Distances to the Span of Sparse Random Matrices, with Applications to Gradient Coding

Patrick DeMichele ∗  Margalit Glasgow †  Alexander Moreira ‡
May 26, 2021

Abstract

We study a natural question about sparse random matrices which arises from an application in distributed computing: what is the distance from a fixed vector to the column span of a sparse random matrix $A \in \mathbb{R}^{n \times m}$? We answer this question for several ensembles of sparse random matrices in which the average number of non-zero entries per column, $d$, is smaller than $\log(n)$. Key to our analysis is a new characterization of linear dependencies in sparse random matrices. We show that with high probability, in certain random matrices, including rectangular matrices with i.i.d. Bernoulli entries and $m \geq (1+\epsilon)n$, and symmetric random matrices with Bernoulli entries, any linear dependency must be caused by one of three specific combinatorial structures. We show applications of our result to analyzing and designing gradient codes, replication schemes used in distributed machine learning to mitigate the effect of slow machines, called stragglers. We give the first known construction for a gradient code that achieves near-optimal error for both random and adversarial choices of stragglers.

1 Introduction

We study a natural question about sparse random matrices which arises from an application in distributed computing:

Question 1.

What is the distance from a fixed vector to the column span of a sparse random matrix $A \in \mathbb{R}^{n \times m}$? That is, for $v \in \mathbb{R}^n$, what is $\min_{w \in \mathbb{R}^m} |Aw - v|_2$?

While related questions such as the rank of a sparse random matrix [14, 7] or the distance from a random vector to the span of a random matrix [20] have received considerable attention in the discrete random matrix literature, surprisingly little is understood about this question.

Our motivation comes from gradient coding in distributed computing, in which the fixed vector is the all-ones vector. In this application, the matrix $A$ defines a redundant distribution of $n$ tasks to $m$ machines, such that machine $j$ completes task $i$ iff $A_{ij} \neq 0$. We are interested in matrices where on average each machine completes $d$ tasks, where $d$ is small. This translates to matrices $A$ with $d$ non-zero entries per column. We elaborate more on this application and the connection to Question 1 shortly.

Our work addresses Question 1 for several ensembles of random matrices where $m \geq n$ and where $d$, the average number of non-zero entries per column, is smaller than $\log(n)$. We emphasise that due to the sparsity of the matrix, with high probability, the matrix does not have full row-rank, and hence Question 1 is nontrivial. Our main tool in analyzing this distance is a new characterization of all linear dependencies that occur with constant probability in sparse random matrices. This characterization builds upon a phenomenon

∗Stanford University. Email: demichel@stanford.edu.
†Department of Computer Science, Stanford University. Email: mglasgow@stanford.edu. Research supported by NSF award DGE-1656518.
‡Stanford University. Email: amoreira@stanford.edu.
that has been explored in prior work on sparse random matrices: linear dependencies come from small combinatorial structures [10, 9, 23].

1.1 Connection to Gradient Coding

Gradient codes are data replication schemes that can be used to provide robustness against stragglers, machines that are slow or unresponsive [21, 19]. Specifically, such replication schemes can be used to approximately compute the sum $\sum_i f(X_i)$ of the evaluations of a function $f : \mathcal{X} \to \mathbb{R}$ over a set of data points $X_1, \ldots, X_n \in \mathcal{X}$. Typically in machine learning, the function $f$ represents the gradient of a loss function.

To distribute the computation of $\sum_i f(X_i)$ while maintaining robustness against stragglers, the data points $\{X_i\}$ are distributed redundantly among $m$ machines according to an assignment matrix $A_0 \in \mathbb{R}^{n \times m}$. Each machine $j$ is tasked with computing $g_j := \sum_{i\in A_0}^m f(X_i)$, a weighted sum of $f$ over the data is stores. A central server then collects these values $g_j$ for all $j$ which are not stragglers, and linearly combines the $g_j$ using weights $w_j$ to best approximate $\sum_i f_i(X_i)$.

Formally, we define the decoding error of an assignment $A_0$ along with a set of stragglers $S \subset [m]$ to be the squared distance between the all-ones vector, $\mathbb{1}$, and the span of the columns $\{A_0\}_j$ for $j \notin S$:

$$\text{err}(A_0, S) := \min_{w: w_j = 0 \forall j \in S} \|A_0 w - \mathbb{1}\|^2_2. \quad (1)$$

The decoding error immediately gives us a bound on the approximation error of $\sum_i f(X_i)$. Let $f^*$ be the vector in $\mathbb{R}^n$, where $f^*_i := f(X_i)$. Then for any $w \in \mathbb{R}^m$,

$$\sum_j w_j g_j - \sum_i f(X_i) = f^{*T} A_0 w - f^{*T} \mathbb{1} = f^{*T} (A_0 w - \mathbb{1}). \quad (2)$$

Hence for any $S$, there exists a vector $w \in \mathbb{R}^{[m]\setminus S}$ such that

$$\left( \sum_{j \in [m] \setminus S} w_j g_j - \sum_i f(X_i) \right)^2 \leq \text{err}(A_0, S) |f^*|^2_2. \quad (3)$$

In this work, we will design assignment matrices which have small error when the set of stragglers $S$ is chosen either adversarially or randomly. In the case where $S$ is an adversarially chosen $p$ fraction of the machines, we consider the adversarial decoding error of $A_0$:

$$\max_{S \in \binom{[m]}{\lfloor \gamma m \rfloor}} \{\text{err}(A_0, S)\}.$$

If $S$ is a random $p$ fraction of the machines, we consider the expected decoding error of $A_0$:

$$\mathbb{E}_{S \sim \binom{[m]}{\lfloor \gamma m \rfloor}} \text{err}(A_0, S).$$

Analyzing the decoding error under a random set of stragglers amounts to understanding Question [1] when the matrix $A$ is the random matrix achieved by deleting $pm$ random columns from $A_0$.

1.2 Contributions

Our work considers three ensembles of sparse random matrices in the regime $d \leq \log(n)$. The first is an $n \times \gamma n$ rectangular matrix with i.i.d. Bernoulli entries with parameter $d/n$ for $\gamma > 1$ and $d \geq d_0(\gamma)$. The second is a symmetric matrix with i.i.d. Bernoulli entries with parameter $d/n$ in the upper diagonal portion, for $d = \omega(1)$. The third ensemble arises from deleting random columns from the bi-adjacency matrix of a random $(d\gamma, d)$-biregular bipartite graph for $d \geq d_0(\gamma)$. The analysis of the third ensemble is useful for constructing improved gradient codes.

Our contributions are as follows.

2
1. A tight combinatorial characterization of linear dependencies in the three ensembles of random matrices listed above. Our results consider the structure of all minimal linear row dependencies in a matrix, sets of rows which are linearly dependent, but for which any strict subset of the rows is linearly independent (see Definition 1). We show in Propositions 1, 2 and 3 that with high probability, for the three ensembles of matrices we consider, all minimal linear dependencies among $k$ rows have exactly $2k - 2, 2k - 1$, or $2k$ total non-zero entries among them. We further characterize the structure of these dependencies by viewing the linearly dependent rows as the incidence matrix of a hypergraph (see Figure 1).

2. Progress on Question 1 for random matrices with $m \geq n$. We show how to bound with high probability the distance from the all-ones vector, or more generally, from any incompressible vector, to the column span of any random matrix where our characterization of row dependencies applies.

3. A novel gradient code construction with improved decoding error guarantees. We construct a gradient code called the Augmented Biregular Code (ABC), and prove that it achieves an expected decoding error of $np^{d-o(d)}$ and adversarial decoding error of $\Theta(np/d)$. To our knowledge, the ABC is the first gradient code proved to simultaneously have $O(np/d)$ adversarial decoding error and $np^{d-o(d)}$ expected decoding error. Existing lower bounds show that the adversarial error of a code must be $\Omega(np/d)$, while the expected decoding error under random stragglers is at least $np^d$. We describe this construction and result in more detail in Section 9.

1.3 Related Work

In the discrete random matrix theory literature, understanding the probability of invertibility and the rank of discrete random matrices is an active area of research [23, 9, 10, 6, 14, 22, 16, 15]. A phenomenon core to many of these works is that small structures are the primary reason for linear dependencies in discrete random matrices. For instance, in matrices with Bernoulli entries with parameter $p$, for $\log(n)/n \leq p \leq 1/2$, [22] and [16] show that the probability of invertibility is nearly exactly the probability that there is an all-zero row or column (or two equivalent rows and columns when $p = 1/2$). Further work [7] shows that the probability of having corank $k$ is nearly the probability that $k$ rows are all zero.

Most similar to our work are two papers by Costello and Vu [10, 9], which consider the adjacency matrices of sparse graphs where $p = \frac{c \log(n)}{n}$ for a constant $c \leq 1$. In this line of work, they show that with high probability, all minimal dependencies are due to non-expanding sets, small groups of $k < 1/c$ vertices which are incident to at most $k - 1$ distinct vertices. Our result for symmetric random matrices, Proposition 2, refines this result. In particular, our result pertains to more broadly to any $p = \omega(\frac{1}{n})$, and we refine the type of minimal dependencies that exist to “tree-dependencies” (Figure 1(a)), which are a subclass of non-expanding sets.

The question of the distance between a fixed unit vector and the span of a random matrix has received little attention in the literature. It is well understood that for rotationally invariant matrices — such as matrices with Gaussian entries — the distance squared is close to the corank of the matrix with high probability. Some work on universality for deformed Wigner matrices [18] suggests that this should extend more generally to mean-zero random matrices with subgaussian entries. However, we are unaware of any progress on this question for sparse random matrices.

Gradient coding, introduced in [21], originally considered regimes where $\sum_i f(X_i)$ could be recovered exactly under adversarial stragglers. For assignment matrices $A \in \mathbb{R}^{n \times n}$ that can recover $\sum_i f(X_i)$ exactly under any choice of $pn$ stragglers, the number of data points at each machine, $d$, must exceed $pn$. Closer to our work is a line of work on approximate gradient codes, introduced in [19], which considers the adversarial or expected decoding error in regimes where $A$ is too sparse to recover $\sum_i f(X_i)$ exactly. One important assignment matrix used here is a block matrix based on a Fractional Repetition Code [21], which achieves the optimal expected decoding error of $np^d$. Despite its optimality under random stragglers, the FRC of [21] achieves an adversarial decoding error of $np$. This is a significantly larger than the adversarial error of $\Theta(np/d)$ achieved by the expander-graph based assignment matrix of [19]. Several works [13, 4] have aimed to design assignment matrices that have small adversarial and expected decoding error, but to our knowledge, no existing assignments have achieved adversarial error that decays in $d$ (at any rate) while simultaneously achieving random error that decays faster than inversely linearly in $d$. In Section 9 we
include a table (Table 2) comparing the adversarial and expected decoding errors of our work to other work on approximate gradient coding.

1.4 Organization

In Section 2, we formally define the three ensembles of random matrices we consider in this work. In Section 3, we formally state our results. In Section 4, we give a high level overview of the proof. We first show how to reduce our theorems on the distance from $v$ to the span of $A$ to our results on characterizing linear dependencies of $A$. We give an overview of the proof of our characterization results, which breaks down into two main cases: a “small” case which concerns kernel vector with small support, and a “large” case which concerns kernel vectors with large support.

In Section 5, we state some lemmas that will be used in the small and large cases for more than one result for gradient codes, and show how the ABC ensemble achieves near-optimal decoding error for both random and adversarial stragglers.

2 Notation and Formal Set-up

In this work, we consider the following three ensembles of random matrices. We refer to two of these ensembles as “codes” to follow the gradient coding literature.

1. Rectangular Bernoulli Matrix (Bernoulli Gradient Code) Let $A \sim \text{BGC}(n, \gamma, d)$ denote a random matrix in $\{0, 1\}^{n \times \gamma n}$ where each entry of $A$ is Bernoulli with parameter $d/n$. We will consider this ensemble for any $\gamma > 1$ and $d \geq d_0(\gamma)$.

2. Symmetric Bernoulli Matrix Let $A \sim \text{SB}(n, d)$ denote a random symmetric matrix in $\{0, 1\}^{n \times n}$ whose upper diagonal entries are i.i.d. Bernoulli random variables with parameter $d/n$, and whose diagonal is 0. We will consider this ensemble of matrices for $d = \omega(1)$.

3. Augmented Biregular Code The Augmented Biregular Code (ABC) is based on the adjacency matrix of a random biregular graph generated from the configuration model.

Formally, consider the following process to generate a random matrix $A_0 \in \{0, 1\}^{n \times \gamma n}$ from the distribution $\text{ABC}(n, \gamma, d)$, for $\gamma > 1$. Create $n$ row-nodes and $\gamma n$ column-nodes and associate to each row-node $\gamma d$ half-edges and to each column node $d$ half-edges. Create a multigraph $G$ by choosing uniformly random pairing of the $\gamma dn$ half-edges from the row-nodes to the $\gamma dn$ half-edges from the column-nodes. Given this bipartite graph, we will take $A_0 \in \{0, 1\}^{n \times \gamma n}$ to be the matrix where $(A_0)_{ij} = 1$ if there is at least one edge from node $i$ to $j$.

We will study the ensemble of random matrices obtained by removing a random $p$ fraction of columns from this ABC matrix $A_0$. Formally, we call this ensemble $\text{ABC}_p(n, \gamma, d)$ such that $A \sim \text{ABC}_p(n, \gamma, d)$ is an $n \times \gamma n(1 - p)$ matrix formed by deleting $\gamma np$ random columns from an ABC matrix $A_0 \sim \text{ABC}(n, \gamma, d)$, which has dimensions $n \times \gamma n$.

For a positive integer $n$, we use $[n]$ to denote the set $\{1, \ldots, n\}$. All big-O notation denotes limiting behaviour as $n \to \infty$.

We use $\text{Span}(A)$ to mean the span of the columns of a matrix $A$. For a vector $v \in \mathbb{R}^n$ and a set $S \subseteq [n]$, let $v_S$ denote the vector restricted to entries indexed by elements of $S$. Similarly, for a matrix $A \in \mathbb{R}^{n \times m}$ and a set $S \subseteq [n]$, let $A_S$ denote the matrix $A$ restricted to the rows in the set $S$. For $j \leq m$, let $A^{(j)}$ denote the matrix restricted to the first $j$ columns of $A$. Let $A_j$ denote the $j$th column of $A$. For a symmetric matrix $A \in \mathbb{R}^{n \times n}$, for $j \in [n]$, let $A^{(j)}$ denote the $n - 1 \times n - 1$ matrix equal to $A$ with its $j$-th row and column removed. We refer to a matrix $A$ as the vertex-edge incidence matrix of a graph $G$ if each there is a bijection between edges of $G$ and columns of $A$ that maps an edge $(i, j)$ to a column with ones in locations $i$ and $j$ and zeros elsewhere. Let $e_i$ denote the $i$th canonical basis vector with a 1 in position $i$.

