

# WELL-INVERTIBLE COLUMN SUBSETS OF SPARSE MATRICES ARE RARE

HAN HUANG, MARK RUDELSON, AND KONSTANTIN TIKHOMIROV

**ABSTRACT.** A random  $n \times k$  matrix  $S$  is an  $(r, \alpha)$ -oblivious subspace injection (OSI) if  $\mathbb{E}\|S^\top x\|_2^2 = \|x\|_2^2$  for every  $x \in \mathbb{R}^n$ , and for every fixed  $r$ -dimensional subspace  $V \subset \mathbb{R}^n$ , with probability close to one, one has  $\alpha\|x\|_2^2 \leq \|S^\top x\|_2^2$  for all  $x \in V$ . In this work, we show that in the regime  $r = \Omega(k)$  and  $\alpha = \Omega(1)$ , and under a mild additional structural assumption, no constant-row-sparsity matrix  $S$  is OSI, thereby answering, in a strong form, a question raised by Camaño, Epperly, Meyer, and Tropp.

We show that the failure of the OSI property for sparse random matrices stems from a general deterministic phenomenon, thereby reducing a probabilistic problem to a non-probabilistic one. This phenomenon is related to the restricted invertibility principle introduced in the seminal work of Bourgain–Tzafriri. Let  $(n_k)_{k \in \mathbb{N}}$  be a sequence of integers satisfying  $\frac{n_k}{k} \rightarrow \infty$ . For each  $k$ , let  $S^{(k)}$  be a  $n_k \times k$  non-random matrix with  $O(1)$  nonzero entries per row, whose nonzero entries have average magnitude  $O(1)$ , and such that the total number of pairs of rows with supports overlapping at two or more indices is  $o(n_k^2/k)$ . We prove that for every constant  $\varepsilon > 0$ , as  $k \rightarrow \infty$ , the overwhelming majority of  $k \times \lfloor \varepsilon k \rfloor$  submatrices of  $(S^{(k)})^\top$  have the smallest singular value  $o(1)$ . Thus, the well-invertible submatrices whose existence is guaranteed by the Bourgain–Tzafriri theorem are rare. The proof is itself based on probabilistic tools.

## CONTENTS

|                                                                    |    |
|--------------------------------------------------------------------|----|
| 1. Introduction                                                    | 1  |
| 1.1. Motivation: oblivious subspace injections                     | 1  |
| 1.2. Well-invertible column subsets                                | 3  |
| 1.3. Technical overview                                            | 5  |
| 1.4. Organization                                                  | 8  |
| 2. From matrices to graphs                                         | 8  |
| 2.1. Graph Conventions and Matching Rooted Trees                   | 9  |
| 2.2. Weighted Graphs                                               | 10 |
| 3. The $C_4$ -free Sampling Theorem                                | 11 |
| 3.1. Setup and Proof Roadmap                                       | 11 |
| 3.2. Three Finite Ingredients                                      | 13 |
| 3.3. The Constant-Ratio $C_4$ -free Case                           | 17 |
| 3.4. The Finite $C_4$ -free Case                                   | 19 |
| 4. Trichotomy and Sampling                                         | 21 |
| 4.1. From Local Four-Cycle Sparsity to a Good $C_4$ -free Subgraph | 21 |
| 4.2. Average-Degree Trichotomy                                     | 23 |
| 5. Proof of Theorem 1.3                                            | 28 |
| 6. SparseStack and OSI                                             | 29 |
| 7. Open Problems                                                   | 30 |
| References                                                         | 31 |

## 1. INTRODUCTION

**1.1. Motivation: oblivious subspace injections.** Randomized sketching has become a fundamental technique in modern numerical linear algebra for lowering the computational cost of large-scale problems; see, for example, [15, 28, 18]. With randomized sketching, a high-dimensional dataset is replaced by a much smaller random projection obtained by applying a dimension reduction via a *sketch matrix*. With an appropriately constructed sketch matrix, the compressed representation retains the key geometric features

of the original data, making it possible to compute fast approximate solutions with rigorous guarantees. A by-now classical notion in randomized sketching is that of *oblivious subspace embedding* (OSE). A random  $n \times k$  sketch matrix  $S$  is an oblivious subspace embedding with parameters  $r$  and  $\varepsilon$ , if for every non-random subspace  $V$  of  $\mathbb{R}^n$  of dimension  $r$ , the linear operator  $S^\top$  restricted to  $V$  acts as a  $(1 + \varepsilon)$ -isometry. Sparse Euclidean subspace embeddings were developed, in particular, in [9, 19, 22, 21, 10, 6, 7, 8]; we refer to bibliography overview in [8] for an account of this very active line of research.

Rather than two-sided bounds on distortion considered in the OSE setting, one can restrict attention to only quantifying operator's injectivity. The motivation for our note comes from this framework, systematically explored by Camaño, Epperly, Meyer, and Tropp under the name of *oblivious subspace injection* (OSI) [5]. The work [5] shows that several randomized linear algebra algorithms succeed under only the OSI hypothesis, as long as one tolerates constant multiplicative approximation error (at the same time, OSI has certain limitations which are not discussed here; see [25]).

**Definition 1.1** (Oblivious subspace injection property; [5, Definition 1.2]). *Fix integers  $r \leq k \leq n$  and a parameter  $\alpha \in (0, 1]$ . A random  $n \times k$  matrix  $S$  is  $(r, \alpha)$ -OSI if it satisfies the following two conditions:*

- (1) Isotropy: *for every vector  $x \in \mathbb{R}^n$ ,*

$$\mathbb{E}\|S^\top x\|_2^2 = \|x\|_2^2.$$

- (2) Injectivity: *for every fixed  $r$ -dimensional subspace  $V \subset \mathbb{R}^n$ , with probability at least  $9/10$ ,*

$$\alpha\|x\|_2^2 \leq \|S^\top x\|_2^2 \quad \forall x \in V.$$

*The probability  $9/10$  in the second condition can be replaced with another number close to one.*

Given a design matrix  $A \in \mathbb{R}^{n \times r}$  and a response matrix  $B \in \mathbb{R}^{n \times m}$ , multivariate regression problem asks to solve  $\min_{X \in \mathbb{R}^{r \times m}} \|AX - B\|_F^2$ . Instead of solving this problem directly, one can draw a random  $n \times k$  sketch matrix  $S$  and compute  $\tilde{X} \in \arg \min_{X \in \mathbb{R}^{r \times m}} \|S^\top (AX - B)\|_F^2$  by taking  $\tilde{X} := (S^\top A)^\dagger S^\top B$ . The authors of [5] show that if  $S$  is a  $(r, \alpha)$ -OSI, then with positive constant probability

$$\|A\tilde{X} - B\|_F^2 \leq \frac{C}{\alpha} \min_X \|AX - B\|_F^2$$

for a universal constant  $C$  [5, Theorem 2.7]. Furthermore, if  $S$  is sparse with  $\zeta$  nonzeros per row, the sketches  $S^\top A$  and  $S^\top B$  can be formed in  $O(\zeta n(r + m))$  arithmetic operations for dense inputs, instead of the  $O(kn(r + m))$  operations required by a dense  $n \times k$  sketch matrix.

Another example is the single-pass Nystrom approximation for a positive semidefinite matrix. Given a PSD  $A \in \mathbb{R}^{n \times n}$  and a sketch matrix  $S \in \mathbb{R}^{n \times k}$ , one forms the sketch  $Y = AS$  and returns the rank- $k$  positive semidefinite approximation  $\hat{A} = Y(S^\top Y)^\dagger Y^\top$ . The authors of [5] prove that if  $S$  is an  $(r, \alpha)$ -OSI with  $k \geq r$ , then, with probability at least  $9/10$ ,

$$\|A - \hat{A}\|_* \leq \frac{C}{\alpha} \|A - [[A]]_r\|_*$$

for a universal constant  $C$  [5, Corollary 2.5], where  $\|\cdot\|_*$  denotes the nuclear norm, and  $[[A]]_r$  is the best rank  $r$  approximation of  $A$  in Frobenius norm. For a sparse OSI with  $\zeta$  nonzeros per row, forming the Nystrom sketch  $Y = AS$  for dense matrix  $A$  costs  $O(\zeta n^2)$ .

In view of the above discussion, the optimal relation between OSI parameters — sketch matrix sparsity and dimensions, as well as dimension of the oblivious subspace, and the magnitude of the injectivity parameter  $\alpha$  — are of central importance. As an example of recent developments, Tropp [26] proved that a  $n \times k$  SparseStack map  $S$  with row sparsity  $\zeta \geq 6\delta^{-1} \log(Cr)$  and with  $k \geq 16\delta^{-2}r$ , is  $(r, 1 - \delta)$ -OSI. Which sparsity level is sufficient for the OSI property is a compelling open problem. In [5], the authors conjectured that constant sparsity *SparseStack* matrices (to be defined later) are  $(r, 1/2)$ -OSI for some embedding dimension  $k = O(r)$ . As a much more general problem, in this paper we consider

**Problem 1.2** (Constant sparsity OSI matrices). *Can an  $n \times k$  random sketch matrix  $S$  with constant row sparsity have  $(r, \alpha)$ -OSI property for constant  $\alpha$  and with  $k = O(r)$ ?*

The main result of this work shows that, under very broad assumptions including, in particular, SparseStack model, constant sparsity sketch matrices *do not* possess OSI property for proportional oblivious subspace

dimension and with constant injectivity parameter. Moreover, we show that the injectivity property (2) in Definition 1.1 fails even if we restrict the test spaces  $V \in \mathbb{R}^n$  to  $r$ -dimensional *coordinate* subspaces.

Problem 1.2 is probabilistic. However, as we explain in the next section, the failure of the OSI property described above is due to a general phenomenon pertaining to deterministic matrices. This phenomenon called *restricted invertibility* is a classical object of study in geometric functional analysis. Roughly speaking, it states that a matrix which is far away from the set of low-rank matrices has a well-invertible submatrix. This phenomenon will be discussed in more detail in the next section. Its relevance to Problem 1.2 follows from our main result, which shows that, under mild conditions, a sparse matrix has very few such well-invertible submatrices.

**1.2. Well-invertible column subsets.** The seminal *restricted invertibility* theorem of Bourgain–Tzafriri [1, 2, 3] states that there are universal constants  $c, \gamma > 0$  such that every  $n \times n$  matrix  $A$  with columns of Euclidean norm one contains a column index set  $\sigma \subset [n]$  with  $|\sigma| \geq cn/\|A\|^2$  and  $s_{\min}(A_\sigma) \geq \gamma$ , where  $\|A\|$  is the spectral norm of  $A$ , and  $A_\sigma$  denotes the corresponding column submatrix. Vershynin [27] later extended this principle to rectangular matrices, with the number of selected columns governed by the stable rank  $\|A\|_{\text{HS}}^2/\|A\|^2$ . The literature on restricted invertibility is extensive and touches on various problems in local Banach space theory, convex geometry, and operator algebras; see, for example, [24, 29, 16, 17, 20], and, in particular, recent work [20] for an in-depth literature overview. The core principle of the theory is that well-invertible column subsets exist under very general spectral assumptions on the input matrix.

In this paper, we focus on well-invertible column subsets of *proportional* size, namely the setting  $|\sigma| = \Omega(k)$ . As a specific illustration, let  $M$  be a  $k \times n$  matrix ( $n \geq k$ ) with unit-length columns. The theory of restricted invertibility then asserts that, under an extra assumption on the matrix norm, there exists a set of columns of  $M$  of size  $\Omega(k)$  such that the smallest singular value of the corresponding submatrix is bounded below by a positive constant. Conversely, our goal in this work is to provide an *upper* bound on the number of well-invertible column subsets of  $M$ , under a sparsity assumption on the input matrix. In this sense, our result can be viewed as a dual to the Bourgain–Tzafriri theory: while restricted invertibility guarantees that good column selections exist, we show that, for sparse matrices, such selections are necessarily rare. Our main result is the following asymptotic statement about the relative proportion of well-invertible column subsets in sparse matrices. In what follows, by  $M_{\cdot j}$  we denote  $j$ -th column of a matrix  $M$ . Further, write  $\text{supp } M_{\cdot j}$  for the support of its  $j$ -th column.

**Theorem 1.3** (Main result: well-invertible column subsets of sparse matrices are rare). *Let  $M^{(k)}$ ,  $k \geq 1$ , be a deterministic sequence of  $k \times n_k$  matrices with  $n_k/k \rightarrow \infty$ . We assume the following:*

- (1) (*Column support*) *The average column support size is of constant order:*

$$\frac{1}{n_k} \sum_{j=1}^{n_k} |\text{supp } M_{\cdot j}^{(k)}| = O(1).$$

- (2) (*Average magnitude*) *The average magnitude of **nonzero** entries is of constant order:*

$$\frac{1}{|\text{supp } M^{(k)}|} \sum_{i=1}^k \sum_{j=1}^{n_k} |M_{ij}^{(k)}| = O(1).$$

- (3) (*Column overlaps*) *The total number of pairs of columns of  $M^{(k)}$  whose supports overlap on two or more indices, is of order*

$$o\left(\frac{n_k^2}{k}\right).$$

*Then for every  $\varepsilon > 0$ , if  $\Omega_k \subset [n_k]$  are chosen uniformly at random with  $|\Omega_k| = \lfloor \varepsilon k \rfloor$ , we have*

$$s_{\min}(M_{\Omega_k}^{(k)}) = o(1)$$

*with probability  $1 - o(1)$ . Equivalently, a uniformly chosen proportional-size column subset is well-invertible with asymptotically vanishing probability.*

The “columns overlap” assumption in the above theorem should not be thought of as a pseudorandomness condition on the input matrix. Indeed, it is not difficult to verify that for classical random matrix models of constant sparsity (including combinatorial random matrices and SparseStack<sup>†</sup>), the typical number of

pairs of columns intersecting on two or more rows, is of order  $O(n_k^2/k^2)$ , i.e almost  $k$  times smaller than the threshold we impose in the theorem.

We emphasize that our result is *universal* rather than existential. Of course, one can easily construct sparse matrices satisfying the assumptions of the restricted invertibility theory, and yet containing very few well invertible column subsets. As an example, one can take a  $k \times k$  identity matrix extended by the  $k \times (n_k - k)$  zero matrix, or concatenate  $n_k/k$  identity matrices of size  $k \times k$  to form a  $k \times n_k$  matrix. In the first example, a random column subset of size  $\Theta(k)$  is singular with high probability due to presence of zero columns. In the second example, the obstruction to invertibility is existence of two identical columns within the selected subset.

There are, however, sparse matrices for which counting well invertible column subsets is far less obvious. Such are, for example, random matrices with independent columns and  $d$  uniformly chosen ones in each column, or random  $\text{SparseStack}^\top$  matrices. As we will show, these matrices satisfy the assumptions of Theorem 1.3 with high probability.

The proof of our theorem shows that, apart from the selected submatrix having a zero column or having two identical columns, there is the third less obvious reason for the scarcity of well-invertible submatrices in the sparse setting. It is the abundance of  $k \times m$  rectangular submatrices having very small least singular values, and comprising certain combinatorial *tree* structure. Here,  $m = o(k)$  and  $m \rightarrow \infty$  as  $n_k/k \rightarrow \infty$ . These matrices are defined in Section 1.3 below. If a selected submatrix  $M_{\Omega_k}^{(k)}$  of  $M^{(k)}$  contains one of these special submatrices, then the least singular value of  $M_{\Omega_k}^{(k)}$  will be small as well, and their profusion guarantees that the number of submatrices not containing any special  $k \times m$  one is small.

Theorem 1.3 directly addresses Problem 1.2 for sparse matrices satisfying the local column-overlap and growth assumptions. As a specific illustration of the theorem, consider the SparseStack model of sketch matrices.

