathrm{sup}}$ such that

$$\boldsymbol{H}_{ij}^{\mathrm{sup}} = \boldsymbol{\Pi}_{i}^{\top} \boldsymbol{H}_{ij} \boldsymbol{\Pi}_{j}. \tag{60}$$

Recall that n_i denotes the number of sets containing element i, and that

$$\boldsymbol{\Sigma} := \begin{bmatrix} \frac{n}{n_1} & & & \\ & \frac{n}{n_2} & & \\ & & \ddots & \\ & & & \frac{n}{n_m} \end{bmatrix}.$$

The assumption that $\mathcal{P}_{T_{\mathrm{gt}}^{\perp}}(\boldsymbol{H}) = \boldsymbol{0}$ can be translated into

$$\left(\boldsymbol{I} - \frac{1}{n}(\boldsymbol{1}_n \otimes \boldsymbol{I}_m) \boldsymbol{\Sigma} (\boldsymbol{1}_n \otimes \boldsymbol{I}_m)\right) \boldsymbol{H}^{\mathrm{sup}} \left(\boldsymbol{I} - \frac{1}{n}(\boldsymbol{1}_n \otimes \boldsymbol{I}_m) \boldsymbol{\Sigma} (\boldsymbol{1}_n \otimes \boldsymbol{I}_m)\right) = \boldsymbol{0}.$$

We can easily compute that

$$\boldsymbol{H}_{ii}^{\mathrm{sup}} - \boldsymbol{\Sigma} \overline{\boldsymbol{H}}_{\cdot i}^{\mathrm{sup}} - \overline{\boldsymbol{H}}_{i \cdot}^{\mathrm{sup}} \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \overline{\boldsymbol{H}}_{\cdot\cdot}^{\mathrm{sup}} \boldsymbol{\Sigma} = \boldsymbol{0}, \quad 1 \leq i \leq n,$$

where

$$\begin{cases} \overline{\boldsymbol{H}}_{\cdot i}^{\mathrm{sup}} & := \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{H}_{ji}^{\mathrm{sup}}, \\ \overline{\boldsymbol{H}}_{i \cdot}^{\mathrm{sup}} & := \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{H}_{ij}^{\mathrm{sup}}, \\ \overline{\boldsymbol{H}}_{\cdot\cdot}^{\mathrm{sup}} & := \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{H}_{ij}^{\mathrm{sup}}. \end{cases}$$

This combined with the identity $\boldsymbol{H}_{ii} = \boldsymbol{0}$ (and hence $\boldsymbol{H}_{ii}^{\mathrm{sup}} = \boldsymbol{0}$) yields

$$\boldsymbol{\Sigma} \overline{\boldsymbol{H}}_{\cdot\cdot}^{\mathrm{sup}} \boldsymbol{\Sigma} = \boldsymbol{\Sigma} \overline{\boldsymbol{H}}_{\cdot i}^{\mathrm{sup}} + \overline{\boldsymbol{H}}_{i \cdot}^{\mathrm{sup}} \boldsymbol{\Sigma}, \quad 1 \leq i \leq n.$$

Summing over all i leads to

$$\boldsymbol{\Sigma} \overline{\boldsymbol{H}}_{\cdot\cdot}^{\mathrm{sup}} \boldsymbol{\Sigma} = \boldsymbol{\Sigma} \left(\frac{1}{n} \sum_{i=1}^{n} \overline{\boldsymbol{H}}_{\cdot i}^{\mathrm{sup}}\right) + \left(\frac{1}{n} \sum_{i=1}^{n} \overline{\boldsymbol{H}}_{i \cdot}^{\mathrm{sup}}\right) \boldsymbol{\Sigma} = \boldsymbol{\Sigma} \overline{\boldsymbol{H}}_{\cdot\cdot}^{\mathrm{sup}} + \overline{\boldsymbol{H}}_{\cdot\cdot}^{\mathrm{sup}} \boldsymbol{\Sigma}.$$

Expanding it yields

$$\frac{n^2}{n_i n_j} \left(\overline{\boldsymbol{H}}_{\cdot\cdot}^{\mathrm{sup}}\right)_{i,j} = \left(\frac{n}{n_i} + \frac{n}{n_j}\right) \left(\overline{\boldsymbol{H}}_{\cdot\cdot}^{\mathrm{sup}}\right)_{i,j}, \quad 1 \leq i, j \leq m.$$

From our assumption that $\frac{n^2}{n_i n_j} \neq \frac{n}{n_i} + \frac{n}{n_j}$, we can derive

$$\overline{\boldsymbol{H}}_{\cdot\cdot}^{\mathrm{sup}} = \boldsymbol{0}. \tag{61}$$

Due to the feasibility constraint, all diagonal entries of $\boldsymbol{H}_{ij}^{\mathrm{sup}}$ are non-positive, and all off-diagonal entries of $\boldsymbol{H}_{ij}^{\mathrm{sup}}$ are non-negative. These conditions together with (61) establish that $\boldsymbol{H} = \boldsymbol{0}$.