We use the notation $D_{KL}(p||q)$ to refer to the KL divergence of two Bernoulli random variables with parameters $p$ and $q$ respectively.
3 Statement of Results

In this section, we state our main results highlighted in the contributions. We begin with a few definitions, which we summarize in Table 1.

**Definition 1.** A matrix in \( B \in \mathbb{R}^{k \times m} \) is a minimal linear dependency if it satisfies the following two properties:

1. There exists \( x \in \mathbb{R}^k \) such that \( x^T B = 0 \) and \( \text{supp}(x) = [k] \).
2. \( B \) has rank \( k - 1 \).

Let \( \mathcal{M}_k \subseteq \bigcup_{m \geq 1} \mathbb{R}^{k \times m} \) denote matrices of height \( k \) which are minimal dependencies. We will distinguish among these matrices three particular types of linear dependencies, which are illustrated in Figure 1.

**Definition 2** (Tree dependency). Define \( \mathcal{T}_k \) as the set of matrices \( B \in \bigcup_{m \geq 1} \{0, 1\}^{k \times m} \) with exactly \( 2k - 2 \) non-zero entries such that the non-zero columns of \( B \) form the vertex-edge incidence matrix of a tree on \( k \) vertices.

**Definition 3** (Two-forest dependency). Let \( \mathcal{T}_k^+ \) be the set of matrices \( B \in \bigcup_{m \geq 1} \{0, 1\}^{k \times m} \) with exactly \( 2k - 1 \) non-zero entries satisfying the following:

1. \( B \) has \( k - 1 \) non-zero columns: \( k - 2 \) columns supported on 2 entries and one column supported on 3 entries.
2. The submatrix of \( B \) restricted to the columns of support 2 is the vertex-edge incidence matrix of a forest \( F \) with two connected components \( F_1, F_2 \).
3. The column of support 3 contains 1’s at rows \( a, b, c \) where \( a, b \in F_1 \) are connected by an even-length path, and \( c \in F_j \) for \( \{i, j\} = \{1, 2\} \).
Set of minimal dependencies with $k$ rows (Definition 1)

Set of tree dependencies of $k$ rows (Definition 2, Figure 1(a))

Set of two-forest dependencies of $k$ rows (Definition 3, Figure 1(b))

Set of tree-with-added-edge dependencies of $k$ rows (Definition 4, Figure 1(c))

Subset of matrices in $\mathcal{M}_k$ with exactly $L$ ones. (Definition 5)

Subset of matrices in $\mathcal{M}_k$ with exactly $L$ ones and exactly $k$ nonzero columns. (Definition 6)

| $\mathcal{M}_k$ | Set of minimal dependencies with $k$ rows (Definition 1) |
|-----------------|---------------------------------------------------------|
| $\mathcal{T}_k$ | Set of tree dependencies of $k$ rows (Definition 2, Figure 1(a)) |
| $\mathcal{T}_k^+$ | Set of two-forest dependencies of $k$ rows (Definition 3, Figure 1(b)) |
| $\mathcal{T}_k^C$ | Set of tree-with-added-edge dependencies of $k$ rows (Definition 4, Figure 1(c)) |
| $\mathcal{S}_{L,k}$ | Subset of matrices in $\mathcal{M}_k$ with exactly $L$ ones. (Definition 5) |
| $\mathcal{S}_{L,k}'$ | Subset of matrices in $\mathcal{M}_k$ with exactly $L$ ones and exactly $k$ nonzero columns. (Definition 6) |

### Table 1: Notation in this work

**Definition 4** (Tree-with-added-edge dependency). Define $\mathcal{T}_k^C$ as the set of $B \in \bigcup_m \{0,1\}^{k \times m}$ with $2k$ non-zero entries such that the non-zero columns of $B$ form the vertex-edge incidence matrix of a tree on $k$ vertices, with an added edge between two vertices in the tree of odd distance from each other. The additional edge may create a multi-edge in the this graph.

#### 3.1 Characterization of Linear Dependencies

Our main technique in addressing Question 1 is a new characterization of the linear dependencies that occur among the rows of a random matrix $A$. The following three theorems describe our main results:

**Theorem 1** (Characterization BGC). There exists a universal constant $c$ such that for any $\gamma > 1$ and $d \geq d_0(\gamma)$, for $A \sim \text{BGC}(n, \gamma, d)$, with probability $1 - o(1)$:

1. All minimal dependencies of $k$ rows of $A$ are in $\mathcal{T}_k \cup \mathcal{T}_k^+ \cup \mathcal{T}_k^C$.

2. The number of rows involved in a linear dependency of $A$ is at most $ne^{-\gamma d + c\log(\gamma d)}$, that is

$$\left| \bigcup_{x : x^T A = 0} \text{supp}(x) \right| \leq ne^{-\gamma d + c\log(\gamma d)}.$$  

**Theorem 2** (Characterization Symmetric Bernoulli). Let $A \sim \text{SB}(n, d)$, where $d = \omega(1)$. Then with probability $1 - o(1)$,

1. All minimal dependencies of $k$ rows of $A$ are in $\mathcal{T}_k$.

2. The number of rows involved in a linear dependency of $A$ is at most $ne^{-d + o(d)}$, that is

$$\left| \bigcup_{x : x^T A = 0} \text{supp}(x) \right| \leq ne^{-d + o(d)}.$$  

**Theorem 3** (Characterization ABC). There exist universal constants $c$ and $\gamma_0$ such that for any constant $p < 1/2$, $\gamma > \gamma_0$, and $d \geq d_0(\gamma)$, for $A \sim \text{ABC}_p(n, \gamma, d)$, with probability $1 - o(1)$:

1. All minimal dependencies of $k$ rows of $A$ are in $\mathcal{T}_k \cup \mathcal{T}_k^+ \cup \mathcal{T}_k^C$.

2. The number of rows involved in a linear dependency of $A$ is at most $np^{\gamma d - c\log(\gamma d)}$, that is

$$\left| \bigcup_{x : x^T A = 0} \text{supp}(x) \right| \leq np^{\gamma d - c\log(\gamma d)}.$$  

**Remark 1.** With high probability, one can show that under the conditions of Theorem 1, $A \sim \text{BGC}(n, \gamma, d)$ does have minimal dependencies in $\mathcal{T}_k$ for $k \lesssim \frac{\log(n)}{\gamma d}$. We expect the same to hold for the symmetric matrix and for the ABC. For constant $d$, with constant probability, $A \sim \text{BGC}(n, \gamma, d)$ has minimal dependencies in $\mathcal{T}_k^+$ and $\mathcal{T}_k^C$ for $k$ sufficiently small, including $k = 5$. We expect the same to hold for the ABC.
Remark 2. While our initial motivation was Question 1, we believe these characterization theorems may have implications for understanding the exact rank of a sparse random matrices. For instance, in Corollary 7 in the appendix, we show for $A \sim \text{SB}(n, d)$ with $d = \omega(1)$, with probability $1 - o(1)$, the rank of $A$ is exactly equal to the graph-theoretic 2-core rank bound.

3.2 Progress on Question 1

The following three bound the distance from $v$ to the column span of the random matrices we study.

**Theorem 4** (BGC Distance). Let $A \sim \text{BGC}(n, \gamma, d)$. For $\gamma > 1$ and $d \geq d_0(\gamma)$, there exists a constant $c$ such that with probability $1 - o(1)$,

$$(1 - o(1)) e^{-\gamma d} \leq \frac{1}{n} \min_w |Aw - 1|_2^2 \leq e^{-\gamma d + c \log(\gamma d)}. \quad (4)$$

**Theorem 5** (Square Bernoulli Distance). Let $A \sim \text{SB}(n, d)$. For $d = \omega(1)$, with probability $1 - o(1)$,

$$(1 - o(1)) e^{-d} \leq \frac{1}{n} \min_w |Aw - 1|_2^2 \leq e^{-d + o(d)}. \quad (5)$$

**Theorem 6** (ABC Distance). Let $A \sim \text{ABC}_p(n, \gamma, d)$. For any $p < 1/2$, there exists constants $c$, $\gamma_0$ and $d_0$ such that for $\gamma \geq \gamma_0$ and $d \geq d_0$, with probability $1 - o(1)$,

$$(1 - o(1)) p^{\gamma d} \leq \frac{1}{n} \min_w |Aw - 1|_2^2 \leq p^{\gamma d + c \log(\gamma d)}. \quad (6)$$

**Remark 3** (Distances to arbitrary vectors: good news). These results can be extended beyond $\mathbb{I}$ to unit vectors $v \in \mathbb{R}^n$ whose mass is well-distributed among the coordinates of $v$. Formally, suppose $v$ is a distance of at least $\rho$ from any $\delta$-sparse unit vector, that is, $v$ is $(\delta, \rho)$-incompressible.

Then under the conditions of Theorem 5, with probability $1 - o(1)$,

$$\frac{\rho^2 \delta}{8} e^{-d} \leq \min_w |Aw - v|_2^2 \leq \max(n|v|_2^2 e^{-d + o(d)}), 1) \quad (7)$$

Similar results hold for $A \sim \text{BGC}$ or $A \sim \text{ABC}_p$, with the exponent of $e$ begin the same exponent as in the respective theorems.

**Remark 4** (Distances to arbitrary vectors: bad news). For arbitrary vectors $v$, we cannot hope to prove similar high probability bounds as for $\mathbb{I}$. For instance, if $v$ is a $O(1)$-sparse unit vector, then under the conditions of Theorems 4, 5, 6 with constant probability, the distance from $v$ to the span of $A$ is 0. Similarly, under the conditions of Theorems 4 or 6 with constant probability (that may depend on $d$), the distance from $v$ to the span of $A$ is 1.

The statements in these remarks are evident from the proofs of the theorems above, described in Section 3.2.2

3.3 Gradient Coding Results

We are able to use Theorem 6 on the distance between $\mathbb{I}$ and $A \sim \text{ABC}_p$ to analyze the expected decoding error of an assignment matrix based on the ABC ensemble.

Since most results in the gradient coding literature concern a square assignment matrices where $m = n$, we design an $n \times n$ assignment matrix $B \sim \text{ABC}_{\text{stacked}}(n, \gamma, d)$ by stacking together $\gamma$ copies of a rectangular matrices $A_\gamma \sim \text{ABC}(n/\gamma, \gamma, d/\gamma)$ for an appropriate choice of $\gamma$. Full details of the construction are given in Section 3. We prove the following theorem about the assignment matrix $B$:

**Theorem 7.** Let $c, \gamma_0, d_0$ be the universal constants from Theorem 6. Choose any $\gamma, d \in \mathbb{Z}^+$ such that $\gamma \geq \gamma_0$, $|d|$ and $\frac{d}{\gamma} \geq d_0$. For any sufficiently large $n$ divisible by $\gamma$, let $B \sim \text{ABC}_{\text{stacked}}(n, \gamma, d)$. Then with constant probability over the choice of $B$:

$$\frac{1}{n} \mathbb{E}_{S \sim (n)} \text{err}(B, S) \leq p^{d - c \log(d)} + o(1), \quad (8)$$
\[ \frac{1}{n} \max_{S \in \binom{[n]}{\ell}} \{ \text{err}(B, S) \} \leq \left( \frac{8 \gamma^3 p}{d} \right) + o(1). \]

\section{Overview of Proofs}

In the next subsection, we give an overview of how we bound the distance from \( \mathbf{1} \) to the span of \( A \) to yield Theorems \ref{thm:small}–\ref{thm:medium} and \ref{thm:large}. In the following subsection, we give an overview of our proof of the characterization results.

\subsection{Bounding the Distance from \( \mathbf{1} \) to the Span of \( A \)}

Given a matrix \( A \), we will partition its rows into two sets: \( D \), the set of all rows involved in a linear dependency, and \( [n] \setminus D \). Formally,

\[ D = \bigcup_{x : x^T A = 0} \text{supp}(x). \]

We will use the following lemma to bound the distance from a vector \( v \) to the column span of \( A \).

\begin{lemma}
Let \( A \in \mathbb{R}^{n \times m} \), and let \( D = \bigcup_{x : x^T A = 0} \text{supp}(x) \) be the set of rows which are involved in a linear dependency. Then for any \( v \in \mathbb{R}^n \), we have

\[ \min_{w \in \mathbb{R}^m} |Aw - v|_2^2 \leq |v_D|_2^2. \]
\end{lemma}

\begin{lemma}
Let \( A \in \mathbb{R}^{n \times m} \), and let \( D = \bigcup_{x : x^T A = 0} \text{supp}(x) \). Then for any \( i \not\in D \) we have \( e_i \in \text{Span}(A) \).
\end{lemma}

To see how Lemma \ref{lem:1} follows from Lemma \ref{lem:2}, for any \( v \), by Lemma \ref{lem:2}, the vector \( v' := \sum_{i \in [n] \setminus D} v_i e_i \) \( \in \text{Span}(A) \). Hence

\[ \min_w |Aw - v'|_2^2 \leq |v - v'|_2^2 = |v_D|_2^2, \]

establishing Lemma \ref{lem:1}. For \( v = \mathbf{1} \), we have

\[ \min_w |Aw - \mathbf{1}|_2^2 \leq |D|. \]

Plugging in the bound on \( |D| \) given in the characterization theorems from Section \ref{sec:char} yields the upper bounds in the distance theorems in Section \ref{sec:dist}.

The lower bounds on the distance are given by counting the number of of all-zero rows in \( A \). Notice that the squared distance between \( \mathbf{1} \) and the span of \( A \) is at least the number of all-zero rows in \( A \). We formally prove these lower bounds using standard concentration tools in Section \ref{sec:conc}.

\subsection{Overview of Proof of Characterization}

Our proof of each characterization result is divided into either two or three main cases: a “small” case, a “large” case, and sometimes a “medium” case. The small case proves that for some constant \( c_s \), for \( k \lesssim \frac{n}{d} \), with high probability, all minimal dependencies of \( k \) rows in \( A \) are in \( T_k \cup T^+_k \cup T^-_k \). The large case proves that for a second constant \( c_L \), for \( k \gtrsim \frac{n}{d^2} \), with high probability there are no minimal dependencies of \( k \) rows in \( A \). The exact constants \( c_s \) and \( c_L \) depend on the particular ensemble of random matrices. We require a medium case when \( c_s > c_L \), in order to account for all \( k \).

While our proofs are organized by the specific ensemble of random matrices, we give here a short overview of the techniques in the small and large cases, as they are similar among all three ensembles we study.

The main idea in proving this characterization of dependencies in the small and medium cases comes from the following two trivial observations.

\begin{observation}
Let \( B \in \{0, 1\}^{k \times m} \) be a matrix. Then \( B \) cannot be a minimal dependency if some column of \( B \) contains exactly one 1.
\end{observation}
Proof. Recall that our matrices are 0/1 valued, and thus, if there is only one row with value 1 at index \(j\), that row is linearly independent to all other rows in our subset. Hence, no such dependency exists. \(\square\)

**Observation 2.** Let \(B \in \{0,1\}^{k \times m}\) be a matrix. If \(B\) has fewer than \(2k - 2\) entries that are 1, then \(B\) is not a minimal dependency.