**Definition 1.4** (SparseStack). *Fix integers  $n, \zeta, b \in \mathbb{N}$ , and set*

$$k := \zeta b.$$

*Partition the column set  $[k]$  into  $\zeta$  consecutive blocks of size  $b$ . Independently choose a uniformly distributed random position in each block and populate these positions with independent  $\pm 1$  entries.*

*More formally, the SparseStack matrix is a random  $n \times k$  matrix  $S = (s_{ij})$  constructed as follows: for every row index  $i \in [n]$  and every block index  $t \in [\zeta]$ , choose  $p_{i,t} \in [b]$ ,  $q_{i,t} \in \{-1, 1\}$ , independently and uniformly at random, and place the value  $q_{i,t}/\sqrt{\zeta}$  in the matrix coordinate of  $S$*

$$(i, (t-1)b + p_{i,t}).$$

*After the construction is complete, set all unassigned entries of  $S$  to zero.*

The authors of [5] formulated the following conjecture:

**Conjecture 1.5** (SparseStack: constant row sparsity; see [5, Conjecture 1.9]). *A SparseStack test matrix is an  $(r, 1/2)$ -OSI for some embedding dimension  $k = O(r)$  and some sparsity level  $\zeta = O(1)$ .*

As an immediate corollary of the main theorem, we obtain

**Corollary 1.6** (No OSI for constant-sparsity SparseStack model). *For every integer  $\zeta \geq 2$  and every choice of  $\varepsilon, \alpha > 0$  there are constants  $B_0, C > 0$  depending only on  $\zeta, \varepsilon, \alpha$  with the following property. Let  $b \geq B_0$ , let  $k = \zeta b$  and  $n \geq Ck$ , and assume that the  $n \times k$  matrix  $S$  is SparseStack with row sparsity  $\zeta$  and with block size  $b$ . Let  $\Omega$  be any non-random subset of  $[n]$  of size at least  $\varepsilon k$ . Then with probability at least 0.99,*

$$s_{\min}((S^\top)_\Omega) < \alpha,$$

*where  $s_{\min}((S^\top)_\Omega)$  is the smallest singular value of the  $k \times |\Omega|$  submatrix  $(S^\top)_\Omega$ . In particular,  $S$  is not  $(\varepsilon k, \alpha)$ -OSI.*

**Remark 1.7.** *The sign distribution of the nonzero entries in the SparseStack model can be replaced by an arbitrary distribution with finite first moment; our theorem then yields the same conclusion for  $(S^\top)_\Omega$ . We leave the details to the interested reader, and refer to Sections 5 and 6 for a discussion of quantitative aspects of our results.*

**Remark 1.8.** *Our main result extends well beyond the SparseStack model. In particular, our theorem covers “combinatorial” random sketch matrices with i.i.d. rows, where the row supports are chosen uniformly at random among all  $d$ -element subsets of  $[k]$  and the nonzero entries are  $\pm 1$ , as well as random matrices with prescribed number of non-zero entries in every row and column.*

### 1.3. Technical overview.

We now explain the graph mechanism behind the main result. To illustrate the idea, in this discussion we consider matrices whose entries only take values in  $\{0, 1, -1\}$  and every column has exactly  $d$  nonzero entries for some fixed  $d \geq 2$ .

1.3.1. *The obstruction pattern.* Given a rectangular matrix  $M$ , we view it as the adjacency matrix of a bipartite graph: rows are left vertices, columns are right vertices, and each nonzero entry corresponds to an edge. Sampling column subsets of  $M$  is equivalent to constructing a “right-induced subgraph”, obtained as one-neighborhood of the selected right vertices. Notation-wise, if  $G = (I, J, E)$  is a bipartite graph on  $I \sqcup J$  and  $\Omega \subset J$ , we write

$$G[\Omega] := (I, \Omega, E \cap (I \times \Omega)).$$

The *local obstruction* to strong quantitative invertibility of the random column subset, which we exploit in the proof, is an  *$m$ -matching rooted tree*. Let  $G[\Omega]$  be the support graph of the sampled matrix  $M_\Omega$ . Suppose further that  $G[\Omega]$  contains an induced depth-three rooted bipartite tree  $T$  with right root  $j_\star$ . Thus  $j_\star$  has left neighbors  $i_1, \dots, i_d$ ; for each  $i_\ell$ , there are  $m$  secondary (auxiliary) vertices  $\mathcal{K}_{i_\ell} = \{c_{\ell,1}, \dots, c_{\ell,m}\}$  adjacent to  $i_\ell$ ; and each auxiliary right vertex has its remaining  $d-1$  left neighbors in a third layer, with all these outer left neighbors distinct. Let us emphasize that the tree contains all graph neighbors of each of its right vertices.

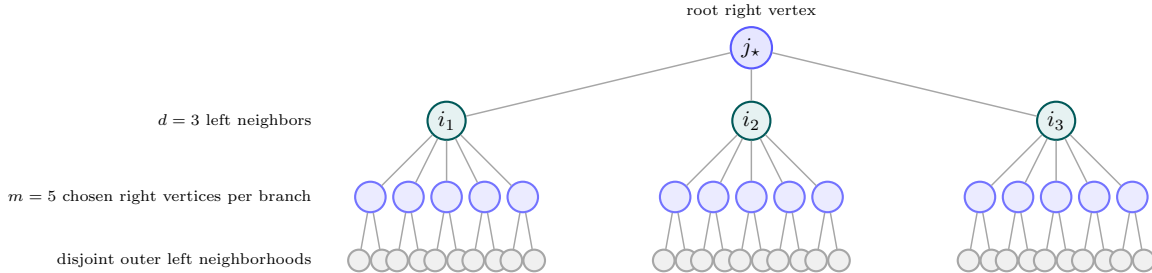


FIGURE 1. A  $m$ -matching rooted tree, shown for right degree  $d = 3$  and width  $m = 5$ . The four tree layers have 1, 3, 15, and 30 vertices, respectively.

This pattern gives an explicit test vector  $v$  supported on the right vertices of  $T$ . Put  $v_{j_\star} = 1$ , and choose the coefficients  $v_{c_{\ell,a}}$  with  $|v_{c_{\ell,a}}| = 1/m$  so that the contribution of the root column cancels on the rows  $i_1, \dots, i_d$ . The only remaining nonzero coordinates of  $M_\Omega v$  lie on the outer left layer. They are disjointly supported, there are  $dm(d-1)$  of them, and each has magnitude  $1/m$ . Hence

$$\|M_\Omega v\|_2 \leq \frac{\sqrt{dm(d-1)}}{m} \leq \frac{d}{\sqrt{m}}.$$

Since  $\|v\|_2 \geq 1$ , this forces  $s_{\min}(M_\Omega) \leq d/\sqrt{m}$ . Thus, whenever we are able to find the described pattern in  $M_\Omega$  with parameter  $m$  large, we immediately certify a strong upper bound on the smallest singular value of  $M_\Omega$ .

|             | $j_\star$ | $c_{1,\cdot}$ |   |   |   |   | $c_{2,\cdot}$ |   |   |   |   | $c_{3,\cdot}$ |   |   |   |   |
|-------------|-----------|---------------|---|---|---|---|---------------|---|---|---|---|---------------|---|---|---|---|
| $i_1$       | 1         | 1             | 1 | 1 | 1 | 1 |               |   |   |   |   |               |   |   |   |   |
| $i_2$       | 1         |               |   |   |   |   | 1             | 1 | 1 | 1 | 1 |               |   |   |   |   |
| $i_3$       | 1         |               |   |   |   |   |               |   |   |   |   | 1             | 1 | 1 | 1 | 1 |
| $M_T =$     |           |               |   |   |   |   |               |   |   |   |   |               |   |   |   |   |
| $o_{1,1,1}$ |           | 1             |   |   |   |   |               |   |   |   |   |               |   |   |   |   |
| $o_{1,1,2}$ |           | 1             |   |   |   |   |               |   |   |   |   |               |   |   |   |   |
| $o_{1,2,1}$ |           |               |   | 1 |   |   |               |   |   |   |   |               |   |   |   |   |
| $o_{1,2,2}$ |           |               |   | 1 |   |   |               |   |   |   |   |               |   |   |   |   |
|             |           |               |   |   |   |   |               |   |   |   |   |               |   |   |   |   |
| $o_{3,5,1}$ |           |               |   |   |   |   |               |   |   |   |   |               |   |   |   | 1 |
| $o_{3,5,2}$ |           |               |   |   |   |   |               |   |   |   |   |               |   |   |   | 1 |

FIGURE 2. The support-pattern matrix associated with the tree in Figure 1. Blank entries are zero; the displayed 1's record support entries. The first three rows are the root-neighbor rows where the test vector cancels, while the lower rows are disjoint outer-neighborhood rows.

For a formal definition of the  $m$ -matching tree, see Definition 2.2 below. A generalized version of the deterministic singular-value statement sketched above is Proposition 2.5. The simplified description above only operates with the matrix support. For a general *weighted* matrix  $M$ , the construction of the target tree pattern is more complicated: we require each edge connecting second and third layers of the tree, to have a magnitude at least as large as that of the corresponding “parent” edge connecting the second layer to the root; this is the weight-compatible condition in Definition 2.4.

**1.3.2. Reduction to  $C_4$ -free support graphs.** To extract an  $m$ -matching tree from a sampled subgraph, we repeatedly use a two-stage sampling strategy. Rather than looking only at the final sample, we first expose a larger intermediate right vertex set. More precisely, we may realize the final uniform sample in two stages: first choose an intermediate set  $\Omega'_k \subset [n_k]$  uniformly, with  $|\Omega'_k|$  larger than the final sample size, and then choose the final set  $\Omega_k \subset \Omega'_k$  uniformly with  $|\Omega_k| = r_k = \Theta(k)$ . This intermediate set has enough room to reveal the desired structure, and after trimming a small number of “bad” right vertices it may have better deterministic properties than the original graph. In fact, we use multiple rounds of intermediate sampling in the proof.

One of the key properties we obtained from an intermediate support graph is  $C_4$ -freeness. Here we illustrate that in detail. For distinct columns  $j, j'$ , the overlap condition appearing in Theorem 1.3 has the following graph-theoretic interpretation, where  $G^{(k)}$  is the support graph of  $M^{(k)}$ :

$$|\text{supp } M_{\cdot j}^{(k)} \cap \text{supp } M_{\cdot j'}^{(k)}| \geq 2 \quad \Leftrightarrow \quad \begin{array}{l} \text{there is a 4-cycle } C_4 \text{ in the support graph } G^{(k)} \\ \text{containing } j \text{ and } j'. \end{array}$$

Specifically, if  $i, i'$  are two distinct common left neighbors of  $j$  and  $j'$ , then  $i - j - i' - j' - i$  is a four-cycle in the support graph. Thus the third condition in the main theorem controls the number of pairs of right vertices that lie together on a four-cycle.

Our construction relies on a somewhat technical *local sparse four-cycle incidence* assumption on the graph, see Definition 4.2. A cleaner, albeit slightly more restrictive, way to ensure that it is satisfied is to assume that the support graph has  $o(n^2/k)$  four-cycles in total:

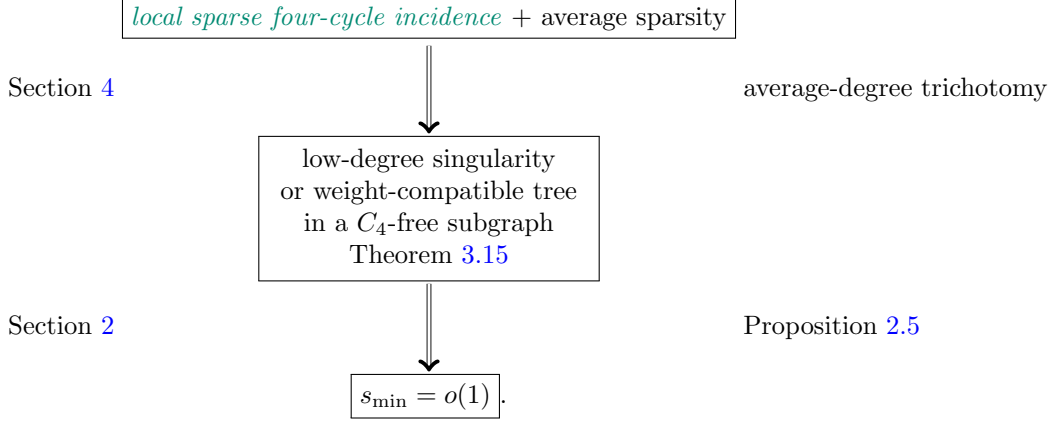
$$(1) \quad \text{the number of 4-cycles in } G^{(k)} = o\left(\frac{n_k^2}{k}\right).$$

Throughout the introduction, the reader can substitute condition (1) for *local sparse four-cycle incidence*. This easier assumption holds for a number of random matrix families, and is used in the (non-technical)

statement of the main result above. For example, in fixed-sparsity models such as SparseStack, the expected number of four-cycles is  $O_d(n_k^2/k^2)$ , so this global condition holds with high probability by Markov's inequality.

Under *local sparse four-cycle incidence*, together with  $n_k/k \rightarrow \infty$  and average right degree  $O(1)$ , the intermediate sampled set can be trimmed to a large  $C_4$ -free bounded-degree subgraph with high probability. From that point, the task is to extract an *m-matching rooted tree* from the  $C_4$ -free subgraph.

The proof runs through the chain



The key technical result of our work is the sampling theorem for  $C_4$ -free graphs in Section 3; here we state an informal version:

**Theorem** (Informal statement of Theorem 3.15). *Fix  $d \geq 2$  and  $\rho > 0$ . Consider three positive integers  $r, k, n$  satisfying  $\rho k \leq r \leq k \leq n$ . Let  $G = (I, J, M)$  be a  $C_4$ -free weighted bipartite graph with  $|I| = k$ ,  $|J| = n$ , and right support degrees between 2 and  $d$ . Set*

$$\Lambda := \min\left(\frac{r}{2}, \frac{n}{k}\right) \quad \text{and} \quad m = \Theta\left(\frac{1}{d^3} \frac{\log \Lambda}{\log(\frac{d}{\rho} \log \Lambda)}\right).$$

*Then for a uniform right sample  $\Omega$  of size  $r$ , we have*

$$\mathbb{P}(G[\Omega] \text{ contains a weight-compatible } m\text{-matching rooted tree}) \geq 1 - O\left(\frac{d}{\log(\Lambda)}\right).$$

The notion of weight-compatible above was a technical requirement to ensure that the test vector construction works in the weighted case. Further, we have the following reduction from *local sparse four-cycle incidence* to the  $C_4$ -free support subgraph.

**Proposition** (Average-degree trichotomy, asymptotic form of Proposition 4.5). *Let  $G_k$  be a sequence of weighted bipartite graphs with  $k$  left vertices,  $n_k$  right vertices, average right support-degree  $O(1)$ , average nonzero weight magnitude at most 1,  $n_k/k \rightarrow \infty$ , and *local sparse four-cycle incidence*. Let  $r_k$  satisfy  $k/r_k = O(1)$ . Then, for a uniform right sample  $\Omega_k$  of size  $r_k$ , with probability  $1 - o(1)$  one of the following holds:*

- (1)  $G_k[\Omega_k]$  has an isolated sampled right vertex;
- (2)  $G_k[\Omega_k]$  has two degree-one sampled right vertices with the same left neighbor;
- (3)  $G_k[\Omega_k]$  contains a *weight-compatible  $m_k$ -matching rooted tree* inside a  $C_4$ -free subgraph of bounded maximum right degree, with nonzero weights bounded by a constant depending only on the average-degree bound, where

$$m_k \rightarrow \infty.$$

The first two cases are the price of working with average right degree, rather than a model in which every sampled column has degree  $d$ . Average control can leave many degree-zero or degree-one columns. If the sample contains an isolated right vertex, or two degree-one right vertices supported on the same left vertex, the corresponding sampled matrix is already singular. If neither low-degree obstruction occurs, the

trichotomy argument truncates to a bounded-degree subgraph and produces the tree case. The bounded-degree cutoff gives at most  $O(m_k)$  third-layer vertices, so  $m_k \rightarrow \infty$  forces the desired small singular value.

### 1.3.3. Extracting the $m$ -matching rooted tree from a $C_4$ -free subgraph.

**Witness blocks.** The  $C_4$ -free property helps us find such trees through a simpler combinatorial object. The rough idea is as follows. We still have the root vertex  $j_\star$  and its left neighbors  $I_\star = N_G(j_\star)$ . To build an  *$m$ -matching rooted tree* directly, one would need to find disjoint right-neighbor subsets  $\mathcal{K}_i \subseteq N_G(i) \setminus \{j_\star\}$  for each  $i \in I_\star$  so that all the outer left neighborhoods  $N_G(\mathcal{K}_i) \setminus \{i\}$  are disjoint. Instead, we first search for larger auxiliary right-neighbor sets  $\mathcal{R}_i$  of each  $i \in I_\star$ , each of size at least  $m(1 + (d-1)^2)$ , which we call *reservoirs*. The  $C_4$ -freeness forces the *reservoirs* to be disjoint regardless of how they are chosen, and  $C_4$ -freeness together with the right-degree bound ensures that for every  $j \in \cup_{i \in I_\star} \mathcal{R}_i$ , the number of other  $j' \in \cup_{i \in I_\star} \mathcal{R}_i$ , sharing a neighbor with  $j$  outside of  $I_\star$ , is bounded. Thus we can greedily prune the *reservoirs* to find  $\mathcal{K}_i \subseteq \mathcal{R}_i$  for each  $i \in I_\star$  so that all the outer left neighborhoods  $N_G(\mathcal{K}_i) \setminus \{i\}$  are disjoint, forming the desired  *$m$ -matching rooted tree*. See Lemma 3.5 for the formal statement.