A.8 Proof of Lemma 9

Lemma 9. *Given any set of n permutation matrices $\boldsymbol{P}_i \in \mathbb{R}^{m\times m}$ $(1 \leq i \leq n)$, generate a random matrix $\boldsymbol{M}$ via the following procedure.*

1. *Generate a symmetric block matrix $\boldsymbol{A} = [\boldsymbol{A}_{ij}]_{1\leq i,j\leq n}$ such that $\boldsymbol{A}_{ii} = \boldsymbol{I}$ for all $1 \leq i \leq n$, and for all $i < j$,*
$$\boldsymbol{A}_{ij} = \begin{cases} \boldsymbol{0}, & \text{if } \mu_{ij} = 0, \\ \boldsymbol{P}_i \boldsymbol{P}_j^\top, & \text{if } \nu_{ij} = 1 \text{ and } \mu_{ij} = 1, \\ \boldsymbol{U}_{ij}, & \text{else,} \end{cases} \tag{62}$$
where $\nu_{ij} \sim \text{Bernoulli}(p)$ and $\mu_{ij} \sim \text{Bernoulli}(\tau)$ are independent binary variables, and $\boldsymbol{U}_{ij} \in \mathbb{R}^{m\times m}$ are independent random permutation matrices obeying $\mathbb{E}\boldsymbol{U}_{ij} = \frac{1}{m}\mathbf{1}_m \cdot \mathbf{1}_m^\top$.

2. *$\boldsymbol{M}$ is a principal minor of $\boldsymbol{A}$ from rows / columns at indices from a set $I \subseteq \{1, 2, \cdots, mn\}$, where each $1 \leq i \leq mn$ is contained in I independently with probability q.*

Then there exist absolute constants $c_1, c_2 > 0$ such that if $p \geq c_1 \frac{\log^2(mn)}{q\sqrt{\tau n}}$, one has
$$\begin{cases} \lambda_i(\boldsymbol{M}) & \geq \left(1 - \frac{1}{\log(mn)}\right)\tau pqn, \qquad \text{if } 1 \leq i \leq m \\ \lambda_i(\boldsymbol{M}) & \leq c_2\sqrt{\tau n}\log(mn) < \frac{\tau pqn}{\log(mn)}, \quad \text{if } i > m \end{cases} \tag{63}$$
with probability exceeding $1 - \frac{1}{m^5 n^5}$. Here, $\lambda_i(\boldsymbol{M})$ represents the ith largest eigenvalue of $\boldsymbol{M}$.

Proof. See Appendix A.8. □

Without loss of generality, we assume that $\boldsymbol{P}_i = \boldsymbol{I}_m$ for all $1 \leq i \leq n$, since row / column permutation $\boldsymbol{A}$ does not change its eigenvalues. For convenience of presentation, we write $\boldsymbol{A} = \boldsymbol{Y} + \boldsymbol{Z}$ such that for all $1 \leq i \leq j \leq n$:
$$\boldsymbol{Y}_{ij} = \begin{cases} \boldsymbol{0}, & \text{if } \mu_{ij} = 0, \\ \boldsymbol{I}_m, & \text{if } \nu_{ij} = 1 \text{ and } \mu_{ij} = 1, \\ \boldsymbol{U}_{ij}, & \text{else,} \end{cases} \tag{64}$$
and
$$\boldsymbol{Z}_{ij} = \begin{cases} \boldsymbol{I}_m, & \text{if } i = j \text{ and } \mu_{ij} = 0, \\ \boldsymbol{I}_m - \boldsymbol{U}_{ii}, & \text{if } i = j,\ \mu_{ij} = 1 \text{ and } \nu_{ii} = 0, \\ \boldsymbol{0}, & \text{else.} \end{cases} \tag{65}$$
Apparently, $\boldsymbol{Z}$ is a block diagonal matrix satisfying
$$\|\boldsymbol{Z}\| \leq 2, \tag{66}$$
which is only a mild perturbation of $\boldsymbol{Y}$. Thus, one has reduced to the case when all blocks are i.i.d., which is slightly easier to analyze.