**Proof.** The requirement that \(B\) is rank \(k - 1\) means that at least \(k - 1\) columns must have at least one 1 in them and it follows from the previous observation that for a minimal dependency to exist, each of these columns must have at least two ones in them. Thus, we need at least \(2k - 2\) entries which are 1 in \(B\) before a minimal dependency can exist. \(\square\)

4.2.1 Small Case

The goal of the small case is to prove that with high probability all small minimal row dependencies \((k \lesssim n/d^c)\) are contained in \(T_k \cup T_k^+ \cup T_k^C\). We begin by selecting an arbitrary set \(S\) of \(k\) rows, which induces a submatrix \(A_S\). Then, by conditioning on \(L\), which is the number of 1s which appear in \(A_S\), we consider a random process derived from the distribution of \(A\) which places these \(L\) 1s in the submatrix one by one. Due to Observation 2, we only must consider the case when \(L \geq 2k - 2\).

By Observation 1, \(A_S\) does not have a minimal dependency if there exists a column with exactly one 1 in \(A_S\). To lower bound the probability of this event, we consider a random walk which increases by 1 every time our random process places a 1 in an already occupied column and stays constant otherwise. As long as the value of this random walk is less than \(L/2\) at the time we have placed of the \(L\) 1s in the submatrix, we know that \(A_S\) does not have a minimal dependency.

For the case of symmetric matrices, we modify this argument slightly by coupling with a random walk which also increases every time a 1 is placed in the “symmetric” portion of \(A_S\), that is, a column of \(A_S\) indexed by an element of \(S\).

4.2.2 Large Case

The goal of the large case is to show that with high probability, for \(k \gtrsim n/d^c\), there are no minimal dependencies of \(k\) rows. Our main tool in ruling out large dependencies is the following set of anti-concentration results, most of which are standard in the literature. Roughly speaking, these results state that the dot product of a random vector and a deterministic vector will not concentrate on any one value with too large probability.

Each ensemble of random matrices we study requires an anti-concentration lemma tailored to the distribution of a column vector from that matrix.

In the BGC matrix, we use the following version of the Littlewood-Offord theorem for sparse random vectors.

**Lemma 3** (c.f. [10] Lemma 8.2). Let \(v \in \mathbb{R}^n\) be a deterministic vector with support at least \(m\). Let \(z \in \mathbb{R}^n\) be the random vector with i.i.d. Bernoulli entries with parameter \(p \leq 1/2\). Then for any fixed \(c\),

\[
\Pr[v^T z = c] \leq \left(\frac{1}{\sqrt{\pi/2}}\right) \frac{1}{\sqrt{mp}} + \left(e^{\ln(2) - 1)mp}\right).
\]

In particular, for \(mp \geq 9\), we have:

\[
\Pr[v^T z = c] \leq \frac{1}{\sqrt{mp}}.
\]

In the symmetric matrix case, we will additionally use a quadratic Littlewood-Offord result, originally due to [8]. We state a version for sparse random vectors from [10].

**Lemma 4** (c.f. [10] Lemma 8.4). Let \(M \in \mathbb{R}^{n \times n}\) be a deterministic matrix with a least \(m\) non-zero entries in each of \(m\) distinct columns of \(M\). Let \(z \in \mathbb{R}^n\) be the random vector with i.i.d. Bernoulli entries with parameter \(p \leq 1/2\). Then for any fixed \(c\),

\[
\Pr[(z^T M z = c) = O\left(\frac{1}{(mp)^{1/4}}\right).
\]

9
For the ABC matrix, in which each column of $A$ is close to $d$-regular, we will use the following weaker anti-concentration lemma, which we prove in Section \[8\]

**Lemma 5** (Anti-concentration for Sparse Regular Vectors). Let $v \in \mathbb{R}^n$ be an arbitrary vector whose most common entry is $a$. Then for any $d \leq \sqrt{n}$, if $z \in \{0,1\}^n$ is sampled uniformly from the set of vectors with exactly $d$ 1s, we have:

$$\Pr[v^T z = c] \leq 1/2 + \frac{d^2}{n},$$

for all $c \in \mathbb{R}\setminus\{da\}$.

Our general strategy for the rectangular BGC and ABC matrices is as follows. Fix a set $S \subset [n]$ of size $k$. We consider the random process $A_S^1, A_S^2, \ldots, A_S^m = A_S$, where we add columns of $A_S$ one at a time. (Recall that $A_S^i$ is the submatrix of $A$ given by restricting to the rows and $S$ and the first $i$ columns.) At each step $i$, we keep track of the left kernel of $A^i$. Since our goal is to show that the left kernel of $A_S$ contains no vectors with support size $k$, we leverage the anti-concentration results above as follows: if the kernel of $A_S^i$ contains a vector $v$ with support $k$, then it is likely that the next column added, $(A_S)_{i+1}$, will not be orthogonal to $v$. After adding enough columns, we show that with high probability, we “knock out” all candidate kernel vectors with large support.

While this approach is relatively straightforward for the BGC matrix where the columns are independent, we must be more careful for the ABC since the columns are not independent. In this case, for each column added, we consider the pairing of the $d$ half-edges of the corresponding column-node, and we show that for at least half of the columns, these pairings are “sufficiently random”.

We use a similar approach to rule out some large dependencies $(n/d^{\epsilon} < k < \Theta(n))$ for the symmetric Bernoulli matrices. However, this approach breaks down for when $k$ becomes close to $n$ since the matrix is square. For instance, in the extreme case when $k = n$, we would need to “knock-out” a kernel vector at every single step of adding columns to $A$. This certainly won’t occur with high probability.

Our strategy for ruling out dependencies on the order of $\Theta(n)$ rows is inspired by a combination of the approaches in \[12\] and \[8\]. Using Markov’s law, we show that if there exists a kernel vector of $A$ with large support, then $A$ must contain many columns $A_i$ which are in the span of the remaining columns $\{A_j\}_{j \neq i}$. We bound this probability that $A_i \in \text{Span}(\{A_j\}_{j \neq i})$ by conditioning on $A^{(i)}$, the matrix formed by removing the $i$th row and column of $A$, and leveraging the randomness of $A_i$.

We consider two main cases: one in which $A^{(i)}$ has a kernel vector with large support, and one in which it doesn’t. If $A^{(i)}$ has a kernel vector $v$ with large support (on the order of $\Theta(n)$), we use Lemma 3 to show that with probability $1 - O(1/\sqrt{d})$, $A_i$ is orthogonal to $v$, and hence $A_i$ is not in the span of the remaining columns $\{A_j\}_{j \neq i}$. If $A^{(i)}$ has no kernel vectors with large support, we are able to construct a dense “pseudoinverse” $B$ for $A^{(i)}$, for which $A_i^T BA_i \neq 0$ implies that $A_i$ is not in the span of $\{A_j\}_{j \neq i}$. Lemma 4 guarantees this occurs with probability $1 - O(1/\sqrt{d})$.

Notice that unlike the results for the BGC and ABC which hold with probability that decays at least polynomially in $n$, our results for square matrices hold with probability that decays polynomially in $d$. For this reason, our results only hold with high probability if $d$ goes to infinity with $n$.

### 4.2.3 Medium Case

We sometimes need an extra medium case to rule out dependencies of $k$ rows where $n/d^{\epsilon} < k < n/d^{\epsilon}$. This is easily accomplished using Observation 1 by showing the existence of a column with a single 1.

## 5 Useful Lemmas

We gather here a few definitions and lemmas that we use in our proofs for multiple of the random matrix ensembles we study. Their proofs are technical, so we defer them to the appendix.

The first lemma allow us to show that the linear dependencies we encounter with constant probability have the graph-structures depicted in Figure 1.

**Definition 5.** Define $S_{L,k}$ to be union over all integers $m$ of the set of matrices $B \in \{0,1\}^{k \times m}$ such that $B$ forms a minimal dependency and $B$ has exactly $L$ entries which are 1.
Definition 6. Define \( S_{L,k}' \subset S_{L,k} \) to be the subset of matrices in \( S_{L,k} \) which have exactly \( k \) non-zero columns.

The following lemma relates these sets to the sets \( T_k, T_k^+ \) and \( T_k^C \) introduced earlier in Definitions 2, 3, and 4.

Lemma 6 (Classification of Dependencies). We have the following three equivalences:

1. \( S_{2k-2,k} = T_k \).
2. \( S_{2k-1,k} = T_k^+ \).
3. \( S_{2k,k}' = T_k^C \).

We will use the next lemma to bound the correlation between the existence of linear dependencies in intersecting submatrices \( A_S \) and \( A_T \).

Lemma 7. Suppose we have two sets \( S \) and \( T \) with \( S \cap T \neq \emptyset \) where \( A_S \in M_{|S|} \) and \( A_T \in M_{|T|} \). Let \( \ell \) be the number of non-zero entries in \( A_{S \cup T} \). Then there are at least \( \max(|S \cup T| - 1, \frac{\ell}{2}) \) non-zero entries in \( A_{S \cup T} \) that are not the first (top) non-zero entry in their column.

Lemmas 6 and 7 are proved in Appendix A.

Several of our proofs use the following tail bound on Binomial distributions with a small parameter.

Lemma 8 (Tail Bound on Binomial). If \( t \geq 2np \), then

\[
\Pr[Bin(n,p) \geq t] \leq 2 \left( \frac{enp}{t} \right)^t.
\]  

(13)

Several of our proofs in the small case will use the following black-box calculation to bound the probability of encountering minimal dependencies.

Lemma 9 (Small Case Binomial Calculation). For constants \( \gamma, d > 0 \), there exists a constant \( c_9 \) such that for any \( j \in \{k-1, k, k+1\} \) and \( \gamma \geq 1/2 \), we have

\[
\sum_{\ell \geq 1} \Pr \left[ Bin \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \max \left( j, \frac{\ell}{2} \right) \right] \Pr[Bin(\gamma nk, d/n) = \ell] \leq e^{-\gamma dk + \log(d) \left( \frac{k}{n} \right)^{\gamma}}.
\]  

(14)

Similarly, several of our large cases will use the following black-box calculation.

Lemma 10 (Large Case Binomial Calculation). There exists constants \( C, d_0, \) and \( c_{10} \) such that for all \( d \geq d_0 \), for any positive integer \( \frac{dn}{d} \leq k \leq n/C \), we have

\[
\sum_{k=\frac{dn}{d}}^{n/C} \binom{n}{k} \Pr \left[ Bin \left( \frac{n-k-1}{\sqrt{kd/n}} \right) \leq k \right] \leq e^{-\gamma dk}
\]  

(15)

Lemmas 8, 9 and 10 are proved in Appendix C.

6 Bernoulli Gradient Codes

In this section, we prove Theorem 1 on the characterization of linear dependencies in matrices \( A \sim BGC(n, \gamma, d) \). As discussed in the overview, we will divide the proof of this result into small, medium, and large cases.
6.1 Small Case

The goal of this section will be to prove the following lemma. Recall that $\mathcal{M}_k$ is the set of matrices with $k$ rows that are minimal dependencies.

**Lemma 11** (BGC Small Case). Let $A \sim \text{BGC}(n, \gamma, d)$. Let $S \subset [n]$ be a set of size $k$ for $k \in [1, \frac{n}{8e^{\gamma d}/d}]$. There exist universal constants $c, d_0$ such that if $d > d_0$, then:

1. $\Pr[A_S \in \mathcal{M}_k \setminus (\mathcal{T}_k \cup \mathcal{T}_k^+ \cup \mathcal{T}_k^C)] = \left( e^{-\gamma d + \log(d)} \right)^k \left( \frac{k}{n} \right)^{k+1}$.
2. $\Pr[A_S \in \mathcal{T}_k] \leq \left( e^{-\gamma d + \log(d)} \right)^k \left( \frac{k}{n} \right)^{k-1}$.
3. $\Pr[A_S \in \mathcal{T}_k^+] \leq \left( \left( e^{-\gamma d + \log(d)} \right) \left( \frac{k}{n} \right) \right)^k$.
4. $\Pr[A_S \in \mathcal{T}_k^C] \leq \left( \left( e^{-\gamma d + \log(d)} \right) \left( \frac{k}{n} \right) \right)^k$.

Further if $S, T \subset [n]$ with $S \cap T \neq \emptyset$ and $j := |S \cup T| \leq \frac{n}{8e^{\gamma d}/d}$, then

$$\Pr[A_S \in \mathcal{M}_{|S|} \cap A_T \in \mathcal{M}_{|T|}] \leq \left( e^{-\gamma d + \log(d)} \right) \left( \frac{j}{n} \right)^{j-1}.$$  

**Proof of Lemma** Let $A \sim \text{BGC}(n, \gamma, d)$. For a row subset $S$ where $|S| = k$, let $E_S$ denote the event that the submatrix $A_S$ induced by the row subset $S$ does not contain a column with exactly one 1. Let $L_S \sim \text{Bin}(\gamma kn, d/n)$ denote the number of 1s in $A_S$. By Observation 1 we know the following:

$$\Pr[A_S \in \mathcal{S}_{\ell, k}] \leq \Pr[E_S|L_S = \ell] \cdot \Pr[L_S = \ell].$$

By Observation 2 we already know the expression on the left hand-side is 0 when $L_S < 2k - 2$. Thus, we only need to address the case where $L_S \geq 2k - 2$.

Recall that each entry in our matrix is an independent Bernoulli random variable. Thus, after conditioning on the event that $L_S = \ell$ for some $\ell \in [2k - 2, \gamma kn]$, the $\ell$ ones in $A_S$ are distributed uniformly at random throughout the matrix $A_S$. We can arbitrarily enumerate the 1s from 1 to $\ell$ and consider the random process that places each 1 into $A_S$ sequentially starting from an $k \times \gamma n$ all zeros matrix.

On each step of this process, we can query whether a column with exactly one 1 has been created. If such a column is created, we will call this step good. All other steps are bad. As we add exactly $\ell$ ones into the matrix, and there are $\gamma n$ columns, we conclude that the probability of each step being a good event is at least $1 - \frac{\ell}{\gamma n}$. As each bad step can remove at most one column with exactly one 1 in it, it is clear that if more than half of the steps are good events, then our resulting submatrix $A_S$ must contain a column with exactly one 1.

Let $\{X_i\}_{i=1}^\ell$ be the random process which counts the number of bad events that have occurred after $i$ ones have been added to the matrix.