The combination of the root and its *reservoirs* is the intermediate object we want the random sample to capture. More precisely, if each  $\mathcal{R}_i$ ,  $i \in I_\star$ , has size at least  $m(1 + (d-1)^2)$ , then the right-vertex set

$$\mathcal{W}(j_\star) := \{j_\star\} \cup \bigcup_{i \in I_\star} \mathcal{R}_i$$

is a *witness block*. Capturing this whole block in a uniform sample is easier to analyze than capturing the already-pruned tree.

**Overall strategy in the  $C_4$ -free case.** Now suppose that  $G = (I, J, E)$  is  $C_4$ -free, with  $|I| = k$  and  $|J| = n$ , where  $n$  is much larger than  $k$ . We choose a uniform random subset  $\Omega \subseteq J$  of size  $r$ , with  $r$  of order  $k$ , and want to show that with high probability the induced subgraph  $G[\Omega]$  contains a *witness block*.

To justify the existence of many pairwise disjoint *witness blocks*, we rely on the property that typical pair of right vertices don't share common left neighbors to guarantee the existence of many root candidates with good properties. Then a multi-matching argument chooses pairwise disjoint heavy *reservoirs* around their neighboring left vertices, producing many disjoint *witness blocks*.

Suppose  $\lambda = n/r$  and we have  $K$  disjoint *witness blocks*, each of size at most  $B$ . A fixed block is then contained in a uniform  $r$ -subset with probability roughly  $\lambda^{-B}$ , and formally at least  $(8\lambda)^{-B}$  in Lemma 3.10. Because the blocks are disjoint, sequential exposure gives the heuristic failure bound

$$(1 - c\lambda^{-B})^K,$$

With compatible choices of  $\lambda$ ,  $B$ , and  $K$ , this probability is very small. This is the main idea behind the proof of Theorem 3.15.

**1.4. Organization.** In Section 2 we prove that matching rooted trees force small least singular values. In Section 3 we prove the  $C_4$ -free sampling theorem. In Section 4 we prove the average-degree trichotomy, reducing *local sparse four-cycle incidence* to the  $C_4$ -free subgraph case and the low-degree singular cases. In Section 5 we combine the graph cases with the test vectors to prove the matrix theorem. Section 6 applies the result to SparseStack and OSI. We close with open problems.

**Funding acknowledgement.** K.T. was partially supported by NSF grant DMS 2452120.

**Acknowledgement of AI Assistance.** The authors used ChatGPT for language editing, stylistic suggestions, and assistance in developing and checking some proof arguments during the preparation of this manuscript. All mathematical statements, proofs, and final wording were independently reviewed and verified by the authors, who take full responsibility for the content of the paper.

## 2. FROM MATRICES TO GRAPHS

This section records the analytic obstruction used throughout the paper. The object is purely local in the support of a weighted bipartite graph: once a sampled support contains a weight-compatible matching rooted tree, the corresponding weighted matrix has an explicit test vector with small image.

**2.1. Graph Conventions and Matching Rooted Trees.** A matrix can also be viewed as a weighted bipartite graph. We will use the following conventions throughout the paper.

**Definition 2.1** (Weighted bipartite graph and support graph). *Let  $G = (I, J, M)$  be a weighted bipartite graph, where*

$$M \in \mathbb{R}^{I \times J}.$$

*We call  $I$  the left vertex set and  $J$  the right vertex set. We declare that*

$$(2) \quad (i, j) \in I \times J \text{ is an edge} \quad \Leftrightarrow \quad M_{ij} \neq 0$$

*and, when  $(i, j)$  is an edge, its weight is  $M_{ij}$ . With this convention, we define the support of  $G$  to be the set of edges*

$$E_M := \{(i, j) \in I \times J : M_{ij} \neq 0\}.$$

*By the support graph of  $G$  we mean the unweighted bipartite graph  $(I, J, E_M)$ .*

Indeed, many technical statements in the paper only depend on the support, and not on the actual weights. We will often blur the distinction between  $G$  and its support graph  $(I, J, E_M)$ . In other words, the majority of the arguments are purely combinatorial, and the weights only come into play when we construct the test vector for the matching rooted tree.

For a vertex  $v$ , write  $N_G(v)$  for its neighbors in this support graph. Degrees always mean support degrees. For  $\Omega \subseteq J$ , write

$$G[\Omega] := (I, \Omega, M_{\cdot, \Omega})$$

for the right-induced weighted subgraph on  $\Omega$ .

The following dictionary will be used repeatedly.

| Matrix language                   | Graph language                                               |
|-----------------------------------|--------------------------------------------------------------|
| row of $M$                        | left vertex                                                  |
| column of $M$                     | right vertex                                                 |
| sampled columns $M_\Omega$        | right-induced weighted subgraph $G[\Omega]$                  |
| column sparsity                   | right degree                                                 |
| zero sampled column               | isolated sampled right vertex                                |
| two degree-one columns on one row | two degree-one right vertices sharing the same left neighbor |

**Definition 2.2** ( $m$ -Matching Rooted Tree). *Let  $G = (I, J, M)$  be a weighted bipartite graph. Fix a right vertex  $j_\star \in J$ , and set*

$$I_\star := N_G(j_\star).$$

*We say that  $G$  contains an  $m$ -matching rooted tree at  $j_\star$  if, for each  $i \in I_\star$ , there is a set*

$$\mathcal{K}_i \subseteq N_G(i) \setminus \{j_\star\}, \quad |\mathcal{K}_i| = m,$$

*such that*

$$N_G(j) \cap I_\star = \{i\} \quad \forall i \in I_\star, \forall j \in \mathcal{K}_i,$$

*the sets  $\mathcal{K}_i$ ,  $i \in I_\star$ , are pairwise disjoint, and the outer left neighborhoods*

$$N_G(j) \setminus I_\star, \quad j \in \bigcup_{i \in I_\star} \mathcal{K}_i,$$

*are pairwise disjoint. Indeed, these vertices form a three-layer rooted tree in the support graph, including all neighbors of the right vertices in the tree:*

- $j_\star$  is the root,
- $I_\star$  is the first left layer,
- the vertices in  $\bigcup_i \mathcal{K}_i$  are the second right layer, and
- the sets  $N_G(j) \setminus I_\star$  for  $j \in \bigcup_{i \in I_\star} \mathcal{K}_i$  form the third left layer.

**Remark 2.3** (Automatic disjointness in the  $C_4$ -free case). *If  $G$  is  $C_4$ -free, then the condition  $N_G(j) \cap I_\star = \{i\}$  and the pairwise disjointness of the sets  $\mathcal{K}_i$ ,  $i \in I_\star$ , are automatic. Indeed, any violation would give a second-layer right vertex  $j$  and the root  $j_\star$  two common left neighbors.*

**Definition 2.4** (Weight-compatible matching rooted tree). *Let  $G = (I, J, M)$  contain an  $m$ -matching rooted tree at  $j_*$ , witnessed by sets  $\mathcal{K}_i$ ,  $i \in I_* = N_G(j_*)$ . We say that this rooted tree is weight-compatible if*

$$|M_{i,j}| \geq |M_{i,j_*}| \quad \forall i \in I_*, \forall j \in \mathcal{K}_i.$$

*We say that  $G$  contains a weight-compatible  $m$ -matching rooted tree if some choice of root and witness sets has this property.*

## 2.2. Weighted Graphs.

**Proposition 2.5** (Weight-Compatible Matching Rooted Tree Implies Small Singular Value). *Let  $G = (I, J, M)$  be a real weighted bipartite graph. Suppose  $G$  contains a weight-compatible  $m$ -matching rooted tree, witnessed by a root  $j_*$  and sets  $\mathcal{K}_i$ ,  $i \in I_* = N_G(j_*)$ . Define the third left layer*

$$\mathcal{L}_{\text{out}} := \bigcup_{i \in I_*} \bigcup_{j \in \mathcal{K}_i} (N_G(j) \setminus I_*),$$

*and define its maximum weight by*

$$A_{\text{out}} := \max \{|M_{\ell,j}| : i \in I_*, j \in \mathcal{K}_i, \ell \in N_G(j) \setminus I_*\},$$

*with the convention  $A_{\text{out}} = 0$  if the set is empty. Then*

$$s_{\min}(M) \leq \frac{A_{\text{out}}}{m} \sqrt{|\mathcal{L}_{\text{out}}|}.$$

*Proof.* For  $j \in \bigcup_{i \in I_*} \mathcal{K}_i$ , let  $i(j)$  be the unique vertex of  $I_*$  with  $j \in \mathcal{K}_{i(j)}$ . By the definition of an  $m$ -matching rooted tree,  $j$  is adjacent to no other vertex of  $I_*$ .

Define  $v \in \mathbb{R}^J$  by

$$v_{j_*} = 1, \quad v_j = -\frac{M_{i(j),j_*}}{mM_{i(j),j}} \quad \text{for } j \in \bigcup_{i \in I_*} \mathcal{K}_i,$$

and put  $v_j = 0$  on all remaining right vertices. The denominators are nonzero because  $j \in N_G(i(j))$ . Also  $\|v\|_2 \geq 1$ , and weight compatibility gives  $|v_j| \leq 1/m$  for every selected second-layer right vertex.

For every  $i \in I_*$ , the second-layer contributions cancel the root contribution:

$$(Mv)_i = M_{i,j_*} + \sum_{j \in \mathcal{K}_i} M_{i,j} \left( -\frac{M_{i,j_*}}{mM_{i,j}} \right) = 0.$$

For  $\ell \notin I_*$ , the root contributes nothing. The outer left neighborhoods of the selected right vertices are pairwise disjoint, so at most one selected right vertex contributes to row  $\ell$ . Hence every nonzero coordinate of  $Mv$  outside  $I_*$  has magnitude at most  $A_{\text{out}}/m$ .

Therefore

$$\|Mv\|_2^2 \leq |\mathcal{L}_{\text{out}}| \frac{A_{\text{out}}^2}{m^2}.$$

Since  $\|v\|_2 \geq 1$ , this proves the claimed bound. □

**Remark 2.6** (Signed right-regular case). *If the support is right  $d$ -regular and all nonzero entries of  $M$  are signs, then  $A_{\text{out}} = 1$  unless the third left layer is empty, in which case the bound is trivial. Moreover, the root has  $d$  first-layer neighbors, so there are  $md$  selected second-layer right vertices. Each has at most  $d - 1$  outer left neighbors, and therefore*

$$|\mathcal{L}_{\text{out}}| \leq md(d - 1).$$

*Thus Proposition 2.5 gives*

$$s_{\min}(M) \leq \sqrt{\frac{d(d-1)}{m}} \leq \frac{d}{\sqrt{m}}.$$

### 3. THE $C_4$ -FREE SAMPLING THEOREM

**3.1. Setup and Proof Roadmap.** In this section we use the graph conventions from Section 2. The goal is to show that a uniform sample of  $r$  right vertices typically contains a *weight-compatible  $m$ -matching rooted tree*, where  $m$  grows with  $r$  and  $n/k$ . (See the informal statement of Theorem 3.15 in the introduction, and the precise statement later in this section below.)

We say that a bipartite graph  $G$  is  $C_4$ -free if it has no 4-cycles, or equivalently if no two distinct right vertices have two common left neighbors:

$$|N_G(j) \cap N_G(j')| \leq 1 \quad \forall j \neq j' \in J.$$

Thus two right vertices can share at most one left neighbor. When  $|J| \geq 2$ , we measure how often such a common neighbor occurs by the *overlap parameter*

$$\eta(G) := \mathbb{P}(N_G(j_1) \cap N_G(j_2) \neq \emptyset),$$

where  $(j_1, j_2)$  is a uniformly chosen ordered pair of distinct right vertices. For bounded right degrees and  $n/k \rightarrow \infty$ ,  $C_4$ -freeness forces  $\eta(G) = o(1)$ ; see Lemma 3.4.

**Definition 3.1** (Roots, reservoirs, and witness blocks). *In this section a root, or candidate root, is a right vertex  $j_\star \in J$ . Its first left layer is*

$$I_\star := N_G(j_\star).$$

*For  $i \in I_\star$ , an  $L$ -reservoir at  $i$  is a set*

$$\mathcal{R}_i \subseteq N_G(i) \setminus \{j_\star\}$$

*of auxiliary right vertices with  $|\mathcal{R}_i| \geq L$ . After choosing, for each  $i \in I_\star$ , an  $L$ -reservoir  $\mathcal{R}_i$ , the associated  $L$ -witness block at  $j_\star$  is the subset of right vertices*

$$\mathcal{W}(j_\star) := \{j_\star\} \cup \bigcup_{i \in I_\star} \mathcal{R}_i.$$

*The reservoirs  $\mathcal{R}_i$ ,  $i \in I_\star$ , witness this block.*

**Remark 3.2.** *Due to the  $C_4$ -free condition, the reservoirs  $\mathcal{R}_i$ ,  $i \in I_\star$ , are pairwise disjoint.*

For a given positive integer  $d \geq 2$ , let

$$\beta_d := 1 + (d - 1)^2.$$

If we have an  $L$ -witness block at  $j_\star$  with  $L \geq \beta_d m$ , then the pruning lemma, Lemma 3.5, relies on the  $C_4$ -free condition to greedily choose subfamilies  $\mathcal{K}_i \subseteq \mathcal{R}_i$  and produces the required  *$m$ -matching rooted tree* at  $j_\star$ . The weighted refinement below chooses *light* roots and heavy *reservoirs* so that this tree is automatically weight-compatible. Informally, a heavy *reservoir* at a left vertex consists of right neighbors with large incident weights, while a *light* root is not among those heavy neighbors at any adjacent left vertex; the precise definition is Definition 3.6.

A high level overview of the proof is as follows. We first show that a random  $\lambda r$ -subset of the right vertices contains many *light* candidate roots with large heavy *reservoirs*, where  $\lambda > 2$  is a parameter that grows slowly with  $r$  and  $n/k$ . Then we show that a uniform  $r$ -subset of the right vertices captures at least one complete *witness block* with high probability. Finally, the pruning lemma and the heavy-reservoir corollary produce a *weight-compatible  $m$ -matching rooted tree* from this captured block, establishing the main result.

The proof uses four scales. First, a small *overlap parameter* produces many candidate right roots. The number of candidate roots is denoted by  $K$ ; its allowable size is limited by the *overlap parameter*  $\eta(G)$  and by the final sample size  $r$ .

Second, for a target width  $m$ , we build heavy *reservoirs* around the left neighbors of each candidate root. The *reservoir* threshold is

$$L = m\beta_d = m(1 + (d - 1)^2).$$

Third, each root together with its *reservoirs* forms a *witness block*. The size of one block is at most

$$B = 1 + dL.$$

Thus  $B$  is of order  $d^3 m$ .

Fourth, the random sample must capture one complete *witness block*. If the ambient right side has size about  $\lambda r$ , then a fixed block of size  $B$  is captured with probability at least  $(8\lambda)^{-B}$ . Intuitively, with  $K$  disjoint attempts, the failure probability is therefore bounded by

$$(1 - (8\lambda)^{-B})^K \leq \exp\{-K(8\lambda)^{-B}\}.$$

The convenient finite condition

$$(8\lambda)^{2B} \leq K$$

turns this into the cleaner error bound  $\exp(-\sqrt{K})$ .

The exact parameter configurations in this section are chosen only to make these four steps compatible.

**Lemma 3.3** (Basic consequences of  $C_4$ -freeness). *Let  $G = (I, J, M)$  be  $C_4$ -free.*

(1) *If  $i \in I$  and  $j, j' \in N_G(i)$  are distinct, then*

$$(N_G(j) \setminus \{i\}) \cap (N_G(j') \setminus \{i\}) = \emptyset.$$

(2) *If  $j_\star \in J$ ,  $I_\star = N_G(j_\star)$ , and*

$$j \neq j_\star, \quad N_G(j) \cap I_\star \neq \emptyset,$$

*then  $j$  is adjacent to a unique vertex of  $I_\star$ .*

(3) *If  $J' \subseteq J$ , then  $G[J']$  is  $C_4$ -free.*

*Proof.* For the first assertion, if a left vertex  $x \neq i$  belonged to both  $N_G(j)$  and  $N_G(j')$ , then  $j$  and  $j'$  would share both  $i$  and  $x$ , contradicting  $C_4$ -freeness.

For the second assertion, if  $j$  were adjacent to two distinct vertices of  $I_\star$ , then  $j$  and  $j_\star$  would have two common left neighbors. This again contradicts  $C_4$ -freeness.