Decompose $\boldsymbol{Y}$ into 2 components $\boldsymbol{Y} = \boldsymbol{Y}^{\text{mean}} + \boldsymbol{Y}^{\text{var}}$ such that for all $1 \leq i \leq j \leq n$,
$$\boldsymbol{Y}_{ij}^{\text{mean}} = \tau\left(\frac{(1-p)}{m}\mathbf{1}_m \cdot \mathbf{1}_m^\top + p\boldsymbol{I}_m\right), \tag{67}$$
$$\boldsymbol{Y}_{ij}^{\text{var}} = \begin{cases} -\tau\left(\frac{(1-p)}{m}\mathbf{1}_m \cdot \mathbf{1}_m^\top + p\boldsymbol{I}_m\right), & \text{if } \mu_{ij} = 0, \\ (1-\tau p)\boldsymbol{I}_m - \frac{(1-p)}{m}\mathbf{1}_m \cdot \mathbf{1}_m^\top, & \text{if } \nu_{ij} = 1 \text{ and } \mu_{ij} = 1, \\ \boldsymbol{U}_{ij} - \tau\left(\frac{(1-p)}{m}\mathbf{1}_m \cdot \mathbf{1}_m^\top + p\boldsymbol{I}_m\right), & \text{else.} \end{cases} \tag{68}$$
In other words, $\boldsymbol{Y}^{\text{mean}}$ represents the mean component of $\boldsymbol{Y}$, while $\boldsymbol{Y}^{\text{var}}$ comprises all variations. It is straightforward to check that $\boldsymbol{Y}^{\text{mean}} \succeq \boldsymbol{0}$ and $\text{rank}(\boldsymbol{Y}^{\text{mean}}) \leq m + 1$. If we denote by $\boldsymbol{Y}_I^{\text{mean}}$ the principal

minor coming from the rows and columns of $\boldsymbol{Y}$ at indices from I, then from Weyl's inequality one can easily see that

$$\lambda_i\left(\boldsymbol{M}\right) \geq \lambda_i\left(\boldsymbol{Y}_I^{\mathrm{mean}}\right) - \|\boldsymbol{Y}^{\mathrm{var}}\| - \|\boldsymbol{Z}\| \geq \lambda_i\left(\boldsymbol{Y}_I^{\mathrm{mean}}\right) - \|\boldsymbol{Y}^{\mathrm{var}}\| - 2, \quad 1 \leq i \leq m \tag{69}$$

and

$$\lambda_i\left(\boldsymbol{M}\right) \leq \|\boldsymbol{Y}^{\mathrm{var}}\| + \|\boldsymbol{Z}\| \leq \|\boldsymbol{Y}^{\mathrm{var}}\| + 2, \qquad i > m. \tag{70}$$

In light of this, it suffices to evaluate $\|\boldsymbol{Y}^{\mathrm{var}}\|$ as well as the eigenvalues of $\boldsymbol{Y}_I^{\mathrm{mean}}$.

We are now in position to quantify the eigenvalues of $\boldsymbol{Y}_I^{\mathrm{mean}}$. Without affecting its eigenvalue distribution, one can perform row / column permutation of $\boldsymbol{Y}_I^{\mathrm{mean}}$ so that

$$\boldsymbol{Y}_I^{\mathrm{mean}} \overset{(\mathrm{permutation})}{=} \tau p \begin{bmatrix} \mathbf{1}_{n_1} \cdot \mathbf{1}_{n_1}^\top & & \\ & \ddots & \\ & & \mathbf{1}_{n_m} \cdot \mathbf{1}_{n_m}^\top \end{bmatrix} + \frac{\tau(1-p)}{m} \mathbf{1}_N \cdot \mathbf{1}_N^\top. \tag{71}$$

Here, n_i $(1 \leq i \leq m)$ denotes the cardinality of a set I_i generated by independently sampling n elements each with probability q, and we set $N := n_1 + \cdots + n_m$ for simplicity. From Bernstein inequality, there exist universal constants $c_5, c_6 > 0$ such that if $q > \frac{c_5 \log(mn)}{n}$, then

$$|n_i - nq| \leq c_6 \sqrt{nq \log(mn)}, \quad 1 \leq i \leq m \tag{72}$$

holds with probability exceeding $1 - (mn)^{-10}$.