Thus we have:

$$\Pr[E_S|L_S = \ell] \leq \Pr[X_\ell \geq \left\lceil \frac{\ell}{2} \right\rceil].$$

Since the probability of $X_i$ increases at each step is at most $\frac{\ell}{\gamma n} < \frac{\ell + k}{\gamma n}$, we can define $Y(\ell) \sim \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right)$, such that

$$\Pr[E_S|L_S = \ell] \leq \Pr\left[Y(\ell) \geq \left\lceil \frac{\ell}{2} \right\rceil \right].$$

This implies that

$$\Pr[A_S \in \mathcal{S}_{\ell, k}] \leq \Pr[Y(\ell) \geq \left\lceil \frac{\ell}{2} \right\rceil] \cdot \Pr[L_S = \ell].$$

We now employ Lemma 5 which we restate here for convenience.
Lemma 9 (Small Case Binomial Calculation). For constants $\gamma, d > 0$, for $k \leq \frac{n}{\log(n)}$, there exists a constant $c$ such that for any $j \in \{k-1, k, k+1\}$ and $\gamma \geq 1/2$, we have

$$\sum_{\ell \geq 1} \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \max \left( \frac{j}{2}, \ell \right) \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell] \leq e^{-\gamma dk + [\log(\gamma d)]} \left( \frac{k}{n} \right)^{k/2}.$$

Using this Lemma 9, we conclude the following results.

For the second statement in the lemma, we have

$$\Pr[A_S \in S_{2k-1,k}] \leq \Pr \left[ Y^{(\ell)} \geq k - 1 \right] \cdot \Pr[L_S = \ell]$$
$$\leq \sum_{\ell \geq 1} \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \max \left( k - 1, \frac{\ell}{2} \right) \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell]$$
$$\leq e^{-\gamma dk + [\log(\gamma d)]} \left( \frac{k}{n} \right)^{k-1}.$$

Similarly for the third and fourth statements, we have

$$\Pr[A_S \in S_{2k-1,k} \cup \tilde{S}_{2k,k}] \leq \Pr \left[ Y^{(\ell)} \geq k \right] \cdot \Pr[L_S = \ell]$$
$$\leq \sum_{\ell \geq 1} \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \max \left( k, \frac{\ell}{2} \right) \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell]$$
$$\leq e^{-\gamma dk + [\log(\gamma d)]} \left( \frac{k}{n} \right)^{k}.$$

For the first statement in the lemma, observe that by Lemma 6, we have

$$\mathcal{M}_k \setminus (T_k \cup \tilde{T}_k^+ \cup \tilde{T}_k^-) = \left( \bigcup_{\ell \geq 2k+1} S_{\ell,k} \right) \cup \left( S_{2k,k} \setminus S'_{2k,k} \right).$$

Now

$$\sum_{\ell = 2k+1}^{\gamma nk} \Pr[A_S \in S_{\ell,k}] \leq \Pr \left[ Y^{(\ell)} \geq \frac{\ell}{2} \right] \cdot \Pr[L_S = \ell]$$
$$\leq \sum_{\ell \geq 1} \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \max \left( k + 1, \frac{\ell}{2} \right) \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell]$$
$$\leq e^{-\gamma dk + [\log(\gamma d)]} \left( \frac{k}{n} \right)^{k+1}.$$

Finally, we note that if $A_S \in S_{2k,k} \setminus S'_{2k,k}$, then there must be at least one column with three ones. Hence, the number of bad events cannot be exactly $k$ as an equal number of good and bad events means that each column has of $A_S$ has exactly 2 ones in it. Thus, we find:

$$\Pr[A_S \in S_{2k,k} \setminus S'_{2k,k}] \leq \Pr \left[ Y^{(\ell)} \geq k + 1 \right] \cdot \Pr[L_S = \ell]$$
$$\leq \sum_{\ell \geq 1} \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \max \left( k + 1, \frac{\ell}{2} \right) \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell]$$
$$\leq e^{-\gamma dk + [\log(\gamma d)]} \left( \frac{k}{n} \right)^{k+1}.$$
Combining these two equations yields the first statement in the lemma.

For the final setting, where we have two sets $S$ with $S \cap T \neq \emptyset$, by Lemma 7, it must be the case that if $A_S$ and $A_T$ are minimal dependencies, then $A_{S \cup T}$ contains at least $\max\left(\frac{\ell}{2}, |S \cup T| - 1\right)$ ones that are not the first in their column, where $\ell$ is the total number of ones in $A_{S \cup T}$. Plugging in Lemma 9 again proves that

$$\Pr[A_S \in \mathcal{M}_{|S|}, A_T \in \mathcal{M}_{|T|}] \leq \left(\frac{|S \cup T|}{n}\right)^{|S \cup T| - 1} e^{-\gamma d |S \cup T| + \gamma |S \cup T| \log(\gamma d)}.$$

Finally, the lemma follows by using the classification lemma, Lemma 6, to replace $S_{2k-2,k}$ with $T_k$, $S_{2k-1,k}$ with $T^+_k$, and $S'_{2k,k}$ with $T^C_k$.

### 6.2 Medium Case

To fill the small gap between where our small case ends and our large case begins, our proof requires the following lemma which shows that, with high probability, medium-sized minimal dependencies do not occur in our matrix $A$. Unlike in the small case, tree-like dependency structures no longer occur with significant probability. Thus, it is sufficient to explicitly calculate the probability that $A_S$ has no columns with exactly 1 one without conditioning on the number of ones in $A_S$.

However, one should observe that this strategy will begin to fail as the expected number of 1s in each column begins to exceed 1. Thus, a more careful bound on the existence of a minimal dependency will be needed when $k$ is greater than approximately $\frac{n}{d}$. This case is covered in the next subsection.

**Lemma 12 (BGC Medium Case).** Let $A \sim \text{BGC}(n, \gamma, d)$. For any constants $\alpha, \beta \in \mathbb{R}^+$, there exists $d(\alpha, \beta)$, such that for any $d > d(\alpha, \beta)$,

$$\Pr\left[\exists x : A^T x = 0, |\text{supp}(x)| \in [\alpha n/d^2, \beta n/d]\right] = o(1).$$

**Proof.** We will take a union bound over all possible sets $S$ of size $k \in [\alpha n/d^2, \beta n/d]$ of the probability that $A_S$ is a minimal dependency.

For a row subset $S$, let $E_S$ be the event that the matrix $A_S$ has no columns with exactly one 1. Then by Observation 1

$$\Pr\left[\exists x : x^T A = 0, |\text{supp}(x)| \in [\alpha n/d^2, \beta n/d]\right] \leq \sum_{k=\alpha n/d^2}^{\beta n/d} \binom{n}{k} \Pr[E_{[k]}].$$

For sufficiently large $n$, it follows:

$$\Pr[E_{[k]}] = \left(1 - k \left(\frac{d}{n}\right) \left(1 - \frac{d}{n}\right)^{k-1}\right)^{\gamma n} = \left(1 - k \left(\frac{d}{n}\right) e^{-2dk/n}\right)^{\gamma n} \leq e^{-\gamma dk\left(e^{-2dk/n}\right)}.$$
So we find:

\[
\Pr[\exists x : A^T x = 0, |\text{supp}(x)| \in [\alpha n/d^2, \beta n/d]] \leq \sum_{k=\alpha n/d^2}^{\beta n/d} \binom{n}{k} \Pr[E_{[k]}]
\]

\[
\leq \sum_{k=\alpha n/d^2}^{\beta n/d} \left( \frac{\epsilon n}{k} \right)^k e^{-\gamma d \left( e^{-2d/k} \right)}
\]

\[
\leq \sum_{k=\alpha n/d^2}^{\beta n/d} \left( e^{-\gamma d \left( e^{-2d/k} \right) + \ln(\epsilon n/k)} \right)^k
\]

\[
\leq n \left( e^{-\gamma d \left( e^{-2d/n} \right) + \ln(\epsilon d^2/\alpha)} \right)^k
\]

\[
= o(1).
\]

for sufficiently large \(d\) depending only on \(\alpha\) and \(\beta\).

\[\square\]

### 6.3 Large Case

The goal of this section is to prove the following lemma:

**Lemma 13 (BGC Large Case).** Let \(A \sim \text{BGC}(n, \gamma, d)\). For any \(\gamma > 1\), there exists \(d_0\) such that for all \(d > d_0\) we have,

\[
\Pr[\exists x : A^T x = 0, |\text{supp}(x)| \in [9n/d, n]] = o(1).
\]  

(18)

Our main tool in proving this Lemma is the following Littlewood-Offord Theorem due to Costello and Vu [10]. However, their paper only states the result up to an implied constant. As our theorems rely on this implied constant, we reprove this result in Appendix G using the same strategy as Costello and Vu, but with an explicit constant and sharper lower order terms.

**Lemma 3 (c.f. [10] Lemma 8.2).** Let \(v \in \mathbb{R}^n\) be a deterministic vector with support at least \(m\). Let \(z \in \mathbb{R}^n\) be the random vector with i.i.d. Bernoulli entries with parameter \(p \leq 1/2\). Then for any fixed \(c\),

\[
\Pr[v^T z = c] \leq \left( \frac{1}{\sqrt{\pi/2}} \right) \frac{1}{\sqrt{mp}} + \left( e^{(\ln(2) - 1)mp} \right).
\]

In particular, for \(mp \geq 9\), we have:

\[
\Pr[v^T z = c] \leq \frac{1}{\sqrt{mp}}.
\]

We now prove Lemma [13]

**Proof of Lemma [13]** We union bound over all \(k \in [9n/d, n]\) and all sets \(S\) of size \(k\) of the probability that \(A_S\) is a minimal dependency.

Fix a set \(S\) of size \(k\). We will consider the random process where we generate the \(\gamma n\) independent columns \((A_S)_i\) for \(i = 1 \ldots \gamma n\) one at a time. For \(i \leq \gamma n\), let \(\mathcal{N}_i \in \mathbb{R}^k\) be the nullspace of the first \(i\) columns drawn, and let \(\mathcal{D}_i \subset \mathcal{N}_i\) be the span of the set of vectors in \(\mathcal{N}_i\) which have no zeros. Let \(R_i\) be the dimension of \(\mathcal{D}_i\).

If \(R_i > 0\), then we can choose an arbitrary vector \(v\) in \(\mathcal{D}_i\) with support \(k\), and by Lemma [3] with probability at least \(1 - \frac{1}{\sqrt{kd/n}}\), the \((i+1)\)th column drawn is not orthogonal to \(v\). In this case \(R_{i+1} = R_i - 1\). If at any point \(R_i\) becomes 0, then this means there can be no dependency involving all the rows. It follows that since \(R_0 = k\), we have

\[
\Pr[A_S \in \mathcal{M}_k] \leq \Pr[R_{\gamma n} \neq 0] \leq \Pr\left[\text{Bin} \left( \gamma n, 1 - \frac{1}{\sqrt{kd/n}} \right) < k \right].
\]  

(19)

Let \(C\) be the universal constant in Lemma [10] which we restate for the reader's convenience:
Lemma 10 (Large Case Binomial Calculation). There exists constants $C$, $d_0$, and $c_{10}$ such that for all $d \geq d_0$, for any positive integer $\frac{9n}{d} \leq k \leq n/C$, we have

$$\sum_{k=\frac{9n}{d}}^{\frac{n}{C}} \binom{n}{k} \Pr \left[ \text{Bin} \left( n - k - 1, 1 - \frac{1}{\sqrt{kd/n}} \right) \leq k \right] \leq e^{-\Omega(n)}$$

(15)

We first handle when $k \leq n/C$. Assuming $d$ is larger than the universal constant in Lemma 10, we can apply Lemma 10 to compute the following union bound:

$$\Pr \left[ \exists S : A_S \in M_{|S|} \land |S| \in [9n/d, n/C] \right] \leq \sum_{k=9n/d}^{n/C} \Pr \left[ A_k \in M_{|S|} \right] \leq \sum_{k=9n/d}^{n/C} \binom{n}{k} \Pr \left[ \text{Bin} \left( n, 1 - \frac{1}{\sqrt{kd/n}} \right) \leq k \right] \leq \sum_{k=9n/d}^{n/C} \binom{n}{k} \Pr \left[ \text{Bin} \left( n - k - 1, 1 - \frac{1}{\sqrt{kd/n}} \right) \leq k \right] \leq e^{-\Omega(n)} = o(1).$$

(20)

Now we handle the case where $k \geq n/C$. Choose $d$ large enough, dependent on $\gamma$ and $C$ to satisfy

$$\left( 1 - 2 \sqrt{\frac{d}{C}} \right) \gamma > 1.$$

This implies that for $\frac{n}{C} \leq k \leq n$, we have

$$\gamma n - k \geq \frac{2\gamma n}{\sqrt{(d/n)k}}.$$

(21)

The inequality in this form allows us to use the the Binomial Tail bound in Lemma 8, which establishes:

$$\Pr \left[ \text{Bin} \left( \gamma n, 1 - \frac{1}{\sqrt{(d/n)k}} \right) \leq k \right] = \Pr \left[ \text{Bin} \left( \gamma n, 1 - \frac{1}{\sqrt{(d/n)k}} \right) \geq \gamma n - k \right] \leq 2 \left( \frac{e^{\gamma n}}{\sqrt{(d/n)k}(\gamma n - k)} \right)^{\gamma n - k} \leq 2 \left( \frac{e^{\gamma n}}{\sqrt{d/C}(\gamma n - k)} \right)^{\gamma n - k} \leq 2 \left( \frac{e^n}{\sqrt{d/C}(\gamma - 1)} \right)^{(\gamma - 1)n}. $$

By choosing $d$ large enough depending on $\gamma$ and the universal constant $C$, we may get $(\frac{e^n}{\sqrt{d/C}(\gamma - 1)})^{\gamma - 1} < 1/3$, so

$$\Pr \left[ \text{Bin} \left( \gamma n, 1 - \frac{1}{\sqrt{(d/n)k}} \right) \leq k \right] \leq (1/3)^n.$$

This overcomes the union bound over at most $2^n$ sets:

$$\Pr \left[ \exists S : A_S \in M_{|S|} \land |S| \in [n/C, n] \right] \leq 2^n (1/3)^n = o(1).$$

(22)
Thus, we may combine Equations 20 and 22 to obtain our result
\[
\Pr \left[ \exists S : A_S \in \mathcal{M}_{|S|} \land |S| \in [9n/d, n] \right] \leq o(1).
\]

6.4 Proof of Theorem 1

We are now ready to prove Theorem 1, which we restate here for the reader’s convenience.

**Theorem 1** (Characterization BGC). There exists a universal constant \( c \) such that for any \( \gamma > 1 \) and \( d \geq d_0(\gamma) \), for \( A \sim \text{BGC}(n, \gamma, d) \), with probability \( 1 - o(1) \):

1. All minimal dependencies of \( k \) rows of \( A \) are in \( T_k \cup T_k^+ \cup T_k^C \).

2. The number of rows involved in a linear dependency of \( A \) is at most \( ne^{-\gamma d + c \log(\gamma d)} \), that is

\[
\left| \bigcup_{x : x^T A = 0} \text{supp}(x) \right| \leq ne^{-\gamma d + c \log(\gamma d)}.
\]

**Proof of Theorem 1**. It follows immediately by combining Lemmas 11, 12 (with \( \alpha = \frac{1}{8e^4}, \beta = 9 \)), and Lemma 13 that with probability \( 1 - o(1) \), there are no sets \( S \) where \( A_S \) is a minimal dependency \( A_S \), but \( A_S \) is not in \( \cup_{k \leq \log(n)} T_k \cup T_k^+ \cup T_k^C \).