The third assertion is immediate because deleting right vertices cannot create a new pair of right vertices with two common left neighbors. □

**Lemma 3.4** ( $C_4$ -free graphs have sparse overlap). *Let  $G = (I, J, M)$  be  $C_4$ -free with  $|I| = k$ ,  $|J| = n \geq 2$ , and every right degree between 2 and  $d$ . Then*

$$\eta(G) \leq 2d \frac{k}{n}.$$

*Consequently, if  $d$  is fixed and  $G^{(k)}$  is a sequence of such graphs with  $n_k/k \rightarrow \infty$ , then  $\eta(G^{(k)}) = o(1)$ .*

*Proof.* For each left vertex  $i \in I$ , let  $D_i := |N_G(i)|$ . An ordered pair  $(j_1, j_2)$  with  $j_1 \neq j_2$  shares a left neighbor only if both vertices belong to  $N_G(i)$  for some  $i$ . Hence the number of such ordered pairs is at most

$$\sum_{i \in I} D_i(D_i - 1).$$

We claim  $D_i \leq k - 1$  for every  $i$ . Indeed, by the first basic consequence of  $C_4$ -freeness, the sets  $N_G(j) \setminus \{i\}$ ,  $j \in N_G(i)$ , are pairwise disjoint. Since every right vertex has degree at least 2, each of these sets is nonempty, so  $D_i \leq k - 1$ .

Since every right vertex has degree at most  $d$ ,  $\sum_i D_i \leq dn$ . Therefore

$$\sum_i D_i(D_i - 1) \leq \sum_i D_i^2 \leq (k - 1)dn.$$

Dividing by the total number  $n(n - 1)$  of distinct ordered pairs gives

$$\eta(G) \leq \frac{d(k - 1)}{n - 1} \leq 2d \frac{k}{n}.$$

□

**3.2. Three Finite Ingredients.** The proof of the sampling theorem uses three finite ingredients: a reservoir-pruning lemma, a root-selection lemma, and a block-capture lemma.

The first ingredient is the local  $C_4$ -free pruning step. Once every neighbor of the root has a large *reservoir*, it extracts the disjoint outer neighborhoods required in the matching rooted tree.

**Lemma 3.5** (Reservoirs produce an  $m$ -matching rooted tree). *Fix integers  $d \geq 1$  and  $m \geq 1$ , and set*

$$\beta_d := 1 + (d - 1)^2.$$

*Let  $G = (I, J, M)$  be  $C_4$ -free and suppose every right vertex has degree at most  $d$ . Fix  $j_\star \in J$ , set  $I_\star = N_G(j_\star)$ , and assume that for every  $i \in I_\star$  there is a *reservoir**

$$\mathcal{R}_i \subseteq N_G(i) \setminus \{j_\star\}$$

*with*

$$|\mathcal{R}_i| \geq m\beta_d.$$

*Then  $G$  contains an  $m$ -matching rooted tree at  $j_\star$  with  $\mathcal{K}_i \subseteq \mathcal{R}_i$  for every  $i \in I_\star$ .*

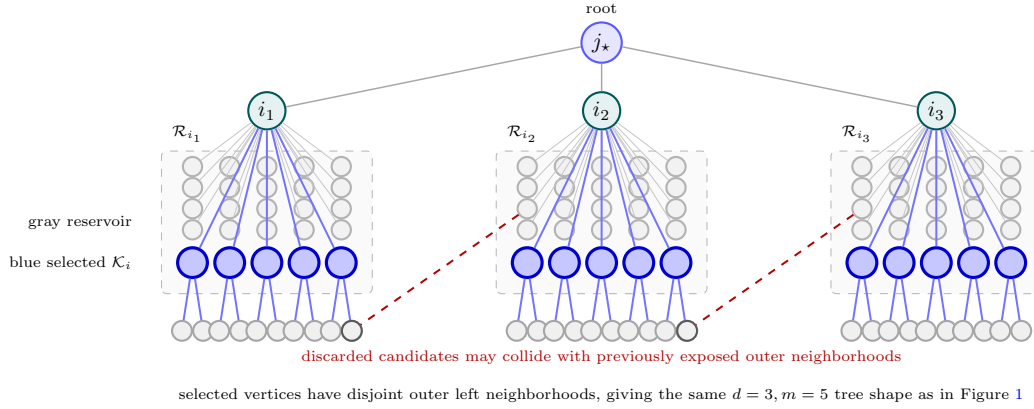


FIGURE 3. Reservoir pruning around a root for  $d = 3$  and  $m = 5$ . Each left neighbor of the root has a reservoir of size  $m(1 + (d - 1)^2) = 25$ . The pruning lemma selects five blue vertices from each reservoir and avoids candidates whose outer left neighborhoods collide with previously selected vertices.

*Proof.* Because  $G$  is  $C_4$ -free, every  $j \in \mathcal{R}_i$  satisfies

$$N_G(j) \cap I_\star = \{i\}.$$

Otherwise  $j$  and  $j_\star$  would have two common left neighbors. In particular, the *reservoirs*  $\mathcal{R}_i$  are pairwise disjoint as  $i$  varies.

For  $j \in \bigcup_{i \in I_\star} \mathcal{R}_i$ , call  $N_G(j) \setminus I_\star$  its outer left neighborhood. Enumerate  $I_\star = \{i_1, \dots, i_q\}$ , where  $q \leq d$ , and choose  $\mathcal{K}_{i_s} \subseteq \mathcal{R}_{i_s}$  greedily.

Suppose  $\mathcal{K}_{i_1}, \dots, \mathcal{K}_{i_{s-1}}$  have already been chosen. Fix a previously chosen vertex  $j'$ . Its outer left neighborhood has size at most  $d - 1$ . For each  $x \in N_G(j') \setminus I_\star$ , there is at most one vertex  $j \in \mathcal{R}_{i_s}$  with  $x \in N_G(j)$ ; two such vertices would share both  $i_s$  and  $x$ , creating a 4-cycle.

Thus each previously chosen vertex rules out at most  $d - 1$  candidates from  $\mathcal{R}_{i_s}$ . Since there are at most  $(s - 1)m \leq (d - 1)m$  previously chosen vertices, at most  $m(d - 1)^2$  candidates are ruled out. The *reservoir* size assumption

$$|\mathcal{R}_{i_s}| \geq m\beta_d = m + m(d - 1)^2$$

leaves at least  $m$  available candidates; take these as  $\mathcal{K}_{i_s}$ . The outer left neighborhoods of distinct vertices inside the same  $\mathcal{R}_{i_s}$  are automatically disjoint: otherwise two such right vertices would share both  $i_s$  and an outer left vertex, contradicting  $C_4$ -freeness.

After all *reservoirs* are processed, the chosen subfamilies have size  $m$  and have pairwise disjoint outer left neighborhoods. Hence they form an  *$m$ -matching rooted tree* at  $j_\star$ . See Figure 3 illustrating the reservoir pruning procedure.  $\square$

**Definition 3.6** (Heavy-neighbor sets and light roots). *For a  $G = (I, J, M)$  with maximum right degree  $d$ . For  $i \in I$ , write*

$$D_i := |N_G(i)|.$$

*Let  $N_G^\dagger(i) \subseteq N_G(i)$  be any set of cardinality  $\lceil \frac{1}{32d} D_i \rceil$ , such that*

$$|M_{ij}| \geq |M_{ij'}| \quad \forall j \in N_G^\dagger(i), \forall j' \in N_G(i) \setminus N_G^\dagger(i).$$

*A right vertex  $j_\star \in J$  is **light** if*

$$j_\star \notin N_G^\dagger(i) \quad \forall i \in N_G(j_\star).$$

**Remark 3.7** (Graph combinatorics versus matrix weights). *The choice of the sets  $N_G^\dagger(i)$  is the only place where the proof uses information beyond the combinatorics of the support graph. Indeed, after these sets are fixed, the rest of the argument is purely graph-theoretic. In particular, for  $\{0, \pm 1\}$  matrices this weighted choice is unnecessary: all nonzero entries have the same magnitude, so any support-theoretic choice of the corresponding auxiliary neighbor sets already gives the required weight-compatibility.*

**Corollary 3.8** (Heavy reservoirs produce a weight-compatible tree). *Fix integers  $d \geq 1$  and  $m \geq 1$ , and set*

$$\beta_d := 1 + (d - 1)^2.$$

*Let  $G = (I, J, M)$  be  $C_4$ -free and suppose every right vertex has degree at most  $d$ . Fix a **light** right vertex  $j_\star \in J$ . Set  $I_\star = N_G(j_\star)$ , and assume that for every  $i \in I_\star$  there is a *reservoir**

$$\mathcal{R}_i \subseteq N_G^\dagger(i) \setminus \{j_\star\}$$

*with*

$$|\mathcal{R}_i| \geq m\beta_d.$$

*Then  $G$  contains a **weight-compatible  $m$ -matching rooted tree** at  $j_\star$  with  $\mathcal{K}_i \subseteq \mathcal{R}_i$  for every  $i \in I_\star$ .*

*Proof.* By Lemma 3.5, the *reservoirs* contain subfamilies  $\mathcal{K}_i \subseteq \mathcal{R}_i$ ,  $i \in I_\star$ , which form an  *$m$ -matching rooted tree* at  $j_\star$ . For every  $i \in I_\star$  and every  $j \in \mathcal{K}_i$ , we have  $j \in N_G^\dagger(i)$ , while  $j_\star \notin N_G^\dagger(i)$  because  $j_\star$  is *light*. Therefore the defining property of  $N_G^\dagger(i)$  gives

$$|M_{ij}| \geq |M_{ij_\star}|.$$

This is exactly the weight-compatibility condition.  $\square$

The second ingredient combines sparse right-pair overlap with the light-root deletion. It produces many candidate right roots whose left neighborhoods are disjoint, whose neighboring left vertices have large degree, and which are *light* relative to the heavy-neighbor sets.

**Lemma 3.9** (Sparse overlap yields many light candidate roots). *Let  $G = (I, J, M)$  have  $|I| = k$  and  $|J| = n$  satisfying  $k/n \leq 1/32$ , and suppose every right degree is at most  $d$ . For  $i \in I$ , write  $D_i := |N_G(i)|$ , and let  $\eta := \eta(G)$  be the *overlap parameter* of  $G$ . Let  $K \geq 1$  satisfy*

$$K(8\eta + 2/n) \leq 1.$$

*Then there exist distinct right vertices  $j_1, \dots, j_K \in J$  such that:*

- (1) *the neighborhoods  $I_s := N_G(j_s)$  are pairwise disjoint;*
- (2) *for every  $s \in [K]$  and every  $i \in I_s$ ,  $D_i \geq \frac{n}{8k}$ .*
- (3) *every  $j_s$  is **light**.*

*Proof.* We will define three undesirable sets of right vertices.

First, let

$$I_{\text{bad}} := \{i \in I : D_i < n/(8k)\}.$$

Then

$$\sum_{i \in I_{\text{bad}}} D_i \leq k \cdot \frac{n}{8k} = \frac{n}{8}.$$

Hence at most  $n/8$  right vertices touch  $I_{\text{bad}}$ . Let  $\mathcal{V}$  be the set of right vertices  $j$  such that every  $i \in N_G(j)$  satisfies  $D_i \geq n/(8k)$ . Then  $|\mathcal{V}| \geq \frac{7}{8}n$ .

For a right vertex  $j$ , let

$$h(j) := |\{j' \in J \setminus \{j\} : N_G(j') \cap N_G(j) \neq \emptyset\}|.$$

Since  $\eta$  is the overlap probability for an ordered pair of distinct right vertices,

$$\frac{1}{n} \sum_{j \in J} h(j) = \eta(n-1) \leq \eta n.$$

Let

$$\mathcal{L} := \{j \in J : h(j) \leq 4\eta n\}.$$

If  $\eta = 0$ , then  $\mathcal{L} = J$ . If  $\eta > 0$ , Markov's inequality gives  $|J \setminus \mathcal{L}| \leq n/4$ .

Let

$$\mathcal{H} := \{j \in J : j \text{ is not } \textit{light}\}.$$

Then every vertex in  $\mathcal{H}$  belongs to  $N_G^\dagger(i)$  for at least one  $i \in I$ . Therefore

$$|\mathcal{H}| \leq \sum_{i \in I} |N_G^\dagger(i)| = \sum_{i \in I} \left\lceil \frac{D_i}{32d} \right\rceil \leq \frac{1}{32d} \sum_{i \in I} D_i + k \leq \left( \frac{1}{32} + \frac{k}{n} \right) n \leq \frac{n}{16}.$$

Set

$$\mathcal{U} := \mathcal{V} \cap \mathcal{L} \cap (J \setminus \mathcal{H}).$$

Our assumptions imply

$$|\mathcal{U}| \geq \frac{9}{16}n.$$

Choose roots greedily from  $\mathcal{U}$ . Suppose  $j_1, \dots, j_{s-1}$  have already been chosen. Each previous root  $j_t$  forbids itself and at most  $h(j_t) \leq 4\eta n$  right vertices whose neighborhoods meet  $N_G(j_t)$ . Thus the number of forbidden vertices is at most

$$(s-1)(4\eta n + 1) \leq K(4\eta n + 1).$$

The hypothesis gives

$$K(4\eta n + 1) \leq \frac{n}{2}.$$

Since  $|\mathcal{U}| \geq 9n/16$ , some vertex  $j_s \in \mathcal{U}$  remains available. By construction,  $j_s$  is distinct from the earlier roots and  $N_G(j_s)$  is disjoint from every earlier root neighborhood. Repeating this  $K$  times gives  $j_1, \dots, j_K$  with pairwise disjoint neighborhoods.

Finally, every chosen root lies in  $\mathcal{V}$ , so every left neighbor of every chosen root has degree at least  $n/(8k)$ . Every chosen root also lies outside  $\mathcal{H}$ , so it is *light*. □

The third ingredient is purely probabilistic: among many disjoint witness blocks, a uniform right-vertex sample captures one complete block with high probability.

**Lemma 3.10** (Capturing one of many disjoint witness blocks). *Fix integers  $B \geq 1$  and  $K \geq 1$ , and real  $\lambda > 2$ . Let  $r$  be a positive integer and  $n_0 = \lceil \lambda r \rceil$ . Assume*

$$(8\lambda)^{2B} \leq K, \quad K \log K \leq \frac{r}{2}.$$

*Let  $J$  be a set of size  $n_0$  and assume there are pairwise disjoint subsets*

$$\mathcal{W}_1, \dots, \mathcal{W}_K \subseteq J,$$

*each of size at most  $B$ . If  $\Omega \subseteq J$  is chosen uniformly from  $\binom{J}{r}$ , then*

$$\mathbb{P}(\forall s \in [K] : \mathcal{W}_s \not\subseteq \Omega) \leq \exp(-\sqrt{K}).$$

*The condition  $K \log K \leq r/2$  keeps the sequential exposure of the  $K$  blocks from depleting the sample. The condition  $(8\lambda)^{2B} \leq K$  ensures that the capture probability of one block, repeated  $K$  times, gives exponent at least  $\sqrt{K}$ .*

**Remark 3.11.** In this paper we repeatedly estimate the probability that a uniformly sampled set captures a fixed subset. If  $W \subseteq J$  has size  $b$  and  $\Omega$  is chosen uniformly from  $\binom{J}{r}$ , then

$$\mathbb{P}(W \subseteq \Omega) = \frac{\binom{|J|-b}{r-b}}{\binom{|J|}{r}} = \frac{(r)_b}{(|J|)_b},$$

where, for integers  $a, q \geq 0$ , we write

$$(3) \quad (a)_q := a(a-1) \cdots (a-q+1)$$

for the falling factorial, with the convention  $(a)_0 = 1$ .

*Proof.* For each  $s \in [K]$ , let  $\Gamma_s$  be the event  $\mathcal{W}_s \subseteq \Omega$ , and put  $b_s := |\mathcal{W}_s| \leq B$ .

We first record the scale bounds used below. From  $(8\lambda)^{2B} \leq K$ ,

$$2B \log(8\lambda) \leq \log K.$$

Since  $\lambda > 2$ , we have  $\log(8\lambda) > 1$ , and hence

$$B \leq \log K.$$

Also

$$K \geq (8\lambda)^{2B} \geq 16^2 > 2.$$

Hence

$$KB \leq K \log K \leq \frac{r}{2}.$$

Also, since  $K \log K \leq r/2$  and  $\log K \geq 1$ , we have  $K \leq r/2$ .

Reveal  $\Omega$  block by block in the order  $\mathcal{W}_1, \dots, \mathcal{W}_K$ . After the first  $s-1$  blocks have been examined, let  $a_{s-1}$  be the number of sampled elements already found in  $\mathcal{W}_1 \cup \dots \cup \mathcal{W}_{s-1}$ . Conditional on the past, the restriction of  $\Omega$  to the complement of these blocks is a uniform subset of size

$$q_s := r - a_{s-1}$$

inside a ground set of size

$$N_s := n_0 - |\mathcal{W}_1 \cup \dots \cup \mathcal{W}_{s-1}|.$$

The scale bounds give

$$q_s \geq r - KB \geq \frac{r}{2}$$

and

$$N_s \leq n_0 \leq 2\lambda r.$$

Moreover,

$$q_s - b_s \geq r - KB - B \geq \frac{r}{2} - \log K \geq \frac{r}{4},$$

because  $K \leq r/2$  implies  $\log K \leq r/4$  for  $K \geq 2$ .