Since $\boldsymbol{Y}_I^{\mathrm{mean}}$ is positive semidefinite, from (71) one can see that all non-zero eigenvalues of $\boldsymbol{Y}_I^{\mathrm{mean}}$ are also eigenvalues of the following $(m+1) \times (m+1)$ matrix

$$\begin{aligned}
\overline{\boldsymbol{Y}}_I^{\mathrm{mean}} &:= \tau \begin{bmatrix} \sqrt{p}\mathbf{1}_{n_1}^\top & & & \\ & \sqrt{p}\mathbf{1}_{n_2}^\top & & \\ & & \ddots & \\ & & & \sqrt{p}\mathbf{1}_{n_m}^\top \\ & \sqrt{\frac{1-p}{m}}\mathbf{1}_N^\top & & \end{bmatrix} \begin{bmatrix} \sqrt{p}\mathbf{1}_{n_1} & & & & \\ & \sqrt{p}\mathbf{1}_{n_2} & & & \\ & & \ddots & & \sqrt{\frac{1-p}{m}}\mathbf{1}_N \\ & & & \sqrt{p}\mathbf{1}_{n_m} & \end{bmatrix} \\
&= \tau \begin{bmatrix} pn_1 & & & & \sqrt{\frac{p(1-p)}{m}}n_1 \\ & pn_2 & & & \sqrt{\frac{p(1-p)}{m}}n_2 \\ & & \ddots & & \vdots \\ & & & pn_m & \sqrt{\frac{p(1-p)}{m}}n_m \\ \sqrt{\frac{p(1-p)}{m}}n_1 & \sqrt{\frac{p(1-p)}{m}}n_2 & \cdots & \sqrt{\frac{p(1-p)}{m}}n_m & \frac{1-p}{m}N \end{bmatrix} \qquad (73) \\
&= \underbrace{\tau qn \begin{bmatrix} p & & & \sqrt{\frac{p(1-p)}{m}} \\ & \ddots & & \vdots \\ & & p & \sqrt{\frac{p(1-p)}{m}} \\ \sqrt{\frac{p(1-p)}{m}} & \cdots & \sqrt{\frac{p(1-p)}{m}} & 1-p \end{bmatrix}}_{\overline{\boldsymbol{Y}}_{I,0}} + \underbrace{\tau \begin{bmatrix} p\Delta_1 & & & \sqrt{\frac{p(1-p)}{m}}\Delta_1 \\ & \ddots & & \vdots \\ & & p\Delta_m & \sqrt{\frac{p(1-p)}{m}}\Delta_m \\ \sqrt{\frac{p(1-p)}{m}}\Delta_1 & \cdots & \sqrt{\frac{p(1-p)}{m}}\Delta_m & \frac{1-p}{m}\Delta_N \end{bmatrix}}_{\overline{\boldsymbol{Y}}_{I,\Delta}} \qquad (74)
\end{aligned}$$

where $\Delta_i = n_i - nq$ $(1 \leq i \leq m)$, and $\Delta_N = N - qnm$ which satisfies $|\Delta_N| \leq m \max_{1 \leq i \leq m} |\Delta_i|$. By Schur complement condition [7], if $\begin{bmatrix} \boldsymbol{C} & \boldsymbol{B} \\ \boldsymbol{B}^\top & \boldsymbol{D} \end{bmatrix} \succ \mathbf{0}$, then $\boldsymbol{C} \succ \mathbf{0}$ and $\boldsymbol{D} - \boldsymbol{B}^\top \boldsymbol{C}^{-1} \boldsymbol{B} \succ \mathbf{0}$. Applying this condition to $\overline{\boldsymbol{Y}}_{I,0}$ suggests that if $\overline{\boldsymbol{Y}}_{I,0} \succ \mathbf{0}$, then

$$(1-p) - \frac{p(1-p)}{m} \frac{1}{p} \mathbf{1}_m^\top \cdot \mathbf{1}_m > 0,$$

which is contradictory since $(1-p) - \frac{p(1-p)}{m} \frac{1}{p} \mathbf{1}_m^\top \cdot \mathbf{1}_m = 0$. Thus, $\overline{\boldsymbol{Y}}_{I,0}$ is rank deficient. Apparently, $\overline{\boldsymbol{Y}}_{I,0}$ has at least m eigenvalues equal to τqpn, which then suggests that $\lambda_{m+1}\left(\overline{\boldsymbol{Y}}_{I,0}\right) = 0$ and

$$\lambda_i\left(\overline{\boldsymbol{Y}}_{I,0}\right) = \tau pqn, \quad 1 \leq i \leq m. \tag{75}$$