We bound the number of rows involved in minimal linear dependencies using the second moment method. For \( S \subset [n] \), let \( X_S \) be the indicator random variable of the event that \( A_S \) is a minimal dependency. Then with probability \( 1 - o(1) \), the number of rows involved in minimal dependencies is at most

\[
X := \sum_{S \subset [n], |S| \leq \log(n)} |S| X_S.
\]

First we consider the expectation of \( X \). Combining the probability that \( A_S \) is in \( T_k \), \( T_k^+ \), or \( T_k^C \), we have from Lemma 11 that for any \( S \) with \( |S| = k \), for the constant \( c_9 \) from that lemma,

\[
E[X_S] \leq \left( \frac{k}{n} \right)^{k-1} e^{-\gamma d k + c_9 k \log(\gamma d)}
\]

Hence

\[
E[X] = \sum_{S \subset [n], |S| \leq \log(n)} |S| E[X_S]
\]
\[
\leq \sum_{k \leq n} \binom{n}{k} \left( \frac{k}{n} \right)^{k-1} e^{-\gamma d k + c_9 k \log(\gamma d)}
\]
\[
\leq \sum_{k \leq n} ne^k e^{-\gamma d k + c_9 k \log(\gamma d)}
\]
\[
\leq \sum_{k \leq n} ne^{-\gamma d k + c_9 k \log(\gamma d)}
\]
\[
\leq ne^{-\gamma d + c_9 \log(\gamma d)}
\]

for some constant \( c_1 \). Here the final inequality is given by the sum of geometric series.

Next we bound the variance of \( X \). Let \( S, T \subset [n] \) with \( |S|, |T| \leq \log(n) \). If \( S \cap T = \emptyset \), by the independence of \( A_S \) and \( A_T \), we have

\[
E[X_S X_T] = E[X_S]E[X_T].
\]
Alternatively, if $S \cap T \neq \emptyset$, then with $R := S \cup T$, by Lemma 11 we have

$$\mathbb{E}[X_S X_T] \leq \left( \frac{|R|}{n} \right)^{|R|-1} e^{-\gamma d|R| + \mathbb{E}[R] \log(\gamma d)}.$$ 

Hence we can compute

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

$$= \sum_{S: |S| \leq \log(n)} \sum_{T: |T| \leq \log(n)} |S||T| (\mathbb{E}[X_S X_T] - \mathbb{E}[X_S] \mathbb{E}[X_T])$$

$$= \sum_{S} \sum_{T: T \cap S = \emptyset} |S||T| (\mathbb{E}[X_S X_T] - \mathbb{E}[X_S] \mathbb{E}[X_T]) + \sum_{S} \sum_{T: T \cap S \neq \emptyset} |S||T| (\mathbb{E}[X_S X_T] - \mathbb{E}[X_S] \mathbb{E}[X_T])$$

$$\leq \sum_{S} \sum_{T: T \cap S \neq \emptyset} |S||T| \mathbb{E}[X_S X_T]$$

$$\leq \sum_{R \subseteq [n]: |R| \geq 1} \sum_{S, T \subseteq R} |R|^2 \left( \frac{|R|}{n} \right)^{|R|-1} e^{-\gamma d|R| + \mathbb{E}[R] \log(\gamma d)}$$

$$\leq \sum_{R \subseteq [n]: |R| \geq 1} 2^{|R||R|^2 \left( \frac{|R|}{n} \right)^{|R|-1} e^{-\gamma d|R| + \mathbb{E}[R] \log(\gamma d)}}$$

$$\leq \sum_{k \geq 1} \binom{n}{k} \left( \frac{k}{n} \right)^{k-1} e^{-\gamma dk + (\mathbb{E}+1)k \log(\gamma d)}$$

$$\leq ne^{-\gamma d + c_2 \log(\gamma d)} \quad (25)$$

for some constant $c_2$. Using Markov’s law, it follows that

$$\Pr[X \geq \mathbb{E}[X] + t] \leq \Pr[(X - \mathbb{E}[X])^2 \geq t^2] \leq \frac{\text{Var}(X)}{t^2}. \quad (26)$$

Plugging in $t = ne^{-\gamma d}$, we obtain that

$$\Pr[X \geq ne^{-\gamma d + c_1 \log(\gamma d)} + ne^{-\gamma d}] \leq \frac{ne^{-\gamma d + c_2 \log(\gamma d)}}{(ne^{-\gamma d})^2} \leq \frac{1}{ne^{-\gamma d - c_2 \log(\gamma d)}} \quad \quad (27)$$

The theorem follows by choosing $c \geq c_1 + \log(2)$. \qed

7 Symmetric Bernoulli Matrices

In this section, we prove Theorem 2 on the characterization of minimal linear dependencies in symmetric Bernoulli matrices. As before, we break down our proof into small, medium, and large cases. The small and medium cases are similar to the last section on the BGC. However, Unlike for the BGC, however, the large case is not self contained: showing that there are no kernel vectors with large support relies on understanding the image of $Ax$ over vectors $x$ of small support (see Lemma 19 below). This will require understanding small and medium dependencies in $A$ and in the matrix $A$ with one column removed, which is why our small case and medium case lemmas contain additional statements to this affect.
7.1 Small Case

Our main goal in the small case is to prove the following lemma.

**Lemma 14 (Symmetric Small Case).** There exists a universal constant $c$ such that the following holds. Let $A \sim \text{SB}(n, d)$ for any $d \geq 1$. Let $S \subset [n]$ be any set of size $k \in [1, \frac{n}{8d^2}]$. Then:

1. $\Pr[A_S \notin M_k^T k] = e^{-dk + c_1 k \log\left(\frac{k}{n}\right) k}.$
2. $\Pr[A_S \notin T_k^k] \leq e^{-dk + c_1 k \log\left(\frac{k}{n}\right) k} - 1.$

The same result applies if $A$ has one column removed, i.e. $A = B^{n-1}$, where $B \sim \text{SB}(n, d)$.

**Proof.** We introduce some notation to prove this, pictured in Figure 2. Fix a set $S$ of $k$ rows and consider the submatrix $A_S$ induced by these rows. Let $E_{\text{Sym}}$ be the set of entries of $A_S$ whose columns are indexed by values in $S$. Let $E_{\text{SymAD}}$ be the subset of entries in $E_{\text{Sym}}$ that are above the diagonal of $A$, and hence mutually independent. Let $E_{\text{Asym}}$ be the set of entries who columns are not in $S$, and finally, let $E = E_{\text{Asym}} \cup E_{\text{SymAD}}$ be the full set of mutually independent entries that determine $A_S$. Formally:

$$
\begin{align*}
E_{\text{Sym}} &:= \{(i, j) : i, j \in S\} \\
E_{\text{SymAD}} &:= \{(i, j) : i, j \in S, i < j\} \\
E_{\text{Asym}} &:= \{(i, j) : i \in S, j \notin S\} \\
E &:= E_{\text{SymAD}} \cup E_{\text{Asym}}
\end{align*}
$$

We will couple the process of putting non-zero entries in these rows with a random walk that counts the number of times a non-zero entry is inserted in $E_{\text{Asym}}$ in a column that already contains a non-zero entry or into $E_{\text{SymAD}}$.

We condition on $L_S$, the number of non-zero entries in $E$. Note that $L_S \sim \text{Bin}(|E|, d/n)$. Conditioned on $L_S = \ell$, the process of choosing random entries in $A_S$ is equivalent to randomly choosing $\ell$ locations in $E$ for these non-zero entries without replacement. (Note that this is true even if we are considering a matrix with the last column removed, even though $E_{\text{SymAD}}$ may include some entries which are not repeated below the diagonal if the index of the last column is in $S$). Let $(X_i)_{i \in \ell}$ be the random walk that increases by 1 if the $i$th random location chosen is in $E_{\text{SymAD}}$ or if the $i$th random location is in $E_{\text{Asym}}$ and is not the first non-zero entry placed in its column. Otherwise, let $X_i = X_{i-1}$.

The following claims say that $X_\ell$ must be large for $A_S$ to be a minimal dependency not in $T_k$. 

![Figure 2: Regions of $A_S$ in a symmetric matrix.](Image)
Claim 1. If $X_{LS} < \max\left( k, \frac{k}{2} \right)$ and there are at least $2k - 1$ non-zero entries total in the rows, then there is a column (whose index is not in $S$) with a single non-zero entry.

Proof. Let $P$ be the number of non-zeros entries in $E_{Asym}$, and let $M$ be the number of non-zero entries in $E_{Sym,AD}$ such that $M + P = L_S$ and $2M + P \geq 2k - 1$. Let $Y := X_{LS} - M$ be the number of non-zero entries in $E_{Asym}$ which are not the first in their column. Now the number of columns not in $S$ which have exactly one non-zero entry is at least

$$(P - Y) - Y = P - 2X_{LS} + 2M \geq \max(2k - 1, L_S) - 2X_{LS} \geq 1.$$

The proof of the following claim is nearly identical.

Claim 2. If $X_{LS} < \max\left( k - 1, \frac{k}{2} \right)$ and there are at least $2k - 2$ non-zero entries total in the rows, then there is a column (whose index is not in $S$) with a single non-zero entry.

Recall from Observations 2 that any minimal dependency $A_S$ must have at least $2k - 2$ non-zero entries total. Further, any minimal dependency not in $T_k$ must have at least $2k - 1$ non-zero entries. Hence by Claim 1

$$\Pr[A_S \in M_k \setminus T_k] \leq \Pr[X_{LS} < \max(k, L_S/2)]$$

Further by Claim 2

$$\Pr[A_S \in T_k] \leq \Pr[X_{LS} < \max(k - 1, L_S/2)].$$

To bound the probability that $X_{LS}$ is large, we couple $X_i$ with a random walk $(Y_i)_{i \in L_S}$ which increases by 1 with probability

$$\frac{k(k + L_S)}{k}$$

and otherwise stays constant. Observe that $Y_i$ stochastically dominates $X_i$, because there are at most $k + L_S$ columns — and hence $k(k + L_S)$ locations in $E$ — in which placing a non-zero entry will increase $X_i$.

Then conditioned on $L_S = \ell$, for any $j$, we have

$$\Pr[X_{\ell} \geq j] \leq \Pr\left[ \text{Bin} \left( \ell, \frac{k + \ell}{n - k} \right) \geq j \right].$$

We now use Lemma 9 which we restate here, to sum this probability over all values of $L_S$.

Lemma 9 (Small Case Binomial Calculation). For constants $\gamma, d > 0$, for $k \leq \frac{n}{\gamma d}$, there exists a constant $c_0$ such that for any $j \in \{ k - 1, k, k + 1 \}$ and $\gamma \geq 1/2$, we have

$$\sum_{\ell \geq 1} \Pr\left[ \text{Bin} \left( \ell, \frac{k + \ell}{\gamma n} \right) \geq \max\left( j, \frac{\ell}{2} \right) \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell] \leq e^{-\gamma dk + \gamma k \log(\gamma d)} \left( \frac{k}{n} \right)^j.$$  (14)

We employ this lemma with $\gamma = \frac{|E|}{nk} \geq 1 - k/n$ and $j = \max(k, \ell/2)$, to achieve

$$\Pr[A_S \in M_k \setminus T_k] \leq e^{-dk + (\gamma + 1)k \log(d)} \left( \frac{k}{n} \right)^k.$$  (32)

Further, setting $j = \max(k - 1, L_S/2)$, we achieve

$$\Pr[A_S \in T_k] \leq e^{-dk + (\gamma + 1)k \log(d)} \left( \frac{k}{n} \right)^{k-1}.$$  (33)

Putting $q_0 = c_0 + 1$ proves the lemma.
7.2 Medium Case

Our main goal of this section is to prove the following lemma.

Lemma 15 (Symmetric Medium Case). Let $A \sim SB(n,d)$ with $d$ at least some universal constant $d_0$. Then

$$\Pr \left[ \exists x : A^T x = 0, \frac{n}{8e^d d^2} \leq \text{supp}(x) < \frac{9n}{d} \right] = o(1).$$

The same result applies if $A$ has one column removed, ie. $A = B^{n-1}$ where $B \sim SB(n,d)$.

Proof. We will take a union bound over all possible sets $S$ of size $k \in \left[ \frac{n}{8e^d d^2}, \frac{9n}{d} \right]$ of the probability that $A_S$ is a minimal dependency. By Observation 1, it suffices to show that with probability $1 - o(1)$, for all such sets $S$, there is a column in $A_S$ with exactly one 1.

Since the columns $(A_S)_i$ for $i \in [n] \setminus S \setminus \{n\}$ are mutually independent, we have

$$\Pr[A_S \in M_{|S|}] \leq \left(1 - \frac{k}{n} \left(1 - \frac{d}{n}\right)^{k-1}\right)^{n-k-1}$$

The calculation in Lemma 12 gives the result that

$$\sum_{k=\frac{n}{8e^d d^2}}^{\frac{9n}{d}} \binom{n}{k} \Pr[A_S \in M_k \text{ where } |S| = k] = o(1).$$

\qed

7.3 Large Case

The first lemma in our large case rules out with high probability minimal dependencies of $k$ rows for $\frac{9n}{d} \leq k < \frac{n}{C}$ for some constant $C$. The proof is similar to the large case for the BGC matrix.

Lemma 16 (Symmetric Large Case 1). Let $A \sim SB(n,d)$. There exist constants $d_0$ and $C$ such that for all $d > d_0$,

$$\Pr \left[ \exists x : A^T x = 0, \frac{9n}{d} \leq \text{supp}(x) < \frac{n}{C} \right] = o(1).$$

The same result applies if $A$ has one column removed, ie. $A = B^{n-1}$ where $B \sim SB(n,d)$.

Proof. We union bound over all $k \in [9n/d, n]$ and all sets $S$ of size $k$ of the probability that $A_S$ is a minimal dependency.

Fix a set $S$ of size $k$. We will consider the random process where we generate the $n - k - 1$ columns $(A_S)_i$ for $i \in [n-1] \setminus S$ one at a time. Note that these columns are all mutually independent since they do not include columns indexed by $S$. Further, they do not include the last column of $A$.

Consider the following process, where we draw these independent columns one at a time. For $i \leq n-k-1$, let $\mathcal{N}_i \in \mathbb{R}^k$ be the nullspace of the first $i$ columns drawn, and let $\mathcal{D}_i \subset \mathcal{N}_i$ be the span of the set of vectors in $\mathcal{N}_i$ which have no zeros. Let $R_i$ be the dimension of $\mathcal{D}_i$.