Therefore, conditional on the past,

$$\mathbb{P}(\Gamma_s \mid \text{past}) = \frac{\binom{N_s - b_s}{q_s - b_s}}{\binom{N_s}{q_s}} = \frac{q_s! (N_s - b_s)!}{(q_s - b_s)! N_s!} = \prod_{\ell=0}^{b_s-1} \frac{q_s - \ell}{N_s - \ell} \geq \left( \frac{q_s - b_s}{N_s} \right)^{b_s} \geq \left( \frac{1}{8\lambda} \right)^{b_s} \geq \left( \frac{1}{8\lambda} \right)^B.$$

Set  $p_\star := (8\lambda)^{-B}$ . Then, by sequential conditioning,

$$\mathbb{P} \left( \bigcap_{s=1}^K \Gamma_s^c \right) \leq (1 - p_\star)^K \leq \exp(-p_\star K).$$

The hypothesis  $(8\lambda)^{2B} \leq K$  gives  $p_\star K = K/(8\lambda)^B \geq \sqrt{K}$ . Thus

$$\mathbb{P} \left( \bigcap_{s=1}^K \Gamma_s^c \right) \leq \exp(-\sqrt{K}).$$

□

**3.3. The Constant-Ratio  $C_4$ -free Case.** The next proposition combines the three ingredients when the ambient right side has size comparable to the final sample size.

**Proposition 3.12** (Constant-ratio  $C_4$ -free sampling theorem). *Fix an integer  $d \geq 2$ , a real  $\varepsilon \in (0, 1]$ , and a number  $\eta \geq 0$ . Set*

$$\beta_d := 1 + (d - 1)^2.$$

*Let  $k, r \geq 1$  satisfy*

$$\varepsilon k \leq r \leq k.$$

*Let*

$$K := \max \left\{ \ell \in \mathbb{N}_{\geq 1} : \ell \log \ell \leq \frac{r}{2} \quad \text{and} \quad \ell \leq \frac{1}{16\eta} \right\},$$

*where the second condition is omitted when  $\eta = 0$ . Assume this set is nonempty.*

*Let  $\lambda > 2$ , and define*

$$m := \left\lfloor \frac{\varepsilon \lambda}{512d^2\beta_d} \right\rfloor, \quad L := m\beta_d, \quad B := 1 + dL.$$

*Assume*

$$\varepsilon \lambda \geq 512d^2\beta_d, \quad (8\lambda)^{2B} \leq K.$$

*Let  $n := \lceil \lambda r \rceil$ , and let  $G = (I, J, M)$  be  $C_4$ -free with*

$$|I| = k, \quad |J| = n,$$

*and right degrees satisfying*

$$2 \leq |N_G(j)| \leq d \quad \forall j \in J.$$

*Assume*

$$\eta(G) \leq \eta.$$

*If  $\Omega$  is chosen uniformly from  $\binom{J}{r}$ , then*

$$\mathbb{P}(G[\Omega] \text{ contains a weight-compatible } m\text{-matching rooted tree}) \geq 1 - \exp(-\sqrt{K}).$$

**Remark 3.13** (Scale of the admissibility condition). *Since*

$$B = 1 + d\beta_d m \asymp d^3 m \quad \text{and} \quad m \asymp \frac{\varepsilon \lambda}{d^4},$$

*the exact condition*

$$(8\lambda)^{2B} \leq K$$

*needed in Lemma 3.10 should be read heuristically as*

$$\left( \frac{Cd^4 m}{\varepsilon} \right)^{Cd^3 m} \leq K$$

*for a universal constant  $C$ .*

*Proof. Step 1: choose candidate roots.* For  $i \in I$ , write  $D_i := |N_G(i)|$ . Lemma 3.9 requires

$$K(8\eta(G) + 2/n) \leq 1.$$

The assumed overlap bound gives  $8K\eta(G) \leq 8K\eta \leq 1/2$  when  $\eta > 0$ , while  $\eta(G) = 0$  when  $\eta = 0$ . Also  $(8\lambda)^{2B} \leq K$  implies  $\log K \geq 1$ , so  $K \log K \leq r/2$  gives  $K \leq r/2$ . Since  $n = \lceil \lambda r \rceil$  and  $\lambda > 2$ , we get  $2K/n \leq 1/2$ . Moreover,

$$\frac{k}{n} \leq \frac{1}{\varepsilon \lambda} \leq \frac{1}{32}.$$

Hence Lemma 3.9 applies.

We obtain  $K$  distinct right vertices  $j_1, \dots, j_K$  such that their neighborhoods

$$I_s := N_G(j_s)$$

are pairwise disjoint and every  $i \in I_s$  satisfies

$$D_i \geq \frac{n}{8k}.$$

Moreover, every  $j_s$  is *light*.

*Step 2: build disjoint heavy reservoirs.* Let

$$I' := \bigsqcup_{s=1}^K I_s, \quad J' := J \setminus \{j_1, \dots, j_K\}.$$

For  $i \in I'$ , let  $s_i$  be the unique index with  $i \in I_{s_i}$ . Since  $j_{s_i}$  is *light* and the root neighborhoods are pairwise disjoint,

$$N_G^\dagger(i) \subseteq J' \quad \forall i \in I'.$$

Let  $H$  be the bipartite graph on  $(I', J')$  whose edges are

$$i \sim_H j \iff j \in N_G^\dagger(i).$$

The right degrees in  $H$  are at most  $d$ . For  $i \in I'$ , the left degree in  $H$  is  $|N_G^\dagger(i)| = \lceil D_i/(32d) \rceil$ , and hence

$$\deg_H(i) \geq \frac{D_i}{32d} \geq \frac{n}{256dk}.$$

The degree estimate shows that any vertex in  $I'$  has a large neighborhood in the heavy-neighbor graph  $H$ , but these neighborhoods can overlap. We need to select, for each left vertex  $i \in I'$ , a subset  $\mathcal{R}_i$  of its neighbors in  $H$  of cardinality  $L$ , with all  $\mathcal{R}_i$  pairwise disjoint. This is a  $b$ -matching problem: the left vertices have demand  $L$ , while the right vertices have capacity 1. If  $L$  were 1, this would have been the standard perfect matching problem whose solvability is described by Hall's Matching theorem. The standard cloning reduction reduces the problem to a perfect matching by replacing each left vertex by  $L$  copies with the same neighborhood. Application of Hall's Matching theorem to the cloned graph shows that a required multi-matching exists if and only if for every subset  $T \subseteq I'$ ,

$$|N_H(T)| \geq L|T|$$

(see, for example, [11, 23]). We verify this Hall-type condition. For any  $T \subseteq I'$ , double-counting edges between  $T$  and  $N_H(T)$  gives

$$\frac{n}{256dk} |T| \leq d |N_H(T)|.$$

Therefore

$$|N_H(T)| \geq \frac{n}{256d^2k} |T| \geq \frac{\varepsilon\lambda}{256d^2} |T| \geq L|T|,$$

where the second inequality uses  $n \geq \lambda r \geq \varepsilon\lambda k$ , and the last follows from

$$L = m\beta_d \leq \frac{\varepsilon\lambda}{512d^2}.$$

Thus Hall's condition holds, and the disjoint heavy *reservoirs* exist.

For each  $s \in [K]$ , define the *witness block*

$$\mathcal{W}_s := \{j_s\} \cup \bigcup_{i \in I_s} \mathcal{R}_i.$$

The blocks are pairwise disjoint, and each has size at most

$$1 + dL = B.$$

*Step 3: capture a witness block.* The  $K$  disjoint *witness blocks* of size at most  $B$ , together with the hypotheses  $K \log K \leq r/2$  and  $(8\lambda)^{2B} \leq K$ , are exactly the input to Lemma 3.10. Hence

$$\mathbb{P}(\exists s \in [K] : \mathcal{W}_s \subseteq \Omega) \geq 1 - \exp(-\sqrt{K}).$$

On this event, fix  $s$  with  $\mathcal{W}_s \subseteq \Omega$ . Inside  $G[\Omega]$ , each  $i \in I_s$  has a *reservoir*

$$\mathcal{R}_i \subseteq N_G^\dagger(i) \setminus \{j_s\}$$

of size  $L = m\beta_d$ . Since  $j_s$  is *light*, Corollary 3.8 gives a *weight-compatible  $m$ -matching rooted tree* at  $j_s$ . All right vertices used by this tree lie in  $\Omega$ , so the tree is contained in  $G[\Omega]$ . □

**3.4. The Finite  $C_4$ -free Case.** The previous proposition assumes that the ambient right side has size  $\lceil \lambda r \rceil$  and that the overlap parameter is explicitly small. The next theorem removes these restrictions for an arbitrary large  $C_4$ -free right side. The proof first thins the right side to an intermediate set  $J_0$  of size  $\lceil \lambda_* r \rceil$ . Uniform thinning preserves the expected overlap parameter, and the bound on the *overlap parameter* for  $C_4$ -free graphs makes the bad thinning event negligible.

**Lemma 3.14** (Uniform thinning preserves sparse overlap). *Let  $G = (I, J, M)$  be finite with  $|J| = n \geq 2$ . Choose  $J_0 \subseteq J$  uniformly from  $\binom{J}{n_0}$ , where  $2 \leq n_0 \leq n$ , and let  $G_0 := G[J_0]$ . Then*

$$\mathbb{E} \eta(G_0) = \eta(G).$$

Consequently, for every  $\alpha > 0$ ,

$$\mathbb{P}(\eta(G_0) > \alpha) \leq \frac{\eta(G)}{\alpha}.$$

*Proof.* Conditional on  $J_0$ , the overlap parameter  $\eta(G_0)$  is the probability that a uniformly chosen ordered pair of distinct right vertices from  $J_0$  overlaps. If we first choose  $J_0$  uniformly and then choose such an ordered pair from  $J_0$ , the resulting ordered pair is uniform over all distinct ordered pairs in  $J$ . Therefore the expected overlap parameter is exactly  $\eta(G)$ . The tail bound is Markov's inequality.  $\square$

**Theorem 3.15** (Finite  $C_4$ -free sampling theorem). *There exists a universal constant  $C_{3.15} \geq 1$  such that the following holds. Fix an integer  $d \geq 2$ , a real  $\rho \in (0, 1]$ , and positive integers  $k, r, n$  with*

$$2 \leq \rho k \leq r \leq k \leq n.$$

*Let  $G = (I, J, M)$  be  $C_4$ -free with  $|I| = k$ ,  $|J| = n$ , and right degrees satisfying*

$$2 \leq |N_G(j)| \leq d \quad \forall j \in J.$$

*Define*

$$K_* := \max \{K \in \mathbb{N}_{\geq 1} : K \log K \leq \Lambda\}, \quad \text{where } \Lambda := \min \left( \frac{r}{2}, \frac{n}{k} \right).$$

*and let  $m_*$  be a positive integer satisfying*

$$\left( \frac{C_{3.15} d^4 m_*}{\rho} \right)^{C_{3.15} d^3 m_*} \leq K_*.$$

*Assume such an  $m_*$  exists. If  $\Omega \subseteq J$  is chosen uniformly from  $\binom{J}{r}$ , then*

$$\mathbb{P}(G[\Omega] \text{ contains a weight-compatible } m_*\text{-matching rooted tree}) \geq 1 - \exp(-\sqrt{K_*}) - \frac{32d}{\log K_*}.$$

**Remark 3.16.** *The constant  $C_{3.15}$  is chosen only so that this simpler displayed condition implies the exact admissibility condition*

$$(8\lambda_*)^{2B_*} \leq K_*$$

*needed in Proposition 3.12, for the  $\lambda_*$  and  $B_*$  defined in the proof below. Thus  $K_*$  is the number of witness-block attempts available after accounting for both the final sample size  $r$  and the ambient ratio  $\theta = n/k$ , while  $m_*$  is the largest tree width whose witness blocks can still be captured with these attempts.*

*After increasing  $C_{3.15}$  if necessary, the existence of  $m_*$  guarantees that both  $K_*$  and  $\Lambda$  are greater than some large universal constants, so that taking logarithms twice is valid and remains positive.*

*Taking logarithms, the admissibility condition for an integer  $m$  is equivalent to*

$$C_{3.15} d^3 m \log \left( \frac{C_{3.15} d^4 m}{\rho} \right) \leq \log K_*.$$

*For  $\Lambda$  large, the definition of  $K_*$  gives*

$$\log K_* \asymp \log \Lambda.$$

*Set  $x = C_{3.15} d^3 m$  and  $y = \log K_*$ . The boundary equation associated with the logarithmic admissibility condition is*

$$y = x \log \left( x \frac{d}{\rho} \right).$$

For  $x$  greater than a sufficiently large universal constant, we have

$$\log\left(\frac{d}{\rho}y\right) = \log\left(\frac{d}{\rho}x \log\left(x\frac{d}{\rho}\right)\right) \asymp \log\left(\frac{d}{\rho}x\right),$$

which in turn implies that

$$x \asymp \frac{y}{\log\left(\frac{d}{\rho}y\right)}.$$

Thus, the largest admissible integer  $m_\star$  satisfies

$$m_\star \asymp \frac{\log \Lambda}{d^3 \log\left(\frac{d}{\rho} \log \Lambda\right)}.$$

*Proof. Step 1: choose the intermediate scale.* We start with selecting the intermediate sample size  $n_0 = \lceil \lambda_\star n \rceil$  which will allow us to apply Proposition 3.12 and checking that the assumptions on  $K = K_\star$  appearing in this proposition are satisfied. Set

$$\lambda_\star := \frac{512d^2\beta_d m_\star}{\rho}, \quad B_\star := 1 + d\beta_d m_\star \quad \theta := \frac{n}{k}.$$

By choosing the universal constant  $C_{3.15}$  sufficiently large, the defining inequality for  $m_\star$ , together with  $\beta_d \leq d^2$ ,  $\rho \leq 1$ , and  $m_\star \geq 1$ , implies

$$(8\lambda_\star)^{2B_\star} \leq K_\star, \quad 8\rho\lambda_\star \leq K_\star, \quad \rho\lambda_\star \geq 512d^2\beta_d.$$

Indeed,  $m_\star \geq 1$  and  $\rho \leq 1$  give  $K_\star \geq (C_{3.15}d^4)^{C_{3.15}d^3}$ . After increasing  $C_{3.15}$  if necessary, the existence of  $m_\star$  therefore implies  $K_\star \geq 3$  and

$$\theta \geq K_\star \log K_\star \geq 32d\beta_d.$$

Since  $K_\star \log K_\star \leq \theta$ , the second inequality gives

$$\lambda_\star r = \rho\lambda_\star k \leq \frac{1}{8}K_\star k \leq \frac{1}{8}\theta k = \frac{n}{8}.$$

With

$$n_0 := \lceil \lambda_\star r \rceil,$$

the bound  $\lambda_\star r \leq n/8$  gives

$$n_0 \leq \lambda_\star r + 1 \leq \frac{n}{8} + 1 \leq n,$$

where the last inequality follows from  $n \geq 2$ .

*Step 2: control overlap after thinning.* Choose  $J_0 \subseteq J$  uniformly from  $\binom{J}{n_0}$  and set  $G_0 := G[J_0]$ . Then  $G_0$  is  $C_4$ -free and its right degrees still lie between 2 and  $d$ . Let

$$\mathcal{E} := \left\{ \eta(G_0) \leq \frac{1}{16K_\star} \right\}.$$

By Lemma 3.14,

$$\mathbb{P}(\mathcal{E}^c) \leq 16K_\star \eta(G).$$

*Step 3: apply the constant-ratio theorem.* Condition on any realization of  $J_0$  in  $\mathcal{E}$ . Also  $\lambda_\star > 2$ , since

$$\lambda_\star = \frac{512d^2\beta_d m_\star}{\rho} \geq 512d^2\beta_d > 2,$$

where  $\rho \leq 1$  and  $m_\star \geq 1$ . Proposition 3.12 applies to  $G_0$  with  $\varepsilon = \rho$ ,  $\lambda = \lambda_\star$ ,  $\eta = 1/(16K_\star)$ , and  $K = K_\star$ : the overlap condition follows from  $\mathcal{E}$ , the bound  $K_\star \log K_\star \leq r/2$  follows from the definition of  $K_\star$ , and the remaining scale conditions were checked in Step 1. Moreover

$$\left\lfloor \frac{\rho\lambda_\star}{512d^2\beta_d} \right\rfloor = m_\star.$$

Thus, if  $\Omega$  is chosen uniformly from  $\binom{J_0}{r}$ , then

$$\mathbb{P}\left( \begin{array}{c} G_0[\Omega] \text{ contains} \\ \text{a weight-compatible } m_\star\text{-matching rooted tree} \end{array} \middle| J_0 \right) \geq 1 - \exp(-\sqrt{K_\star}).$$

Step 4: return to the original right side. For the nested sampling procedure,

$$\mathbb{P}\left(\begin{array}{c} G[\Omega] \text{ fails to contain} \\ \text{a weight-compatible } m_\star\text{-matching rooted tree} \end{array}\right) \leq \mathbb{P}(\mathcal{E}^c) + \exp(-\sqrt{K_\star}) \leq 16K_\star\eta(G) + \exp(-\sqrt{K_\star}).$$

By Lemma 3.4,

$$\eta(G) \leq 2d \frac{k}{n} = \frac{2d}{\theta}.$$

Hence

$$16K_\star\eta(G) \leq 32d \frac{K_\star}{\theta} \leq \frac{32d}{\log K_\star},$$

where the last inequality uses  $K_\star \log K_\star \leq \theta$ .