Besides, the residual component $\overline{\boldsymbol{Y}}_{I,\Delta}$ can be bounded as follows

$$\begin{aligned}\left\|\overline{\boldsymbol{Y}}_{I,\Delta}\right\| &\leq \tau\left\|\begin{bmatrix} p\Delta_1 & & & \\ & \ddots & & \\ & & p\Delta_m & \\ & & & \frac{1-p}{m}\Delta_N \end{bmatrix}\right\| + \tau\left\|\begin{bmatrix} 0 & & & \sqrt{\frac{p(1-p)}{m}}\Delta_1 \\ & \ddots & & \vdots \\ & & 0 & \sqrt{\frac{p(1-p)}{m}}\Delta_m \\ \sqrt{\frac{p(1-p)}{m}}\Delta_1 & \cdots & \sqrt{\frac{p(1-p)}{m}}\Delta_m & 0 \end{bmatrix}\right\|_{\mathrm{F}} \\ &\leq \tau\max\left\{p\max_{1\leq i\leq m}\left|\Delta_i\right|, \frac{1-p}{m}\left|\Delta_N\right|\right\} + \tau\sqrt{2p\left(1-p\right)}\max_{1\leq i\leq m}\left|\Delta_i\right| \\ &\leq 2\tau\max_{1\leq i\leq m}\left|\Delta_i\right| \leq 2c_6\tau\sqrt{nq\log\left(mn\right)},\end{aligned}$$

where the last inequality follows from (72). This taken collectively with (74) and (75) yields that: when $p > \frac{2c_6\log^2(mn)}{\sqrt{nq}}$ or, equivalently, when $2c_6\sqrt{nq\log\left(mn\right)} < \frac{1}{\log^{1.5}(mn)}npq$, one has

$$\begin{cases}\lambda_i\left(\boldsymbol{Y}_I^{\text{mean}}\right) & \geq \left(1-\frac{1}{\log^{\frac{3}{2}}\left(mn\right)}\right)\tau pqn, \qquad 1\leq i\leq m, \\ \lambda_i\left(\boldsymbol{Y}_I^{\text{mean}}\right) & \leq 2c_6\tau\sqrt{nq\log\left(mn\right)} \leq \frac{1}{\log^{\frac{3}{2}}\left(mn\right)}\tau pqn, \qquad i>m.\end{cases} \tag{76}$$

Finally, observe that $\mathbb{E}\boldsymbol{Y}_{ij}^{\text{var}} = 0$, $\mathbb{E}\left\|\frac{1}{2\sqrt{\tau}}\boldsymbol{Y}_{ij}^{\text{var}}\right\|^2 \leq 1$, and $\frac{1}{2\sqrt{\tau}}\left\|\boldsymbol{Y}_{ij}^{\text{var}}\right\| \leq \frac{1}{\sqrt{\tau}}$. When $\tau > \frac{1}{n}$, Lemma 2 yields that

$$\left\|\boldsymbol{Y}^{\text{var}}\right\| \leq 2c_0\sqrt{\tau n}\log\left(mn\right) \tag{77}$$

with probability at least $1-(mn)^{-5}$. The claim then follows by substituting (76) and (77) into (69) and (70).

References

[1] A. Jalali, Y. Chen, S. Sanghavi, and H. Xu, "Clustering partially observed graphs via convex optimization," *International Conf. on Machine Learning (ICML)*, 2011. 3, 1.2, 1.3

[2] A. Jalali and N. Srebro, "Clustering using max-norm constrained optimization," *International Conference on Machine Learning (ICML)*, June 2012. 1.2, 1.3

[3] Y. HaCohen, E. Shechtman, D. Goldman, and D. Lischinski, "Non-rigid dense correspondence with applications for image enhancement," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 70:1–70:10, Jul. 2011. 1.4

[4] N. Alon and J. H. Spencer, *The probabilistic method (3rd Edition)*. Wiley, 2008. 4.1

[5] T. Tao, *Topics in random matrix theory.* AMS Bookstore, 2012, vol. 132. A.2

[6] J. A. Tropp, "User-friendly tail bounds for sums of random matrices," *Foundations of Computational Mathematics*, vol. 12, no. 4, pp. 389–434, 2012. A.5

[7] S. P. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge university press, 2004. A.8