If $R_i > 0$, then we can choose an arbitrary vector $v$ in $\mathcal{D}_i$ with support $k$, and by Lemma 3 with probability at least $1 - \frac{1}{\sqrt{kd/n}}$, the $(i+1)$th column drawn is not orthogonal to $v$. In this case $R_{i+1} = R_i - 1$.

If at any point $R_i$ becomes 0, then this means there can be no dependency involving all the rows. It follows that since $R_0 = k$, we have

$$\Pr[R_{n-k-1} \neq 0] \leq \Pr \left[ \text{Bin} \left( n - k - 1, 1 - \frac{1}{\sqrt{kd/n}} \right) < k \right].$$

Next we take a union bound over all $k$ in the desired range and all $\binom{n}{k}$ sets $S$ of size $k$. Employing the calculation from Lemma 10 for universal constants $\frac{m}{C}$ and $d_0$, for $d \geq d_0$, we have
\[ \Pr[\exists x : Ax = 0, 0.9n/d \leq \text{supp}(x) \leq n/C] \leq e^{-\Omega(1)}. \] (38)

This concludes the proof of the lemma.

The remainder of the large case is based off of ideas from the works of Ferber, Kwan, and Sauer- mann \[12\] (see Lemma 2.1), and Costello, Tau, and Vu \[5\] (see proof of Lemma 2.8).

We will use the following two linear-algebraic lemmas, proved in Appendix \[A\].

**Lemma 17.** Let \( A \) be a matrix with columns \( A_i \) for \( i \in [n] \). Let \( H_i \) be the space spanned by the column vectors \( A_1, A_2, \cdots A_{i-1}, A_{i+1}, \cdots A_n \). Let \( S \) be the set of all \( i \) such that \( A_i \in H_i \). Then there exists some \( y \) with \( \text{supp}(y) = S \) such that \( Ay = 0 \).

**Lemma 18.** With the terminology of the previous lemma, \( e_i \in \text{Span}(A^T) \), if and only if \( A_i \notin H_i \).

The following lemma is our main tool in ruling out large dependencies and proving Theorem \[2\]. It breaks down the probability that there is a large linear dependency consisting of more than \( t \) rows into the sum of the probabilities of several other events. Two of these probabilities (lines 1 and 5 of the right hand side of Equation \[39\] in the lemma) we can show to be small via anti-concentration lemmas. Two of the probabilities (lines 2 and 4 of the right hand side of Equation \[39\] in the lemma), we can show to be small using lemmas we proved in the small case. This is unlike the previous large case lemma, where weren’t concerned with any small structures that might exist in \( A \).

**Lemma 19** (Symmetric Large Case Main). Let \( A \sim SB(n,d) \). Let \( A^{(i)} \) denote the submatrix of \( A \) given by the \( i \)th row and column removed. Let \( A_i \) denote the \( i \)th column of \( A \), and let \( A'_i \) be this vector with the \( i \)th entry removed. For any \( u, r, s, t \in [n] \) with \( r < s \),

\[
\Pr[\exists x : Ax = 0, \text{supp}(x) > t] \leq \frac{n}{t} \max_{x \in \mathbb{R}^{n-1}: \text{supp}(x) \geq s} \Pr \left[ x^T A'_n = 0 \right] + \frac{n}{t} \Pr \left[ \exists x : A^{(n)} x = 0, r < |\text{supp}(x)| < s \right] \\
+ \frac{n}{t} \frac{d}{n} + \frac{n}{t} \frac{n}{u} \Pr \left[ \exists x \neq 0 : A^{(n)} x = e_1, |\text{supp}(x)| < s \right] + \frac{n}{t} \max_{x \in T_{p,q}^{m}} \Pr \left[ A'^T_n X A'_n = 0 \right],
\]

where \( T_{p,q}^{m} \) denotes the set of matrices in \( \mathbb{R}^{m \times m} \) with some set of \( p \) columns that each have at least \( q \) non-zero entries.

**Proof.** For \( i \in [n] \), let \( H_i \) be the space spanned by the column vectors \( A_1, A_2, \cdots A_{i-1}, A_{i+1}, \cdots A_n \).

Then \( Ax = 0 \) for some \( x \) implies that for all \( i \in \text{supp}(x) \), \( A_i \in H_i \). Let \( E_i \) denote the event that \( A_i \in H_i \). Let \( X_i \) be the indicator of this event, and let \( X = \sum_i X_i \). Then by Markov's inequality and the exchangeability of the columns,

\[
\Pr[\exists x : Ax = 0, |\text{supp}(x)| \geq t] = \Pr[X \geq t] \leq \frac{E[X]}{t} = \frac{n \Pr[E_i]}{t}.
\]

We will break down the probability \( \Pr[E_i] \) into several cases, depending on the size of the support of vectors in the kernel of \( A^{(n)} \). Let \( S \subset [n-1] \) be the set of all \( i \) such that \( e_i \in \text{Span}(A^{(n)}) \), such that by Lemma \[17\] and Lemma \[18\]

\[
k := \max(\text{supp}(x) : A^{(n)} x = 0) = n - 1 - |S|.
\]

**Case 1:** \( A^{(n)} \) has a kernel vector \( x \) with large support, that is, \( k \geq s \).
Case 2: $A^{(n)}$ has a kernel vector $x$ with medium support, that is $r < k < s$.

Case 3: $A^{(n)}$ does not have any kernel vectors with large or medium support vectors in its kernel, that is, $k \leq r$.

We can expand

$$
\Pr[E_n] = \Pr[E_n|k \geq s] \Pr[k \geq s] \\
+ \Pr[E_n|r < k < s] \Pr[r < k < s] \\
+ \Pr[E_n|k \leq r] \Pr[k \leq r].
$$

(42)

For simplicity, define $a := A'_n$ to be the first $n - 1$ entries of the column $A_n$.

To evaluate the probability of the first case, we condition on $A^{(n)}$ and let $x$ be any vector of support at least $s$ in the kernel of $A^{(n)}$. Observe that $E_n$ cannot hold if $x^T a$ is non-zero. Indeed, if $x^T a \neq 0$, then let $x' = (x_1, x_2, \ldots, x_{n-1}, 0)/(x^T a)$ such that $A x' = e_n$. Then by Lemma 18 $A_n \notin H_n$ and hence $E_n$ does not occur. Since $a$ is independent from $x$, we have

$$
\Pr[E_n|k \geq s] \Pr[k \geq s] \leq \max_{x: \text{supp}(x) \geq s} \Pr[x^T a = 0].
$$

(43)

Combined with Equation 40, the contribution from this case yields the first term in the right hand side of Equation 39.

For the second case, we bound

$$
\Pr[E_n|r < k < s] \Pr[r < k < s] \leq \Pr[r < k < s] \leq \Pr\left[\exists x : A^{(n)} x = 0, r < |\text{supp}(x)| < s \right].
$$

(44)

Combined with Equation 40, the contribution from this case yields the second term in the right hand side of Equation 39.

The third case will lead to the final three terms in the right hand side of Equation 39. In this case, we will show conditions under which we can algebraically construct a vector $v$ such that $A v = e_n$. This will imply by Lemma 18 that $A_n \notin H_n$.

Recall that $S \subset [n - 1]$ is the set of all $i$ such that $e_i \in \text{Span}(A^{(n)})$. For $i \in S$, let $w_i$ be any vector such that $A^{(n)} w_i = e_i$. We next construct a sort of “pseudoinverse” matrix $B \in \mathbb{R}^{n-1 \times n-1}$ as follows: For $i \in S$, define $B_{ij}$ to be the $i$th entry of $w_i$. That is, for $i \in S$, the $i$th column of $B$ is $w_i$. Define all other entries of $B$ to be zero.

The following claim shows a condition for $E_n$ not holding.

Claim 3. If supp($a$) $\subset S$ and $a^T B a \neq 0$, then $e_n \in \text{Span}(A)$.

Proof. Let

$$
w' := B a = \sum_{i \in S} a_i w_i
$$

such that

$$A^{(n)} w' = \sum_{i \in S} a_i e_i.
$$

Hence if supp($a$) $\subset S$,

$$A^{(n)} w' = a.
$$

In this case, define $w \in \mathbb{R}^n$ to be the vector with $w'$ in the first $n - 1$ entries and $-1$ in the final entry. Then the first $n - 1$ entries of $A w$ are $0$, and the last entry is $a^T w' = a^T B a$. Evidently, if $a^T B a \neq 0$, then

$$
\frac{A w}{a^T B a} = e_n,
$$

so $e_n \in \text{Span}(A)$. \qed
By definition, in the third case, we have $|S| \geq n - 1 - r$. Hence by Claim 3

$$\Pr[\mathcal{E}_n \wedge k \leq r] \leq \Pr[\text{supp}(a) \not\subseteq S \wedge |S| \geq n - 1 - r] + \Pr[a^TBa = 0 \wedge |S| \geq n - 1 - r].$$

(45)

Notice that $S$ is a function of $A^{(n)}$ and so $a$ is independent from $S$. It is easy to check that for any set $S$ of size at least $n - 1 - r$,

$$\Pr[\text{supp}(a) \not\subseteq S] \leq 1 - \left(1 - \frac{d}{n}\right)^r \leq \frac{dr}{n}.$$

(46)

We will break up the second term in Equation (45) by conditioning on whether the support of $B$ has many entries or not, and using the independence of $a$ from $B$:

$$\Pr[a^TBa = 0] \leq \Pr[B \not\subseteq T_{n-1-r-u,s}^{n-1} \wedge |S| \geq n - 1 - r] + \max_{x \in T_{n-1-r-u,s}^{n-1}} \Pr[a^TXa = 0]$$

(47)

To further bound the first probability on the right hand side, observe that if $|S| \geq n - 1 - r$ and $B \not\subseteq T_{n-1-r-u,s}^{n-1}$, there must exist at least $u$ different $i \in S$ such that $\text{supp}(w_i) \subseteq s$.

So

$$\Pr[B \not\subseteq T_{n-1-r-u,s}^{n-1} \wedge |S| \geq n - 1 - r] \leq \Pr\left[\{|i : \exists x \neq 0 : A^{(n)}x = e_i, |\text{supp}(x)| < s\} \geq u\right]$$

$$\leq \frac{n}{u} \Pr\left[\exists x \neq 0 : A^{(n)}x = e_1, |\text{supp}(x)| < s\right],$$

(48)

where the last inequality follows by Markov’s inequality. Combining this with Equations (47), (46), and (45) yields

$$\Pr[\mathcal{E}_n \wedge k \leq r] \leq \frac{dr}{n} + \max_{x \in T_{n-1-r-u,s}^{n-1}} \Pr[a^TXa = 0] + \frac{n}{u} \Pr\left[\exists x \neq 0 : A^{(n)}x = e_1, |\text{supp}(x)| < s\right].$$

(49)

Plugging this and Equations (43) and (44) into Equation (42) and finally Equation (40) yields the lemma. □

We instantiate Lemma 19 with $t = \frac{n}{C}$, $s = \frac{n}{C}$, $r = \frac{n}{d \log(d)}$, $u = \frac{n}{2}$, where $C$ is the constant from Lemma 16 to obtain the following lemma.

**Lemma 20.** Let $A \sim SB(n, d)$ for $d = o(1)$. With $C$ equal to the constant from Lemma 16,

$$\Pr[\exists x : Ax = 0, \text{supp}(x) > n/C] \leq C \Pr[\exists x : A^{(n)}x = 0, \frac{n}{d \log(d)} \leq |\text{supp}(x)| \leq n/C]$$

$$+ 2C \Pr[\exists x : A^{(n)}x = e_1, |\text{supp}(x)| \leq n/C]$$

$$+ o(1).$$

(50)

**Proof.** This follows immediately from plugging in these values of $t, s, r$ and $u$ into Lemma 19 and applying the anti-concentration results in Lemmas 3 and 4 to the first and last terms. Indeed, Lemmas 3 shows that

$$n \max_{x \in R^{n-1}, \text{supp}(x) \geq s} \Pr[x^TA'_{n} = 0] \leq \frac{n}{t} \frac{1}{\sqrt{sd/n}} = o(1).$$

(51)

Lemma 4 shows that

$$n \max_{x \in T_{n-1-r-u,s}^{n-1}} \Pr[A'_{n}^T X A'_{n} = 0] \leq O\left(\frac{1}{\sqrt{\min(s, n-1-r-u)d/n}}\right) = o(1).$$

(52)

The third term is at most $O(1/\log(d))$ which is also $o(1)$. □
7.4 Proof of Theorem 2

We are now ready to put the results of the small, medium, and large cases together to prove Theorem 2. For the convenience of the reader, we restate this theorem:

**Theorem 2** (Characterization Symmetric Bernoulli). Let $A \sim SB(n, d)$, where $d = \omega(1)$. Then with probability $1 - o(1)$,

1. All minimal dependencies of $k$ rows of $A$ are in $T_k$.
2. The number of rows involved in a linear dependency of $A$ is at most $n e^{-d + o(d)}$, that is

$$
\left| \bigcup_{x : x^TA = 0} \text{supp}(x) \right| \leq n e^{-d + o(d)}.
$$

We will need the following two lemmas to show that the first two terms in the right hand size of Lemma 20 are $o(1)$.

**Lemma 21.** Let $A \sim SB(n, d)$ with $d = \omega(1)$ and let $C$ be as in Lemma 16. Then

$$
\Pr \left[ \exists x : Ax = 0, \frac{n}{d \log(d)} < |\text{supp}(x)| < \frac{n}{C} \right] = o(1).
$$

**Proof.** This is immediate from the medium and first large case Lemmas 15 and 16 and the fact that $d = \omega(1)$, which implies that $\frac{n}{d \log(d)} \geq \frac{n}{8e^4}$. \hfill \square

**Lemma 22.** Let $A \sim SB(n, d)$ with $d = \omega(1)$ and let $C$ be as in Lemma 16. Then

$$
\Pr \left[ \exists x : Ax = e_1, |\text{supp}(x)| < \frac{n}{C} \right] = o(1)
$$

We reduce Lemma 22 to Lemma 23, which are better suited to prove with our medium and small case lemmas.

**Lemma 23.** Let $A \sim SB(n, d)$ with $d = \omega(1)$ and let $C$ be as in Lemma 16. Let $K = \{ x : Ax = 0, |\text{supp}(x)| \leq \frac{n}{C} \}$. Then

$$
\Pr \left[ \left| \bigcup_{x \in K} \text{supp}(x) \right| \geq \frac{n}{d \log(d)} \right] = o(1).
$$

**Proof of Lemma 23.** First we consider vectors $x$ with $1 \notin \text{supp}(x)$. Applying the first part of Lemma 23 to $A^{(1)}$, with probability $1 - o(1)$, there exists some set $T \subset [n] \setminus \{1\}$ with $|T| \leq \frac{n}{d \log(d)}$ such that $\text{supp}(x) \subset T$ for all $x$ satisfying that $A^{(1)}x = 0$ and $\text{supp}(x) \leq \frac{n}{C}$. With probability $1 - o(1)$, $\text{supp}(A_1) \cap T = \emptyset$, so for all vectors $x$ with support in $[n] \setminus \{1\}$ and of size less than $\frac{n}{C}$, we do not have $Ax = e_1$.