Finally, nested uniform sampling gives a uniform  $r$ -subset of  $J$ : if first  $J_0$  is uniform in  $\binom{J}{n_0}$  and then  $\Omega$  is uniform in  $\binom{J_0}{r}$ , then  $\Omega$  is uniform in  $\binom{J}{r}$ . This proves the stated bound.  $\square$

#### 4. TRICHOTOMY AND SAMPLING

This section reduces *local sparse four-cycle incidence* to the  $C_4$ -free sampling theorem proved in Section 3. The reduction has two parts. First, we choose an intermediate right set and delete the few sampled vertices that participate in sampled four-cycles; the remaining good set is  $C_4$ -free and still has asymptotically full size. Second, a transfer lemma says that a uniform sample from the original graph contains the same matching rooted tree patterns as a uniform sample from the good  $C_4$ -free subgraph, up to a negligible loss.

##### 4.1. From Local Four-Cycle Sparsity to a Good $C_4$ -free Subgraph.

**Definition 4.1** (Double-overlap). *Column  $j'$  of a matrix  $M$  is a double-overlap partner of  $j$  if*

$$(4) \quad |\text{supp } M_{\cdot j} \cap \text{supp } M_{\cdot j'}| \geq 2.$$

*For a deterministic sequence of matrices  $M^{(k)} \in \mathbb{R}^{k \times n_k}$  for  $k \in \mathbb{N}$ , we say that it satisfies the local sparse double-overlap condition if there are sets*

$$\mathcal{G}_k \subseteq [n_k] \quad \text{and} \quad \varepsilon_k \searrow 0$$

*such that*

$$|\mathcal{G}_k| = (1 - o(1))n_k$$

*and, for every  $j \in \mathcal{G}_k$ ,*

$$\left| \left\{ j' \in [n_k] \setminus \{j\} : |\text{supp } M_{\cdot j}^{(k)} \cap \text{supp } M_{\cdot j'}^{(k)}| \geq 2 \right\} \right| \leq \varepsilon_k \frac{n_k}{k}.$$

Observe that the assumption of the main theorem that the number of pairs of columns overlapping on two or more rows is of order  $o(n_k^2/k)$ , immediately implies *local sparse double-overlap condition* for some sequence of numbers  $\varepsilon_k \rightarrow 0$ .

In graph language, if a pair of columns  $(j, j')$  are *double-overlap partners*, then the corresponding right vertices in the support graph have at least two common left neighbors. In other words, they form a four-cycle  $C_4$  in the support graph. So from the graph-theoretic perspective, we call it the *local sparse four-cycle incidence* condition:

**Definition 4.2** (Local sparse four-cycle incidence). *Let  $G^{(k)} = (I^{(k)}, J^{(k)}, E^{(k)})$  be a sequence of bipartite graphs with*

$$|I^{(k)}| = k, \quad |J^{(k)}| = n_k.$$

*For right vertices  $j, j' \in J^{(k)}$ , we say that  $j'$  is a four-cycle partner of  $j$  if*

$$j' \neq j \quad \text{and} \quad |N_{G^{(k)}}(j) \cap N_{G^{(k)}}(j')| \geq 2.$$

*Equivalently, the two right vertices lie together in at least one 4-cycle.*

*The sequence  $G^{(k)}$  has local sparse four-cycle incidence if there are numbers  $\varepsilon_k \rightarrow 0$  such that the bad set*

$$B_{\text{fc}}^{(k)} := \left\{ j \in J^{(k)} : j \text{ has more than } \varepsilon_k \frac{n_k}{k} \text{ four-cycle partners} \right\}$$

satisfies  $|B_{\text{fc}}^{(k)}| = o(n_k)$ .

Thus the condition is local in the right vertex  $j$ : it allows some exceptional right vertices, but a typical right vertex has very few *four-cycle partners* at the scale  $n_k/k$ .

**Remark 4.3** (Local versus global four-cycle count). *A stronger global hypothesis would be*

$$\text{the number of 4-cycles in } G^{(k)} = o\left(\frac{n_k^2}{k}\right).$$

*This global condition implies local sparse four-cycle incidence by Markov's inequality.*

The next proposition shows that the *local sparse four-cycle incidence* can be upgraded to  $C_4$ -freeness by passing to a subset of a large size in a uniformly chosen sample.

**Proposition 4.4** (Producing a large  $C_4$ -free good subgraph). *Let  $G^{(k)} = (I^{(k)}, J^{(k)}, E^{(k)})$  be a sequence of finite bipartite graphs with*

$$|I^{(k)}| = k, \quad |J^{(k)}| = n_k, \quad \frac{n_k}{k} \rightarrow \infty.$$

*Assume  $G^{(k)}$  has local sparse four-cycle incidence, with corresponding bad sets  $B_{\text{fc}}^{(k)} \subseteq J^{(k)}$  and numbers  $\varepsilon_k \rightarrow 0$ . There is an absolute constant  $c_{4.4} > 0$  with the following property. Let*

$$\tilde{n}_k := \min \left\{ \lceil c_{4.4} \varepsilon_k^{-1} k \rceil, n_k \right\},$$

*and let  $\tilde{J}^{(k)} \subseteq J^{(k)}$  be uniform of size  $\tilde{n}_k$ . With probability  $1 - o(1)$ , there is a subset  $\tilde{J}_{\text{good}}^{(k)} \subseteq \tilde{J}^{(k)}$  such that*

$$\tilde{J}_{\text{good}}^{(k)} \subseteq \tilde{J}^{(k)} \setminus B_{\text{fc}}^{(k)}, \quad G^{(k)}[\tilde{J}_{\text{good}}^{(k)}] \text{ is } C_4\text{-free} \quad \text{and} \quad |\tilde{J}_{\text{good}}^{(k)}| \geq \frac{9}{10} \tilde{n}_k.$$

*Proof.* Let

$$c := c_{4.4}, \quad n := n_k, \quad \varepsilon := \varepsilon_k, \quad B := B_{\text{fc}}^{(k)}.$$

Set  $m := \tilde{n}_k$  and choose  $\tilde{J} := \tilde{J}^{(k)}$  uniformly from  $\binom{J^{(k)}}{m}$ . Since  $m \leq c\varepsilon^{-1}k + 1$ , we have

$$\varepsilon \frac{m}{k} \leq c + o(1).$$

First remove the four-cycle-bad vertices. Let  $Z := |\tilde{J} \cap B|$ . Then

$$\mathbb{E}Z = m \frac{|B|}{n} = o(m),$$

so  $Z = o(m)$  with probability  $1 - o(1)$  by Markov's inequality.

Let  $H$  be the graph on  $J^{(k)} \setminus B$  in which two vertices are adjacent if they are *four-cycle partners*. By the definition of  $B$ ,

$$\Delta(H) \leq \varepsilon \frac{n}{k}, \quad |E(H)| \leq \frac{1}{2} n \varepsilon \frac{n}{k},$$

where  $\Delta(H)$  denotes the maximal degree, and  $E(H)$  is the set of edges of  $H$ . Let  $W$  be the number of edges of  $H$  contained in  $\tilde{J}$ . Then

$$\mathbb{E}W = |E(H)| \frac{\binom{m}{2}}{\binom{n}{2}} \leq C\varepsilon \frac{m^2}{k} \leq C(c + o(1))m,$$

where we recall that  $(a)_q := a(a-1)\cdots(a-q+1)$  is the falling factorial (see (3)). Choosing  $c_{4.4}$  sufficiently small, we may assume that  $\mathbb{E}W \leq m/100$  for all sufficiently large  $k$ .

We next bound the variance of  $W$ . Write  $W = \sum_{e \in E(H)} X_e$ , where  $X_e$  is the indicator of the event  $e \subseteq \tilde{J}$ . If  $e$  and  $f$  are disjoint edges, then sampling without replacement gives

$$\mathbb{E}X_e X_f = \frac{\binom{m}{4}}{\binom{n}{4}} \leq \left( \frac{\binom{m}{2}}{\binom{n}{2}} \right)^2 = \mathbb{E}X_e \mathbb{E}X_f.$$

Thus disjoint edge pairs have nonpositive covariance. Hence

$$\text{Var}(W) \leq \mathbb{E}W + \sum_{\substack{e, f \in E(H): e \neq f \\ e \cap f \neq \emptyset}} \mathbb{E}X_e X_f.$$

The number of ordered pairs of distinct adjacent edges in  $H$  is at most  $n\Delta(H)^2$ , and for each such pair

$$\mathbb{E}X_e X_f = \frac{\binom{m}{3}}{\binom{n}{3}} \leq C \left(\frac{m}{n}\right)^3.$$

Therefore

$$\text{Var}(W) \leq Cm + Cn \left(\varepsilon \frac{n}{k}\right)^2 \left(\frac{m}{n}\right)^3 \leq Cm + C\varepsilon^2 \frac{m^3}{k^2} = o(m^2),$$

where the last step uses  $m \rightarrow \infty$  and  $m \leq c\varepsilon^{-1}k + 1$ . By Chebyshev's inequality,

$$\mathbb{P}\left(W > \frac{1}{20}m\right) = o(1).$$

On the event  $Z = o(m)$  and  $W \leq m/20$ , remove from  $\tilde{J}$  all vertices in  $B$  and then delete one endpoint from each remaining edge of  $H[\tilde{J} \setminus B]$ . The resulting set  $\tilde{J}_{\text{good}}^{(k)}$  is contained in  $\tilde{J}^{(k)} \setminus B_{\text{fc}}^{(k)}$ , has size at least  $m - o(m) - m/20 \geq 9m/10$  for all large  $k$ , and is independent in  $H$ . Hence no two of its vertices are *four-cycle partners*, so  $G^{(k)}[\tilde{J}_{\text{good}}^{(k)}]$  is  $C_4$ -free.  $\square$

**4.2. Average-Degree Trichotomy.** The next proposition is the main structural result. It states that if a sequence of graphs satisfies the conditions of Theorem 1.3, then, with high probability, a random sample contains either an empty column, or two columns with only one entry in each row, or a *weight-compatible  $m_k$ -matching rooted tree* with  $m_k \rightarrow \infty$ .

**Proposition 4.5** (Average-degree trichotomy). *Fix  $\rho \in (0, 1]$  and a constant  $d_{\text{av}} > 0$ . Let*

$$G^{(k)} = (I^{(k)}, J^{(k)}, M^{(k)})$$

*be a deterministic sequence of weighted bipartite graphs, where  $M^{(k)} \in \mathbb{R}^{I^{(k)} \times J^{(k)}}$ . Assume*

$$|I^{(k)}| = k, \quad |J^{(k)}| = n_k, \quad \frac{n_k}{k} \rightarrow \infty.$$

*Assume the following properties hold for all sufficiently large  $k$ :*

- *The support graph of  $G^{(k)}$  has local sparse four-cycle incidence with parameter  $\varepsilon_k \rightarrow 0$ .*
- *The average right degree is at most  $d_{\text{av}}$ :*

$$\frac{1}{n_k} \sum_{j \in J^{(k)}} |N_{G^{(k)}}(j)| \leq d_{\text{av}}.$$

- *The support is nonempty, and the average magnitude of the nonzero weights is at most 1:*

$$\frac{1}{|E^{(k)}|} \sum_{(i,j) \in E^{(k)}} |M_{ij}^{(k)}| \leq 1,$$

*where  $E^{(k)} = \{(i, j) \in I^{(k)} \times J^{(k)} : M_{ij}^{(k)} \neq 0\}$  is the edge set of  $G^{(k)}$ ; see Definition 2.1.*

*Let  $r(k)$  be an integer sequence with  $\rho k \leq r(k) \leq k$ . Then there is a sequence  $m_k \rightarrow \infty$  such that, for a uniform sample*

$$\Omega_k \subseteq J^{(k)}, \quad \text{of size } |\Omega_k| = r(k),$$

*with probability  $1 - o(1)$  at least one of the following holds:*

- (1)  $G^{(k)}[\Omega_k]$  contains an isolated right vertex.
- (2)  $G^{(k)}[\Omega_k]$  contains distinct degree-one right vertices  $j, j'$  with the same unique left neighbor:

$$N_{G^{(k)}}(j) = N_{G^{(k)}}(j') = \{i\}$$

*for some  $i \in I^{(k)}$ .*

- (3)  $G^{(k)}[\Omega_k]$  contains a *weight-compatible  $m_k$ -matching rooted tree* whose right vertices have support degree at most  $40d_{\text{av}}$  and whose incident nonzero weights have absolute value at most  $10d_{\text{av}}$ .

The proof uses the following deterministic bad right-vertex sets. The first two are the low-degree obstruction sets. For a degree cutoff  $\Delta \geq 2$ , the third set is the high-degree bad set. For a weight cutoff  $A > 0$ , the fourth set is the large-entry bad set. After fixing a witness  $\varepsilon_k \rightarrow 0$  for *local sparse four-cycle incidence*, the fifth set is the four-cycle-bad set:

$$\begin{aligned}
B_0^{(k)} &:= \{j \in J^{(k)} : |N_{G^{(k)}}(j)| = 0\}, \\
B_1^{(k)} &:= \{j \in J^{(k)} : |N_{G^{(k)}}(j)| = 1\}, \\
(5) \quad B_{\deg}^{(k)}(\Delta) &:= \{j \in J^{(k)} : |N_{G^{(k)}}(j)| > \Delta\}, \\
B_{\text{amp}}^{(k)}(A) &:= \{j \in J^{(k)} : \exists i \in N_{G^{(k)}}(j) \text{ with } |M_{ij}^{(k)}| > A\}, \\
B_{\text{fc}}^{(k)} &:= \left\{j \in J^{(k)} : j \text{ has more than } \varepsilon_k \frac{n_k}{k} \text{ four-cycle partners}\right\}.
\end{aligned}$$

For all the above bad sets, the last three can be chosen to have relatively small sizes. The size of  $B_0^{(k)}$  and  $B_1^{(k)}$  can be large, but they create the first two alternatives in the conclusion of Proposition 4.5, as shown in the next two lemmas.

**Lemma 4.6** (Many isolated right vertices). *Work under the assumptions and notation of Proposition 4.5. Let  $\Omega_k \subseteq J^{(k)}$  be uniform of size  $r(k)$ , and let  $\delta_k \downarrow 0$  satisfy  $\delta_k k \rightarrow \infty$ . If  $|B_0^{(k)}| \geq \delta_k n_k$ , then  $G^{(k)}[\Omega_k]$  contains an isolated right vertex with probability  $1 - o(1)$ .*

*Proof.* If  $|B_0^{(k)}| + r(k) > n_k$ , then  $\Omega_k$  must contain an isolated right vertex. Assume now that  $|B_0^{(k)}| + r(k) \leq n_k$ . Then the probability that  $\Omega_k$  contains no isolated right vertex is

$$\mathbb{P}(\Omega_k \cap B_0^{(k)} = \emptyset) = \prod_{s=0}^{r(k)-1} \left(1 - \frac{|B_0^{(k)}|}{n_k - s}\right) \leq \left(1 - \frac{|B_0^{(k)}|}{n_k}\right)^{r(k)} \leq \exp\left(-\frac{r(k)|B_0^{(k)}|}{n_k}\right) \leq \exp(-c\delta_k k) = o(1),$$

where  $c > 0$  depends only on  $\rho$ .  $\square$

For the next lemma, we need a stronger condition on  $\delta_k$ , instead of  $\delta_k k \rightarrow \infty$ , we require  $\delta_k^2 k \rightarrow \infty$ .

**Lemma 4.7** (Many degree-one right vertices). *Work under the assumptions and notation of Proposition 4.5. Let  $\Omega_k \subseteq J^{(k)}$  be uniform of size  $r(k)$ , and let  $\delta_k \downarrow 0$  satisfy*

$$\delta_k \frac{n_k}{k} \rightarrow \infty, \quad \delta_k^2 k \rightarrow \infty.$$

*If  $|B_1^{(k)}| \geq \delta_k n_k$ , then with probability  $1 - o(1)$ , the graph  $G^{(k)}[\Omega_k]$  contains distinct degree-one right vertices  $j, j'$  with the same unique left neighbor.*

*Proof.* For  $i \in I^{(k)}$ , let

$$t_i := |\{j \in B_1^{(k)} : N_{G^{(k)}}(j) = \{i\}\}|, \quad T := \sum_i t_i = |B_1^{(k)}|.$$

Since  $\delta_k R_k \rightarrow \infty$ , we have  $T \gg k$ . Let

$$S := |\Omega_k \cap B_1^{(k)}|.$$

Then

$$\mathbb{E}S = r(k) \frac{T}{n_k} \geq \rho \delta_k k \rightarrow \infty.$$

By Hoeffding's comparison theorem for sampling without replacement [13, Theorem 4] (see also [4] for a modern proof), the binomial Chernoff lower-tail bound applies to the hypergeometric random variable  $S$ . Hence

$$\mathbb{P}\left(S \leq \frac{1}{2} \mathbb{E}S\right) \leq \exp(-c\mathbb{E}S) = o(1).$$

In particular,  $S \geq c' \delta_k k$  with probability  $1 - o(1)$ .