Next we consider vectors $x$ with $1 \in \text{supp}(x)$. If such an $x$ exists, that is, $Ax = e_1$ and $1 \in \text{supp}(x)$, then it must be the case that $A'x = 0$, where $A'$ is $A$ with the first row removed. By the second part of Lemma 23, the probability that such an $x$ exists is $o(1)$. \hfill \square
Proof of Lemma 23. Let \( \mathcal{L}_2 \) be the event that

\[
\exists x : Ax = 0, \quad \frac{n}{d \log(d)} < |\text{supp}(x)| < \frac{n}{C}.
\]

Recall that by Lemma 21, this event occurs with probability \( o(1) \). To prove the first part, we will show that conditioned on \( \mathcal{L}_2 \) not occurring, we have \( |\bigcup_{x \in \mathcal{K}} \text{supp}(x)| < \frac{n}{d \log(d)} \).

Index the set \( \mathcal{K} \) as follows: \( \mathcal{K} = \{x^{(1)}, \ldots, x^{(|\mathcal{K}|)}\} \). For \( i = 1 \) to \( |\mathcal{K}| \), let \( y^{(i)} = \sum_{j=1}^{i} c_j x^{(i)} \), where \( c_i \) is chosen uniformly from the interval \([0, 1]\). It follows that with probability 1, for all \( i \leq |\mathcal{K}| \),

\[
\text{supp}(y^{(i)}) = \bigcup_{j \leq i} \text{supp}(x^{(i)}).
\]

If \( \mathcal{L}_2 \) does not occur, then \( |\text{supp}(x^{(i)})| \leq \frac{n}{d \log(d)} \) for all \( i \), so we have that \( |\text{supp}(y^{(i+1)})| \leq |\text{supp}(y^{(i)})| + \frac{n}{d \log(d)} \). It follows that if \( |\text{supp}(y^{(|\mathcal{K}|)})| \geq \frac{n}{d \log(d)} \), then there exists some \( i \leq |\mathcal{K}| \) such that \( |\text{supp}(y^{(i)})| \in \left[\frac{n}{d \log(d)}, \frac{2n}{d \log(d)}\right] \). However, this would imply that \( \mathcal{L}_2 \) holds, which is a contradiction.

For the second part, by Lemmas 14, 15, and 16 with probability \( 1 - o(1) \), the only possible minimal dependencies in the columns of \( A \) must have a total of between \( 2(k - 1) \) and \( 2k \) non-zero entries. (Note that we are applying the results of these lemmas to columns instead of rows.) There are \( \binom{n}{k-1} \) possible sets of \( k \) columns which include the first column. By Lemma 14 the probability of a dependency occurring in one of those sets of columns is at most

\[
\left(\frac{k}{n}\right)^{k-1} e^{-kd + \Theta(k \log(d))}.
\]

Hence by a union bound, with probability at most

\[
\binom{n}{k-1} \left(\frac{k}{n}\right)^{k-1} e^{-kd + \Theta(k \log(d))} = o(1),
\]

there are no small dependencies involving the first column. \( \square \)

We restate Theorem 2 for the reader’s convenience before proving it.

**Theorem 2** (Characterization Symmetric Bernoulli). Let \( A \sim \text{SB}(n, d) \), where \( d = \omega(1) \). Then with probability \( 1 - o(1) \),

1. All minimal dependencies of \( k \) rows of \( A \) are in \( \mathcal{T}_k \).
2. The number of rows involved in a linear dependency of \( A \) is at most \( ne^{-d + o(d)} \), that is

\[
\left| \bigcup_{x:A^T x = 0} \text{supp}(x) \right| \leq ne^{-d + o(d)}.
\]

**Proof of Theorem 2.** Combining Lemma 20 with the Lemmas 21 and 22 presented above, with probability \( 1 - o(1) \), for any \( k \geq \frac{n}{d \log(d)} \), there are no minimal dependencies in \( A \) of \( k \) rows. Further applying Lemma 14 we see that with probability \( 1 - o(1) \), all minimal dependencies of size \( k \) must be in \( \mathcal{T}_k \). This proves the first part of the theorem.

It remains to bound with high probability the number of rows involved in minimal dependencies in \( \mathcal{T}_k \).

For \( S \subset [n] \), let \( X_S \) be the indicator variable that \( A_S \in \mathcal{T}_{|S|} \). Then the total number of rows in minimal dependencies is at most

\[
X = \sum_{S \subset [n]} |S| X_S.
\]  \( (57) \)

By Lemma 14, for the universal constant \( c \), we have
\[
E[X] = \sum_{k \leq \frac{n}{e + d + c \log(d)}} \binom{n}{k} ke^{-dk + c \log(d)} \left( \frac{k}{n} \right)^{k-1}
\]
\[
\leq \sum_{k \leq \frac{n}{e + d + c \log(d)}} \binom{en}{k} e^{-dk + c k \log(d)} \left( \frac{k}{n} \right)^{k-1}
\]
\[
= \sum_{k \leq \frac{n}{e + d + c \log(d)}} ne^k e^{-dk + c k \log(d)} + k
\]
\[
\leq n e^{-d + 1 + c \log(d)}
\]
\[
\leq ne^{-d + c \log(d)}
\]

for some constant \( c \).

By Markov’s law, we have

\[
\Pr \left[ X \geq ne^{-d + (c+1) \log(d)} \right] \leq \frac{E[X]}{ne^{-d + (c+1) \log(d)}} \leq \frac{1}{d}.
\] (59)

Since \( d = \omega(1) \), this proves the theorem.

\[ \square \]

8 Augmented Biregular Codes

In this section, we prove Theorem 3 on the characterization of minimal linear dependencies in matrices from \( \text{ABC}_p \). Our proof is broken down into a small and large case.

The ABC distribution via the Configuration Model

Recall that we use the following process to generate a random matrix \( A_0 \in \{0,1\}^{n \times \gamma n} \) from the distribution \( \text{ABC}(n, \gamma, d) \):

Create \( n \) row-nodes and \( \gamma n \) column-nodes and associate to each row-node \( \gamma d \) half-edges and to each column node \( d \) half-edges. Create a multi-graph \( G \) by choosing a uniformly random pairing of the \( \gamma dn \) half-edges from the row nodes to the \( \gamma dn \) half-edges from the column nodes. Given this bipartite graph, we will take \( A_0 \in \{0,1\}^{n \times \gamma n} \) to be the matrix where \( (A_0)_{ij} = 1 \) iff there is at least one edge from node \( i \) to \( j \).

Because we will study the resulting matrix \( A_0 \) via the half-edge pairing process, in our proofs we will sometimes consider the permutation \( \rho \in \mathcal{S}_{\gamma dn} \) which defines the random pairing of half-edges. From \( \rho \), we can construct an additional matrix \( H_0 = H_0(\rho) \), which we call the row half-edge occupancy matrix. Let \( P \in \{0,1\}^{\gamma dn \times \gamma dn} \) be the permutation matrix of \( \rho \). From \( P \), construct \( H_0 \in \{0,1\}^{\gamma dn \times \gamma n} \) by summing adjacent columns of \( P \) corresponding to half-edges of the same column nodes. Symbolically,

\[
(H_0)_{ij} = \sum_{k=(j-1)d+1}^{jd} P_{ik}.
\]

Note that \( H_0 \) has \( \gamma n \) columns; column \( i \) represents column-node \( i \) in the configuration model. \( H_0 \) has \( \gamma dn \) rows: rows \( (j-1)\gamma d + 1 \) through \( j\gamma d \) represent the \( \gamma d \) half edges of row-node \( j \), for \( 1 \leq j \leq n \). Each column contains exactly \( d \) 1’s, corresponding to that column-node’s half edges.

We may write \( A_0 = A_0(\rho) \) in terms of the random matrix \( H_0 \). Because \( (A_0)_{ij} = 1 \) iff there is at least one edge from vertex \( i \) to \( j \), we have

\[
(A_0)_{ij} = \mathbb{I} (\exists k, d(i-1) < k \leq di, H_{kj} = 1).
\]
When generating a matrix $A \sim \text{ABC}_p(n, \gamma, d)$, we can without loss of generality generate $A_0 \sim \text{ABC}(n, \gamma, d)$ and take $A \coloneqq A_0^{\gamma n(1-p)}$ to be the matrix containing the first $\gamma n(1-p)$ columns of $A_0$. We define $H \coloneqq H_0^{\gamma n(1-p)}$ to be the first $\gamma n(1-p)$ columns of the row half-edge occupancy matrix associated with $A_0$.

For a set $S \subset [n]$ with $|S| = k$, let $H(S) \in \mathbb{R}^{\gamma d k \times \gamma n(1-p)}$ be the matrix $H$ restricted to the $\gamma d k$ rows corresponding to half-edges of the row-nodes in $S$.

### 8.1 Small Case

The goal of this section will be to prove the following lemma:

**Lemma 24.** Let $A \sim \text{ABC}_p(n, \gamma, d)$ for $\gamma \geq 1$. Let $S \subset [n]$ be any set of size $k \in [1, \frac{n}{\log \gamma d}]$. There exists universal constants $c_3, d_0$ such that if $d > d_0$, then:

1. $\Pr[A_S \in M_k \setminus (T_k \cup T_k^+ \cup T_k^C)] = O\left(e^{-k} \left(\frac{k}{n}\right)^{k+1/2}\right)$.
2. $\Pr[A_S \in T_k] \leq (p^{\gamma d} + e\log(\gamma d))^k \left(\frac{k}{n}\right)^{k-1}$.
3. $\Pr[A_S \in T_k^+] \leq (p^{\gamma d} + e\log(\gamma d))^k \left(\frac{k}{n}\right)^k$.
4. $\Pr[A_S \in T_k^C] \leq (p^{\gamma d} + c\log(\gamma d))^k \left(\frac{k}{n}\right)^k$.

Further if $S, T \subset [n]$ with $S \cap T \neq \emptyset$. Let $|S| = k$, $|T| = j$, and $R \coloneqq S \cup T$, then

$$\Pr[A_S \in T_k \cup T_k^+ \cup T_k^C \land A_T \in T_j \cup T_j^+ \cup T_j^C] \leq \left((p^{\gamma d} + e\log(\gamma d))\left(\frac{|R|}{n}\right)\right)^{|R|^{-1}}.$$  

To prove Lemma 24, we shall condition on $L_S$, the number of 1s in $H(S)$. We observe that $L_S \sim \text{HyperGeom}(\gamma d n, \gamma d(1-p)n, \gamma dk)$.

Note that $L_S$ is greater than or equal to the number of entries which are 1 in the submatrix $A_S$, with equality if and only if there does not exist a row node in $S$ which pairs multiple half-edges to the same column-node in $[\gamma n(1-p)]$.

We are now ready to prove Lemma 24.

**Proof of Lemma 24**

Let $E_S$ denote the event that there are no columns in $H(S)$ with exactly 1 one. It follows from Observation [14] that

$$\Pr[A_S \in S_{\ell,k}] \leq \Pr[E_S|L_S = \ell] \cdot \Pr[L_S = \ell]$$

Our general strategy to bound $\Pr[E_S|L_S = \ell]$ relies on the following claim.

**Claim 4.** Let $X$ be the number of non-zero columns in $H(S)$, or equivalently, the number of 1s in $H(S)$ which are the first (top) 1 in their column. If $E_S$ occurs, then $X \leq |L_S/2|$.

**Proof.** If $X \geq |L_S/2|$, then there are strictly less than $L_S/2$ ones in $H(S)$ which are not the first 1 in their column. By the pigeonhole principle, this means there must be at least one column with a “top” 1 but no 1s below it, i.e. this column has exactly one 1.

We will bound the probability that $X$ is small by considering a random walk which counts the number of non-zero columns as half-edges are paired one at a time. We formalize this random walk as follows.

Conditioned on $L_S$, the matrix $H(S) \in \{0, 1\}^{\gamma d k \times \gamma n(1-p)}$ is distributed like a uniformly random matrix on the set of all matrices in $\{0, 1\}^{\gamma d k \times \gamma n(1-p)}$ with exactly $L_S$ ones. We construct $H(S)$ via the following random process: $M^0, M^1, \ldots, M^{L_S} = H(S)$ on matrices in $\{0, 1\}^{\gamma d k \times \gamma n(1-p)}$, in which the $L_S$ half-edges represented in $H(S)$ are paired one at a time. Formally, at each step $i$, we construct $M^i$ from $M^{i-1}$ by placing a 1 in a uniformly random location in $M^{i-1}$ without a 1. For $i \in 0, 1, \ldots, L_S$, let $X_i$ equal the number of non-zero columns in $M^i$, such that $X_{L_S} = X$.

When placing the $i$-th 1, there are at most $i-1$ non-zero columns so far. In particular, since $i \leq L_S$, there are at most $L_S$ non-zero columns throughout this process. Because $H(S)$ has $\gamma(1-p)n$ columns, we
thus have the bound $\Pr[X_i - X_{i-1} = 1] \geq 1 - \frac{L_S}{\gamma(1-p)n}$. It follows that the random variable $X$ first-order stochastically dominates the random variable $\sum_{i=1}^{L_S} \text{Bernoulli}(1 - \frac{L_S}{\gamma(1-p)n})$. So in particular, we have:

$$\Pr[X \leq \lfloor L_S/2 \rfloor] \leq \Pr \left[ \text{Bin} \left( L_S, 1 - \frac{L_S}{\gamma(1-p)n} \right) \leq \lfloor L_S/2 \rfloor \right],$$

which implies

$$\Pr[A_S \in \mathcal{S}_{\ell,k}] \leq \Pr \left[ \text{Bin} \left( \ell, 1 - \frac{\ell}{\gamma(1-p)n} \right) \leq \lfloor \ell/2 \rfloor \right] \cdot \Pr[L_S = \ell].$$

We will use the following claim proved in Appendix D:

**Claim 5.** Let $p < 1/2$. Let $K \leq \frac{3}{2}pN$. There exists constants $c_5$ and $d_0$ such that for all $\gamma > 1$ and $d > d_0$, the following two bounds hold.