Conditional on  $S = s$ , the set  $\Omega_k \cap B_1^{(k)}$  is a uniform  $s$ -subset of  $B_1^{(k)}$ . The no-collision event is the event that this subset contains at most one vertex from each block

$$\{j \in B_1^{(k)} : N_{G^{(k)}}(j) = \{i\}\}, \quad i \in I^{(k)}.$$

Thus

$$\mathbb{P}(\text{no collision} \mid S = s) = \frac{\sum_{A \subseteq I^{(k)}, |A|=s} \prod_{i \in A} t_i}{\binom{T}{s}},$$

where the numerator counts the choices of  $s$  blocks and one degree-one vertex from each chosen block. To bound this probability, note that the numerator is the  $s$ -th elementary symmetric sum of the numbers  $(t_i : i \in I^{(k)})$ . Under the constraint that  $T = \sum_i t_i$  is fixed and  $t_i \geq 0$ , the elementary symmetric sums are maximized when all  $t_i$  are equal, which can be established by showing that averaging any two  $t_i$  and applying the AM-GM inequality will only increase the symmetric sums. This is known as Maclaurin's inequality for elementary symmetric means [12, p. 52], which yields

$$\sum_{A \subseteq I^{(k)}, |A|=s} \prod_{i \in A} t_i \leq \binom{k}{s} \left(\frac{T}{k}\right)^s.$$

The right-hand side is exactly the value when  $t_i = T/k$  for all  $i$ .

An estimate of the conditional probability of the lack of collisions follows the standard birthday paradox argument. More precisely,

$$\mathbb{P}(\text{no collision} \mid S = s) \leq \frac{\binom{k}{s} (T/k)^s}{\binom{T}{s}} = \prod_{i=0}^{s-1} \frac{k-i}{k} \prod_{i=0}^{s-1} \frac{T}{T-i},$$

where

$$\prod_{i=0}^{s-1} \frac{k-i}{k} = \prod_{i=0}^{s-1} \left(1 - \frac{i}{k}\right) \leq \exp\left(-\sum_{i=0}^{s-1} \frac{i}{k}\right) = \exp\left(-\frac{s(s-1)}{2k}\right).$$

Our assumption on  $\delta_k$  implies that  $s = o(T)$ , because  $s \leq r(k) \leq k$ , while

$$\frac{T}{k} \geq \frac{\delta_k n_k}{k} \rightarrow \infty.$$

Hence,

$$\prod_{i=0}^{s-1} \frac{T}{T-i} = \prod_{i=0}^{s-1} \left(1 + \frac{i}{T-i}\right) \leq \exp\left(\sum_{i=0}^{s-1} \frac{i}{T-i}\right) = \exp\left(\frac{s(s-1)}{T}\right).$$

This shows that for every  $c' \delta_k k \leq s \leq r(k)$ ,

$$\mathbb{P}(\text{no collision} \mid S = s) \leq \exp\left(-s(s-1) \left[\frac{1}{2k} - \frac{1}{T}\right]\right) \leq \exp\left(-c'' \frac{s^2}{k}\right)$$

for some universal constant  $c'' > 0$ .

Therefore, on the likely event  $S \geq c' \delta_k k$ , the exponent is at least a constant times  $\delta_k^2 k \rightarrow \infty$ . Thus a collision occurs with probability  $1 - o(1)$ .  $\square$

**Lemma 4.8** (Large-entry right vertices). *Work under the assumptions and notation of Proposition 4.5. Then*

$$|B_{\text{amp}}^{(k)}(10d_{\text{av}})| \leq \frac{1}{10} n_k.$$

*Proof.* Recall that  $E^{(k)} = \{(i, j) \in I^{(k)} \times J^{(k)} : M_{ij}^{(k)} \neq 0\}$ . For each  $j \in B_{\text{amp}}^{(k)}(10d_{\text{av}})$ , choose one edge  $(i_j, j)$  with  $|M_{i_j j}^{(k)}| > 10d_{\text{av}}$ . The chosen edges are distinct. Hence

$$10d_{\text{av}} |B_{\text{amp}}^{(k)}(10d_{\text{av}})| \leq \sum_{j \in B_{\text{amp}}^{(k)}(10d_{\text{av}})} |M_{i_j j}^{(k)}| \leq \sum_{(i, j) \in E^{(k)}} |M_{ij}^{(k)}| \leq |E^{(k)}|,$$

where the last inequality follows from the average magnitude assumption of Proposition 4.5. Further, the average right-degree bound implies

$$|E^{(k)}| = \sum_{j \in J^{(k)}} |N_{G^{(k)}}(j)| \leq d_{\text{av}} n_k.$$

Combining the last two inequalities and dividing by  $d_{\text{av}}$  gives the desired bound.  $\square$

**Lemma 4.9** (Weighted subgraph after trimming). *Work under the assumptions and notation of Proposition 4.5. Let  $\delta_k \downarrow 0$  and assume*

$$|B_0^{(k)}| < \delta_k n_k, \quad |B_1^{(k)}| < \delta_k n_k.$$

*Fix a witness  $\varepsilon_k \rightarrow 0$  for local sparse four-cycle incidence. Set*

$$\tilde{n}_k := \min \{ \lceil c_{4.4} \varepsilon_k^{-1} k \rceil, n_k \}.$$

*Let*

$$\tilde{J}^{(k)} \subseteq J^{(k)}$$

*be uniform with  $|\tilde{J}^{(k)}| = \tilde{n}_k$ .*

*Then, with probability  $1 - o(1)$ , there is a set  $J_{\text{sub}}^{(k)} \subseteq \tilde{J}^{(k)}$  such that*

$$|J_{\text{sub}}^{(k)}| \geq \frac{2}{3} \tilde{n}_k,$$

*the graph  $G^{(k)}[J_{\text{sub}}^{(k)}]$  is  $C_4$ -free, all right support degrees in this subgraph lie between 2 and  $40d_{\text{av}}$ , and all its nonzero weights have absolute value at most  $10d_{\text{av}}$ .*

**Remark 4.10** (Two-stage exposure). *The original uniform choice of  $\Omega_k \in \binom{J^{(k)}}{r^{(k)}}$  is realized equivalently in two stages. First choose  $\tilde{J}^{(k)} \in \binom{J^{(k)}}{\tilde{n}_k}$  uniformly, and then choose  $\Omega_k \in \binom{\tilde{J}^{(k)}}{r^{(k)}}$  uniformly.*

*Proof.* Since the support is nonempty and the average right support degree is at most  $d_{\text{av}}$ , the case  $d_{\text{av}} = 0$  is impossible. Set

$$B_{\text{deg}}^{(k)} := B_{\text{deg}}^{(k)}(40d_{\text{av}}) \quad \text{and} \quad B_{\text{amp}}^{(k)} := B_{\text{amp}}^{(k)}(10d_{\text{av}}).$$

Markov's inequality and the average right support-degree bound give

$$|B_{\text{deg}}^{(k)}| \leq \frac{1}{40} n_k.$$

By Lemma 4.8,

$$|B_{\text{amp}}^{(k)}| \leq \frac{1}{10} n_k.$$

The same witness  $\varepsilon_k$  defines the four-cycle-bad set  $B_{\text{fc}}^{(k)}$  in (5). Set

$$B_{\text{trim}}^{(k)} := B_0^{(k)} \cup B_1^{(k)} \cup B_{\text{deg}}^{(k)} \cup B_{\text{amp}}^{(k)}.$$

The assumptions on  $B_0^{(k)}$  and  $B_1^{(k)}$ , together with the degree bound and the large-entry lemma, imply that for all sufficiently large  $k$ ,

$$|B_{\text{trim}}^{(k)}| \leq \frac{3}{20} n_k.$$

Applying Proposition 4.4 to the support graph of  $G^{(k)}$ , with probability  $1 - o(1)$  there is a set

$$J_{C_4}^{(k)} \subseteq \tilde{J}^{(k)}$$

such that

$$|J_{C_4}^{(k)}| \geq \frac{9}{10} \tilde{n}_k, \quad G^{(k)}[J_{C_4}^{(k)}] \text{ is } C_4\text{-free}.$$

Also, since  $|B_{\text{trim}}^{(k)}| \leq 3n_k/20$ , Hoeffding's comparison theorem for sampling without replacement [13, Theorem 4] (see also [4]) and the binomial Chernoff bound give

$$|\tilde{J}^{(k)} \cap B_{\text{trim}}^{(k)}| \leq \frac{1}{5} \tilde{n}_k$$

with probability  $1 - o(1)$ .

On the intersection of these two high-probability events, define

$$J_{\text{sub}}^{(k)} := J_{C_4}^{(k)} \setminus B_{\text{trim}}^{(k)}.$$

Then in fact

$$|J_{\text{sub}}^{(k)}| \geq \frac{7}{10} \tilde{n}_k,$$

and hence  $|J_{\text{sub}}^{(k)}| \geq (2/3)\tilde{n}_k$  for all large  $k$ . The remaining structural conclusions follow from the definitions of  $B_{\text{trim}}^{(k)}$  and  $J_{C_4}^{(k)}$ .  $\square$

*Proof of Proposition 4.5.* Choose  $\delta_k \downarrow 0$  so that

$$\delta_k \frac{n_k}{k} \rightarrow \infty, \quad \delta_k^2 k \rightarrow \infty.$$

For instance, one can take  $\delta_k = \min\left(\frac{n_k}{k}, \sqrt{k}\right)^{-1/2}$ .

If  $|B_0^{(k)}| \geq \delta_k n_k$ , then Lemma 4.6 gives the first case with probability  $1 - o(1)$ . If  $|B_1^{(k)}| \geq \delta_k n_k$ , then Lemma 4.7 gives the second case with probability  $1 - o(1)$ . It remains to consider the complementary case

$$|B_0^{(k)}| < \delta_k n_k, \quad |B_1^{(k)}| < \delta_k n_k.$$

In particular, this case implies

$$(6) \quad 40d_{\text{av}} \geq 2.$$

Indeed, if  $d_{\text{av}} < 1/20$ , then a standard Markov's inequality argument gives  $|B_0^{(k)}| > 19n_k/20$ , contradicting  $|B_0^{(k)}| < \delta_k n_k$  for all large  $k$ .

Let  $\varepsilon_k \rightarrow 0$  be the diminishing sequence for the *local sparse four-cycle incidence* condition for  $M^{(k)}$ . Set

$$\tilde{n}_k := \min\left\{\lceil c_{4.4} \varepsilon_k^{-1} k \rceil, n_k\right\}, \quad \theta_k := \frac{\tilde{n}_k}{k}.$$

Since  $n_k/k \rightarrow \infty$  and  $\varepsilon_k \rightarrow 0$ , we have  $\theta_k \rightarrow \infty$ . Since  $r(k) \leq k$ , this also gives  $\tilde{n}_k \geq r(k)$  for all sufficiently large  $k$ . Set

$$(7) \quad L_k := \min\left\{k, \frac{n_k}{k}, \frac{1}{\varepsilon_k}\right\}.$$

Then  $L_k \rightarrow \infty$ .

Use the two-stage exposure from Remark 4.10, with  $\tilde{n}_k = \theta_k k$ . The final set  $\Omega_k$  still has the original uniform law on  $\binom{J_{\text{sub}}^{(k)}}{r(k)}$ . Applying Lemma 4.9 to the first stage, with probability  $1 - o(1)$  there is a set  $J_{\text{sub}}^{(k)} \subseteq \tilde{J}^{(k)}$  such that

$$|J_{\text{sub}}^{(k)}| \geq \frac{2}{3}\tilde{n}_k,$$

the graph  $G^{(k)}[J_{\text{sub}}^{(k)}]$  is  $C_4$ -free, every right support degree in this subgraph lies between 2 and  $40d_{\text{av}}$ , and every nonzero weight in this subgraph has absolute value at most  $10d_{\text{av}}$ . Writing

$$Z := |\Omega_k \cap J_{\text{sub}}^{(k)}|,$$

we have, that conditionally on the subgraph,  $Z$  is a hypergeometric random variable with mean at least  $2r(k)/3$ . Hoeffding's comparison theorem [13, Theorem 4] and the binomial Chernoff bound give

$$Z \in [r(k)/2, r(k)]$$

with probability  $1 - o(1)$ . Conditionally on the subgraph and on this value of  $Z$ , the set  $\Omega_k \cap J_{\text{sub}}^{(k)}$  is a uniform  $Z$ -subset of  $J_{\text{sub}}^{(k)}$ . It remains to show that such a uniform subset contains a large *weight-compatible  $m_k$ -matching rooted tree*.

*Finite-subgraph parameters.* Let us fix any such realization  $J_{\text{sub}}^{(k)}$  and  $Z = r_{\text{sub}} \in [r(k)/2, r(k)]$ . The conditional law of  $\Omega_k \cap J_{\text{sub}}^{(k)}$  is a subset of size  $r_{\text{sub}}$  chosen uniformly from  $J_{\text{sub}}^{(k)}$ . We will apply Theorem 3.15 to this conditional law, with the subgraph  $G^{(k)}[J_{\text{sub}}^{(k)}]$  and sample size  $r_{\text{sub}}$ .

Since  $J_{\text{sub}}^{(k)}$  has cardinality at least  $\frac{2}{3}\theta_k k$ , the ratio  $\theta_{\text{sub}} := \frac{|J_{\text{sub}}^{(k)}|}{k}$  satisfies

$$\frac{2}{3}\theta_k \leq \theta_{\text{sub}} \leq \theta_k.$$

Set

$$\Lambda_{\text{sub}} := \min(r_{\text{sub}}/2, \theta_{\text{sub}}).$$

The corresponding parameter  $K_\star$  in Theorem 3.15 is the largest integer  $K$  such that

$$K \log K \leq \Lambda_{\text{sub}}.$$

Moreover,

$$\Lambda_{\text{sub}} \asymp \min(r(k)/2, \theta_k) \asymp L_k,$$

where  $\asymp$  denotes comparability up to a constant factor depending only on  $\rho$  and  $c_{4.4}$ . Consequently,

$$\log K_\star \asymp \log L_k \rightarrow \infty.$$

Set  $d_\star := \lfloor 40d_{\text{av}} \rfloor$ . By (6),  $d_\star \geq 2$ , and every right degree in  $G^{(k)}[J_{\text{sub}}^{(k)}]$  is at most  $d_\star$ . From Remark 3.16, applied with  $d = d_\star$  and  $\rho_{\text{sub}} = r_{\text{sub}}/k$ , the finite theorem's parameter  $m_\star$  is uniformly bounded below by a quantity comparable to

$$\frac{\log L_k}{d_{\text{av}}^3 \log(\frac{d_{\text{av}}}{\rho} \log L_k)} \rightarrow \infty,$$

because  $\rho_{\text{sub}} \geq \rho/2$ , and  $\log K_\star \asymp \log L_k$ . Hence we may choose a deterministic integer sequence  $m_k \rightarrow \infty$  such that, for all relevant finite-subgraph applications,

$$m_k \leq m_\star(r_{\text{sub}}, J_{\text{sub}}^{(k)}).$$

Thus, applying Theorem 3.15 with  $m_k$ , the conditional probability that a uniform  $r_{\text{sub}}$ -subset of  $J_{\text{sub}}^{(k)}$  contains a *weight-compatible  $m_k$ -matching rooted tree* is at least

$$1 - \exp(-\sqrt{K_\star}) - \frac{32d_\star}{\log K_\star} = 1 - o(1),$$

for all relevant  $r_{\text{sub}}$  and subgraph realizations.

Combining this with the conditional distribution of  $\Omega_k \cap J_{\text{sub}}^{(k)}$ , we obtain that the third case occurs with probability  $1 - o(1)$  in the two-stage construction. In particular, we conclude that with high probability  $\Omega_k$  contains a *weight-compatible  $m_k$ -matching rooted tree* with right support degrees at most  $40d_{\text{av}}$  and incident nonzero weights bounded by  $10d_{\text{av}}$ . Since the two-stage construction gives the original uniform law of  $\Omega_k$ , the complementary case is proved. Combining the three cases proves the proposition.  $\square$

## 5. PROOF OF THEOREM 1.3

This section combines the graph cases with the matrix test-vector step.

*Proof of Theorem 1.3.* Denote by  $d_{\text{av}}$  the uniform upper bound on the average degree. We can also assume without loss of generality that the average magnitude of non-zero entries is at most one. If the support of  $M^{(k)}$  is empty along a further subsequence, then  $M_{\Omega_k}^{(k)}$  is the zero matrix and the conclusion is immediate. Thus we may also assume the support is nonempty.

Let  $G^{(k)} = ([k], [n_k], M^{(k)})$  be the weighted bipartite graph associated with  $M^{(k)}$ .

The average column-sparsity bound is exactly the average right support-degree bound for  $G^{(k)}$ , the weight hypothesis is exactly the average-magnitude hypothesis in Proposition 4.5, and the *local sparse double-overlap condition* is exactly the *local sparse four-cycle incidence* condition for its support graph. By Proposition 4.5, there is a sequence  $m_k \rightarrow \infty$  such that, with probability  $1 - o(1)$ , one of the three trichotomy cases occurs for  $G^{(k)}[\Omega_k]$ .

If the sample contains an isolated right vertex, then  $M_{\Omega_k}^{(k)}$  has a zero column, so its least singular value is 0.

If the sample contains distinct degree-one right vertices  $j, j'$  with the same unique left neighbor  $i$ , then the two sampled columns are both nonzero multiples of  $e_i$ . Hence they are linearly dependent, so the sampled matrix has a nontrivial kernel and again has least singular value 0.

It remains to consider the tree case. The *weight-compatible  $m_k$ -matching rooted tree* has right support degrees at most  $40d_{\text{av}}$  and incident nonzero weights bounded by  $10d_{\text{av}}$ . Thus this tree has  $A_{\text{out}} \leq 10d_{\text{av}}$ , and its third left layer has size at most  $(40d_{\text{av}})m_k(40d_{\text{av}} - 1)$ . Use the test vector from Proposition 2.5 on

the selected tree columns and set all other sampled coordinates to zero. Extra sampled columns outside the tree have coefficient zero. Therefore

$$s_{\min}(M_{\Omega_k}^{(k)}) \leq 10d_{\text{av}} \sqrt{\frac{(40d_{\text{av}})(40d_{\text{av}} - 1)}{m_k}} \leq \frac{400d_{\text{av}}^2}{\sqrt{m_k}} = o(1)$$

with probability  $1 - o(1)$ . □

**Remark 5.1** (Slowly growing average degree). *For readability, the main theorems are stated for fixed average degree  $d_{\text{av}}$ . However, the same proof gives Proposition 4.5 and Theorem 1.3 with  $d_{\text{av}}$  replaced by a sequence of non-negative integers  $d_k$ , provided that*

$$d_k^4 \ll m_k \asymp \frac{\log L_k}{d_k^3 \log(\frac{d_k}{\rho}) \log L_k},$$

where we need  $d_k^4 = o(m_k)$  to ensure that the test vector from Proposition 2.5 gives a vanishing upper bound on the least singular value. In particular, this works when  $d_k \leq \log^c L_k$  for some small constant  $c > 0$ , where  $L_k$  is defined in (7).

## 6. SPARSESTACK AND OSI

We now apply Theorem 1.3, together with Remark 5.1 for growing sparsity, to the SparseStack construction from Definition 1.4 and the OSI property from Definition 1.1. The orientation is different from the matrix theorem: SparseStack is an  $n \times k$  matrix  $S$ , while the theorem applies to  $M := S^\top \in \mathbb{R}^{k \times n}$ . Thus rows of  $S$  become columns of  $M$ . Restricting  $S^\top$  to a fixed set of  $r$  coordinates in  $\mathbb{R}^n$  is the same as selecting  $r$  columns of  $M$ . The SparseStack rows are exchangeable, so a deterministic coordinate set can be considered in place of the uniform column sample used in the matrix theorem. The next proposition immediately implies corollary from the introduction. We deliberately consider a more general setting allowing for a growing sparsity parameter:

**Proposition 6.1** (SparseStack satisfies the local sparse double-overlap condition). *Let  $\zeta_k \geq 2$  be an integer sequence such that  $k$  is an integer multiple of  $\zeta_k$ , set  $b(k) := k/\zeta_k$ , and let  $n_k$  satisfy  $n_k/k \rightarrow \infty$ . Further, assume that  $\zeta_k^4/k \rightarrow 0$ , and let  $a_k$  be any sequence such that*

$$\varepsilon_k := a_k \frac{\zeta_k^4}{k} \rightarrow 0; \quad a_k \rightarrow \infty.$$

*Let  $S^{(k)} \in \mathbb{R}^{n_k \times k}$  be a SparseStack matrix with row sparsity  $\zeta_k$  and block size  $b(k)$ , and define  $M^{(k)} := (S^{(k)})^\top \in \mathbb{R}^{k \times n_k}$ . Then the matrix  $M^{(k)}$  has column sparsity  $\zeta_k$  and, with probability  $1 - o(1)$ , satisfies the local sparse double-overlap condition with parameter  $\varepsilon_k$ .*

*Proof.* Each column of  $M^{(k)}$  is the transpose of a row of  $S^{(k)}$ , so it has exactly one nonzero in each of the  $\zeta_k$  blocks and hence exactly  $\zeta_k$  nonzero entries.

Fix distinct column indices  $j, j' \in [n_k]$ . The overlap  $|\text{supp } M_{\cdot j}^{(k)} \cap \text{supp } M_{\cdot j'}^{(k)}|$  counts the number of blocks in which the corresponding SparseStack rows choose the same coordinate. Therefore

$$\mathbb{P}(|\text{supp } M_{\cdot j}^{(k)} \cap \text{supp } M_{\cdot j'}^{(k)}| \geq 2) \leq \binom{\zeta_k}{2} \frac{1}{b(k)^2} \leq \frac{\zeta_k^4}{2k^2}.$$

For each  $j \in [n_k]$ , let  $B_j$  be the number of  $j' \neq j$  whose support intersects the support of  $j$  in at least two coordinates. Conditional on the  $j$ -th SparseStack row, the rows indexed by  $j' \neq j$  remain independent, so the corresponding indicators are independent and have probability at most  $\zeta_k^4/(2k^2)$ . Hence

$$\mathbb{E}B_j \leq \frac{\zeta_k^4 n_k}{2k^2}.$$

Therefore

$$\frac{\mathbb{E}B_j}{\varepsilon_k n_k/k} \leq \frac{1}{2a_k} = o(1)$$

uniformly in  $j$ . By Markov's inequality,  $\mathbb{P}(B_j > \varepsilon_k n_k/k) = o(1)$  uniformly in  $j$ .

Let  $Y_k$  be the number of indices  $j$  for which  $B_j > \varepsilon_k n_k / k$ . Then  $\mathbb{E}Y_k = o(n_k)$ , so  $Y_k = o(n_k)$  with probability  $1 - o(1)$ . On this event, the set

$$\mathcal{G}_k := \left\{ j \in [n_k] : B_j \leq \varepsilon_k \frac{n_k}{k} \right\}$$

has cardinality  $(1 - o(1))n_k$ , and every  $j \in \mathcal{G}_k$  has at most  $\varepsilon_k n_k / k$  *double-overlap partners*. This is exactly the *local sparse double-overlap condition*.  $\square$

Observe that, for any constant sparsity parameter  $\zeta := \zeta_k$ , and in the regime  $n_k/k \rightarrow \infty$ , the above proposition shows that the transpose of the SparseStack matrix model satisfies all the assumptions of Theorem 1.3 with high probability, and Corollary 1.6 follows. Moreover, combined with a quantitative version of the main theorem (see Remark 5.1), it gives the following result where the sparsity parameter is allowed to grow with  $k$  as a small power of  $\log \min(n_k/k, k)$ , assuming polynomial-order aspect ratio:

**Proposition 6.2** (SparseStack with a polylogarithmic row sparsity). *There is a universal constant  $c > 0$  with the following property. Let  $\zeta_k \geq 2$  be an integer sequence with  $\zeta_k \mid k$ , set  $b(k) := k/\zeta_k$ , and let  $r = r(k)$  and  $n_k$  satisfy*

$$\rho k \leq r(k) \leq k \quad \text{for some fixed } \rho \in (0, 1], \quad \frac{n_k}{r(k)} \rightarrow \infty.$$

Set

$$s_k := \min \left\{ k, \frac{n_k}{k} \right\},$$

and assume

$$\zeta_k = O(\log s_k)^c.$$

For each such  $k$ , let  $S^{(k)} \in \mathbb{R}^{n_k \times k}$  be a SparseStack matrix with row sparsity  $\zeta_k$  and block size  $b(k)$ . Let  $T_k \subseteq [n_k]$  be any deterministic subset with  $|T_k| = r(k)$ . Then

$$s_{\min} \left( ((S^{(k)})^\top)_{T_k} \right) = o(1)$$

with probability  $1 - o(1)$ . In particular, the SparseStack with the given parameters is not  $(r(k), 1/2)$ -OSI.

We leave proof details to the interested reader.

## 7. OPEN PROBLEMS

The main structural assumption in this note controls double overlaps between typical pairs of columns. The most natural question is whether this assumption is actually needed.

**Conjecture 7.1** (Unrestricted deterministic sparse sketches). *Fix an integer  $d \geq 2$  and a constant  $C > 0$ . Let  $r(k), n_k \in \mathbb{N}$  satisfy*

$$\frac{k}{r(k)} \rightarrow C, \quad \frac{n_k}{r(k)} \rightarrow \infty.$$

For each  $k$ , let

$$M^{(k)} \in \{-1, 0, 1\}^{k \times n_k}$$

be a deterministic matrix whose every column has exactly  $d$  nonzero entries. If  $\Omega_k \subset [n_k]$  is chosen uniformly from  $\binom{[n_k]}{r(k)}$ , then

$$s_{\min}(M_{\Omega_k}^{(k)}) = o(1)$$

with probability  $1 - o(1)$ .

Our proof does not address the above conjecture since it uses a locally tree-like cancellation pattern built from typical columns having very few *double-overlap partners*. However, we expect that double overlaps cannot create an obstruction to vanishing  $s_{\min}$ .

Another question is to obtain optimal interdependence between the sparsity parameter (number of non-zero elements per column) and the magnitude of the smallest singular value. The question is open even in the setting of random matrices:

**Problem 7.2** (Optimal OSI sparsity for combinatorial random matrices). *Assume that  $n_k \geq k^{1+c}$ , for a small positive constant  $c > 0$ . For each  $k$  let  $M^{(k)}$  be a chosen uniformly among  $\{-1/\sqrt{d_k}, 0, +1/\sqrt{d_k}\}$ -matrices of dimension  $k \times n_k$ , having  $d_k$  non-zero elements in every column. What are optimal necessary and sufficient conditions on the growth of sequence  $d_k$  to guarantee that uniform random  $k \times k/100$  submatrix of  $M^{(k)}$  has  $s_{\min}$  of order  $\Omega(1)$  with high probability?*

In regard to the last problem, it is natural to expect that there is a critical power  $\alpha_* > 0$  such that whenever  $d_k = O(\log^{\alpha_*-\varepsilon})$  for some  $\varepsilon > 0$ , the uniform random  $k \times k/100$  submatrix of  $M^{(k)}$  is poorly invertible, whereas for  $d_k = \Omega(\log^{\alpha_*+\varepsilon})$  it is well invertible (has constant  $s_{\min}$ ) with high probability. We further expect that the parameter  $\alpha_*$  is the same for the above model of randomness and for SparseStack<sup>†</sup>. Determining the exact value of  $\alpha_*$  is an intriguing open question.

## REFERENCES

- [1] J. Bourgain and L. Tzafriri, *Invertibility of “large” submatrices with applications to the geometry of Banach spaces and harmonic analysis*, Israel J. Math. **57** (1987), no. 2, 137–224.
- [2] J. Bourgain and L. Tzafriri, *Restricted invertibility of matrices and applications*, in *Analysis at Urbana, Vol. II (Urbana, IL, 1986–1987)*, London Math. Soc. Lecture Note Ser., vol. 138, Cambridge Univ. Press, Cambridge, 1989, pp. 61–107.
- [3] J. Bourgain and L. Tzafriri, *On a problem of Kadison and Singer*, J. Reine Angew. Math. **420** (1991), 1–43.
- [4] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, Oxford, 2013.
- [5] Chris Camaño, Ethan N. Epperly, Raphael A. Meyer, and Joel A. Tropp, *Faster linear algebra algorithms with structured random matrices*, arXiv:2508.21189, revised August 28, 2025.
- [6] Shabarish Chenakkod, Michał Dereziński, Xiaoyu Dong, and Mark Rudelson, *Optimal embedding dimension for sparse subspace embeddings*, in *Proceedings of the 56th ACM Symposium on Theory of Computing*, 2024.
- [7] Shabarish Chenakkod, Michał Dereziński, and Xiaoyu Dong, *Optimal oblivious subspace embeddings with near-optimal sparsity*, arXiv:2411.08773, 2024.
- [8] Shabarish Chenakkod, Michał Dereziński, and Xiaoyu Dong, *Optimal subspace embeddings: Resolving Nelson–Nguyen conjecture up to sub-polylogarithmic factors*, in *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms*, 2026, pp. 4501–4559.
- [9] Kenneth L. Clarkson and David P. Woodruff, *Low-rank approximation and regression in input sparsity time*, in *Proceedings of the 45th ACM Symposium on Theory of Computing*, 2013, pp. 81–90.
- [10] Michael B. Cohen, *Nearly tight oblivious subspace embeddings by trace inequalities*, in *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2016, pp. 278–287.
- [11] Paul R. Halmos and Herbert E. Vaughan, *The marriage problem*, Amer. J. Math. **72** (1950), 214–215.
- [12] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, 2nd ed., Cambridge University Press, Cambridge, 1952.
- [13] Wassily Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association **58** (1963), no. 301, 13–30.
- [14] Daniel M. Kane and Jelani Nelson, *Sparser Johnson–Lindenstrauss transforms*, Journal of the ACM **61** (2014), no. 1, Article 4.
- [15] Michael W. Mahoney, *Randomized algorithms for matrices and data*, Foundations and Trends in Machine Learning **3** (2011), no. 2, 123–224.
- [16] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava, *Interlacing families II: Mixed characteristic polynomials and the Kadison–Singer problem*, Ann. of Math. (2) **182** (2015), no. 1, 327–350.
- [17] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava, *Interlacing families III: Sharper restricted invertibility estimates*, arXiv:1712.07766, 2017.
- [18] Per-Gunnar Martinsson and Joel A. Tropp, *Randomized numerical linear algebra: Foundations and algorithms*, Acta Numerica **29** (2020), 403–572.
- [19] Xiangrui Meng and Michael W. Mahoney, *Low-distortion subspace embeddings in input-sparsity time and applications to robust linear regression*, in *Proceedings of the 45th ACM Symposium on Theory of Computing*, 2013, pp. 91–100.
- [20] Assaf Naor and Pierre Youssef, *Restricted invertibility revisited*, in *A Journey Through Discrete Mathematics*, Springer, Cham, 2017, pp. 657–691.
- [21] Jelani Nelson and Huy L. Nguyen, *Lower bounds for oblivious subspace embeddings*, arXiv:1308.3280, 2013.
- [22] Jelani Nelson and Huy L. Nguyen, *OSNAP: Faster numerical linear algebra algorithms via sparser subspace embeddings*, arXiv:1211.1002, 2012.
- [23] Alexander Schrijver, *Combinatorial Optimization: Polyhedra and Efficiency*, Algorithms and Combinatorics, vol. 24, Springer-Verlag, Berlin, 2003.
- [24] Daniel A. Spielman and Nikhil Srivastava, *An elementary proof of the restricted invertibility theorem*, Israel J. Math. **190** (2012), 83–91.
- [25] Alex Townsend and Chris Wang, *Oblivious subspace injection is not enough for relative error*, arXiv:2604.10215, 2026.
- [26] Joel A. Tropp, *Comparison theorems for the extreme eigenvalues of a random symmetric matrix*, arXiv:2603.04365, 2026.
- [27] R. Vershynin, *John’s decompositions: selecting a large part*, Israel J. Math. **122** (2001), 253–277.

- [28] David P. Woodruff, *Sketching as a tool for numerical linear algebra*, Foundations and Trends in Theoretical Computer Science **10** (2014), no. 1–2, 1–157.
- [29] Pierre Youssef, *Restricted invertibility and the Banach–Mazur distance to the cube*, Mathematika **60** (2014), no. 1, 201–218.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MISSOURI, COLUMBIA  
*Email address:* `hhuang@missouri.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN  
*Email address:* `rudelson@umich.edu`

DEPARTMENT OF MATHEMATICAL SCIENCES, CARNEGIE MELLON UNIVERSITY  
*Email address:* `ktikhomi@andrew.cmu.edu`