For $\ell \in \{2K - 2, 2K - 1, 2K\}$, and $j \leq \lfloor \ell/2 \rfloor$, we have:

$$\Pr \left[ \text{Bin} \left( \ell, 1 - \frac{\ell}{\gamma(1-p)n} \right) \leq j \right] \cdot \Pr[\text{HyperGeom}(\gamma dN, \gamma d(1-p)N, \gamma dK) = \ell] \leq \left( p^{\gamma d - c_5 \log(\gamma d)} \right)^K \left( \frac{K}{N} \right)^{\ell-j}.$$

Further,

$$\sum_{\ell=1}^{4K} \Pr \left[ \text{Bin} \left( \ell, \frac{4K}{\gamma(1-p)n} \right) \geq K - 1 \right] \cdot \Pr[\text{HyperGeom}(\gamma dN, \gamma d(1-p)N, \gamma dK) = \ell] \leq \left( p^{\gamma d - c_5 \log(\gamma d)} \right)^K \left( \frac{K}{N} \right)^{K-1}.$$

Since $L_S \sim \text{HyperGeom}(\gamma dN, \gamma d(1-p)n, \gamma dk)$, it must be the case that

$$L_S \geq \gamma dk - p\gamma dn.$$

Hence if $L_S \in \{2k - 2, 2k - 1, 2k\}$, then necessarily

$$k \leq \frac{p\gamma dn}{\gamma d - 2} \leq \frac{3}{2}pn,$$

for $\gamma d \geq 6$, which is guaranteed if $d \geq 6$.

Thus we can apply the first bound in Claim 5 with $K = k$ and $N = n$ to study $\Pr[A_S \in \mathcal{S}_{\ell,k}] \leq \Pr[Y \leq \lfloor \ell/2 \rfloor] \cdot \Pr[L_S = \ell]$, yielding:

$$\Pr[A_S \in \mathcal{S}_{2k-2,k}] = \left( p^{\gamma d - c_5 \log(\gamma d)} \right)^k \left( \frac{k}{n} \right)^{k-1}.$$

$$\Pr[A_S \in \mathcal{S}_{2k-1,k}] = \left( p^{\gamma d - c_5 \log(\gamma d)} \right)^k \left( \frac{k}{n} \right)^{k}.$$

$$\Pr[A_S \in \mathcal{S}_{2k,k}] \leq \Pr[A_S \in \mathcal{S}_{2k-2,k}] = \left( p^{\gamma d - c_5 \log(\gamma d)} \right)^k \left( \frac{k}{n} \right)^{k}. $$

We use the equivalences of $\mathcal{S}_{2k-2,k} = \mathcal{T}_k$, $\mathcal{S}_{2k-1,k} = \mathcal{T}_k^+$, and $\mathcal{S}_{2k,k} = \mathcal{T}_k^C$ from Lemma 6 to yield statements 2, 3, and 4 in this lemma.

Next we obtain a bound on the event that $A_S \in \mathcal{M}_k \backslash (\mathcal{T}_k \cup \mathcal{T}_k^+ \cup \mathcal{T}_k^C)$. Observe that

$$A_S \in \mathcal{M}_k \backslash (\mathcal{T}_k \cup \mathcal{T}_k^+ \cup \mathcal{T}_k^C) \Rightarrow A_S \in (\mathcal{S}_{2k,k} \setminus \mathcal{S}_{2k,k}^c) \cup \bigcup_{\ell \geq 2k+1} \mathcal{S}_{\ell,k}.$$

If $A_S \in \mathcal{S}_{2k,k} \setminus \mathcal{S}_{2k,k}^c$, there must be a column with at least three 1s. So by a similar argument to Claim 4 it must be the case that $X \leq k - 1$. 

29
Hence
\[ \Pr[A_S \in S_{2k,k} \setminus S'_{2k,k}] \leq \Pr \left[ \text{Bin} \left( 2k, 1 - \frac{2k}{\gamma(1-p)n} \right) \leq k - 1 \right] \cdot \Pr[L_S = 2k] \]

Another application of the Claim gives us:
\[ \Pr[A_S \in S_{2k,k} \setminus S'_{2k,k}] \leq \left( p^{\gamma d + \log(\gamma d)} \right)^k \left( \frac{k}{n} \right)^{k+1}. \]

It follows that
\[
\Pr[A_S \in M_k \setminus (T_k \cup T^+_k \cup T^-_k)] \leq \left( p^{\gamma d + \log(\gamma d)} \right)^k \left( \frac{k}{n} \right)^{k+1} + \sum_{\ell=2k+1}^{\gamma dk} \Pr[A_S \in S_{k,k}]
\]
\[
\leq \left( p^{\gamma d + \log(\gamma d)} \right)^k \left( \frac{k}{n} \right)^{k+1} + \sum_{\ell=2k+1}^{\gamma dk} \Pr \left[ \text{Bin} \left( \ell, 1 - \frac{\ell}{\gamma(1-p)n} \right) \right] \cdot \Pr[L_S = \ell]
\]

To bound the final term, we use the following claim, which we prove in Appendix D.

**Claim 6.** Let \( p < 1/2 \). There exists \( \gamma_0 \) and \( d_0 \) such that for all \( \gamma > \gamma_0 \) and \( d > d_0 \), if \( k \leq \frac{n}{18e\gamma d^2} \), we have:
\[
\sum_{\ell=2k+1}^{\gamma dk} \Pr \left[ \text{Bin} \left( \ell, 1 - \frac{\ell}{\gamma n} \right) \right] \leq \left. \frac{\ell}{2} \right] \cdot \Pr[\text{HyperGeom}(\gamma dn, \gamma d(1-p)n, \gamma dk) = \ell] \leq e^{-k} \left( \frac{k}{n} \right)^{k+1/2}
\]

This gives us the desired result for sufficiently large \( \gamma d \):
\[
\Pr[A_S \in M_k \setminus (T_k \cup T^+_k \cup T^-_k)] \leq \left( p^{\gamma d + \log(\gamma d)} \right)^k \left( \frac{k}{n} \right)^{k+1} + e^{-k} \left( \frac{k}{n} \right)^{k+1/2}
\]
\[
\leq O \left( e^{-k} \left( \frac{k}{n} \right)^{k+1/2} \right).
\]

This yields the first statement in the lemma.

Finally we prove the final statement about pairs of sets \( S \) and \( T \) of size \( k \) and \( j \) respectively. Recall that \( R = S \cup T \). Since \( A_S \in T_k \cup T^+_k \cup T^-_k \), and \( A_S \in T_j \cup T^+_j \cup T^-_j \), \( A_S \) contains at most \( 2k \) non-zero entries and \( A_T \) contains at most \( 2j \) non-zero entries. Hence \( A_R \) contains at most \( 2(k+j) \leq 4|R| \) non-zero entries. By Lemma, \( A_R \) must contain at least \( |R| - 1 \) non-zero entries that are not the first in their column.

By the same argument as before, conditioned on \( L_R = \ell \), the probability that \( A_S \in T_k \cup T^+_k \cup T^-_k \), and \( A_S \in T_j \cup T^+_j \cup T^-_j \) is at most \( \Pr[\text{Bin}(\ell, 4\ell R / \gamma(1-p)n) \geq |R| - 1] \). We check that the conditions of Claim hold with \( K = |R| \) and \( N = n \). Since we only consider the case when \( L_R \leq 4|R| \), we must have \( K = |R| \leq \frac{p^{\gamma d n}}{2\gamma d^2} = \frac{p^{\gamma d N}}{\gamma d^2} \leq \frac{3}{2} pn \) for \( d \geq 12 \). Hence employing Claim yields the lemma for \( d_0 \geq 12 \) and \( \gamma_n = \gamma \).

**8.2 Large Case**

The main goal of the large case is to prove the following lemma.

**Lemma 25 (ABC Large Case).** Let \( A \sim \text{ABC}_p(n, \gamma, d) \) for some constant \( \gamma \geq 16 \). Then there exists a constant \( d_0(\gamma) \) such that for \( d \geq d_0 \),
\[
\Pr \left[ \exists x : x^T A = 0, \text{supp}(x) \geq \frac{n}{18e\gamma d^2} \right] \leq o(1).
\]

(61)
Our main tool in proving this Lemma is an anti-concentration Lemma that is based on the sparse Littlewood-Offord Theorem from [9].

While typically such anti-concentration lemmas concern the dot product of a deterministic vector and a random vector with independent entries, we derive a weaker result which concerns the dot product of a deterministic vector with a vector in which there are a fixed number of non-zero entries whose positions are random.

**Lemma 26 (Anti-concentration for Sparse Regular Vectors (Same as Lemma 5)).** Let \( v \in \mathbb{R}^N \) be an arbitrary vector whose most common entry is \( a \). Then for any \( d \leq \sqrt{\frac{N}{2}} \), if \( x \in \{0,1\}^N \) is sampled uniformly from the set of vectors with exactly \( d \) 1s, we have:

\[
\Pr[x \cdot v = w] \leq 1/2 + \frac{d^2}{N}.
\]

for all \( w \in \mathbb{R}\{da\} \).

**Remark 5.** As \( d \to \infty \), the anti-concentration probability approaches the smaller value of \( 1/e \), though \( 1/2 \) is tight for \( d = 2 \).

**Proof.** For \( i \in \{1,...,d\} \), let \( j_i \sim \text{Uniform}([N]) \), and let \( X_i := v_{j_i} \). Then

\[
\Pr[X \cdot v = w] = \Pr \left[ \sum_{i=1}^{d} X_i = w \mid \text{All } j_i \text{ are unique} \right] 
\leq \frac{\Pr \left[ \sum_{i=1}^{d} X_i = w \right]}{\Pr[\text{All } j_i \text{ are unique}]}.
\]

**Claim 7.**

\[
\Pr[\text{All } j_i \text{ are unique}] \geq 1 - \frac{d^2}{N}.
\]

**Proof.** By a union bound,

\[
\Pr[\exists i \neq \ell : j_i = j_\ell] \leq \binom{d}{2} \cdot \Pr[j_1 = j_2] 
= \binom{d}{2} \left( \frac{1}{N} \right) 
\leq \frac{d^2}{N}.
\]

Next we show by induction on \( d \) that \( \Pr \left[ \sum_{i=1}^{d} X_i = w \right] \leq 1/2 \) for all \( w \in \mathbb{R}\{da\} \). For \( d = 1 \), we note that the chosen element \( w \) cannot be the most common element \( a \). Thus, \( w \) is at worst equally as common as the most common element \( a \). This implies that the number of times \( w \) appears in \( v \) is at most \( \frac{N}{2} \), hence \( \Pr[X_1 = w] \leq \frac{1}{2} \).

Now assume that \( \Pr \left[ \sum_{i=1}^{d-1} X_i = u \right] \leq 1/2 \) holds for all \( u \neq (d-1)a \). For \( w \neq da \), we write:

\[
\Pr \left[ \sum_{i=1}^{d} X_i = w \right] = \sum_{x \in \text{Supp}(X_d)} \Pr[X_d = x] \cdot \Pr \left[ \sum_{i=1}^{d-1} X_i = w - x \right]
= \Pr[X_d = a] \cdot \Pr \left[ \sum_{i=1}^{d-1} X_i = w - a \right] + \sum_{x \in \text{Supp}(X_d) \setminus \{a\}} \Pr[X_d = x] \cdot \Pr \left[ \sum_{i=1}^{d-1} X_i = w - x \right]
\]
Let \( p_x := \Pr \left[ \sum_{i=1}^{d} X_i = w - x \right] \) such that by the induction hypothesis \( p_a \leq 1/2 \) as \( w - a = (d - 1)a \) if and only if \( w = da \). Thus, we conclude:

\[
\Pr \left[ \sum_{i=1}^{d} X_i = w \right] = \max_{p_x \leq 1/2} \left( \Pr \left[ X_d = a \right] \cdot p_a + \sum_{x \in \text{Supp}(X_d) \setminus \{a\}} \Pr \left[ X_d = x \right] \cdot p_x \right) \\
\leq \max_{\sum_{x \neq a} p_x \leq 1/2} \left( \Pr \left[ X_d = a \right] \cdot (1/2) + \sum_{x \in \text{Supp}(X_d) \setminus \{a\}} \Pr \left[ X_d = x \right] \cdot p_x \right) \\
\leq \max_{\sum_{x \neq a} p_x \leq 1/2} \left( \Pr \left[ X_d = a \right] \cdot (1/2) + \left( \sum_{x \in \text{Supp}(X_d) \setminus \{a\}} P(X_d = x) \right) \cdot \left( \sum_{x \in \text{Supp}(X_d) \setminus \{a\}} p_x \right) \right) \\
= \Pr \left[ X_d = a \right] \cdot (1/2) + (1 - \Pr[X_d = a]) \cdot (1/2) \\
= 1/2.
\]

Here the second inequality follows from the fact that the optimum over the \( p_x \) is achieved by putting the maximum possible mass on \( p_a \), that is, \( p_a = 1/2 \).

Returning to Equation 62 we have for all \( w \neq ad \) and \( d \leq \sqrt{\frac{N}{2}} \),

\[
\Pr[x \cdot v = w] \leq \frac{\Pr \left[ \sum_{i=1}^{d} X_i = w \right]}{\Pr[\text{All } j_i \text{ are unique}]} \leq \frac{1/2}{1 - \frac{d^2}{N}} \leq 1/2 + \frac{d^2}{N},
\]

where the final inequality follows from the fact that \( \frac{1}{1-x} \leq 1 + 2x \) for \( 0 \leq x \leq 1/2 \). \( \square \)

To prove Lemma 25 we take a similar approach to the large case for the BGC. For a fixed set \( S \) of size \( k \), we consider the stochastic process \( A_S^3, A_S^2, \ldots \) in which we add the columns of \( A_S \) one by one.

We define the space:

\[
D(A_S^2) = \text{Span} \left( \{ v \in (\mathbb{R} \setminus \{0\})^k : vA_S^2 = 0 \} \right),
\]

and let

\[
R_j := \text{Rank}(D(A_S^2)).
\]

If \( R_{\gamma n/2} = 0 \), then \( A_S \) is not a minimal dependency.

Each time we add a new column, \( R_j \) either stays constant or decreases by at least 1 (note that \( R_j \) can decrease by more than 1: if a new column \( j \) has exactly one non-zero entry, then \( R_j = R_{j+1} = \ldots = 0 \) no matter what \( R_{j-1} \) was). We will use the following lemma to show for \( j \leq \gamma n/2 \), each column we add is close to random. Then, using Lemma 26 we show that with decent probability, \( R_j \) decreases. We can then apply a Chernoff Bound to show that \( R_{\gamma n/2} = 0 \) with high probability.

More formally, we consider the process of constructing \( H \in \{0,1\}^{dn \times \gamma n(1-p)} \) (the row half-edge occupancy matrix) one column at a time by pairing the \( d \) half-edges from each column-node at each step to a random set of \( d \) unpaired row-half-edges.

Define \( e(S, j) := \gamma dk - |H^j(S)| \), that is the number of unpaired half-edges among the \( k \) row-nodes in \( S \) after the first \( j \) column-nodes have randomly paired their half edges. The following lemma uses standard concentration bounds to show that for the first \( \gamma n/2 \) columns, there are many unpaired half-edges out of row-nodes in \( S \).

**Lemma 27.** Let \( A \sim \text{ABC}_{p}(n, \gamma, d) \). Let \( \Omega \) be the event in which for any set \( S \) with \( |S| \geq \frac{n}{18 \gamma d^2} \) and for any \( j \in [1, \frac{\gamma}{2} n] \), we have \( \frac{e(S, j)}{\gamma \eta_n} \geq \frac{k}{d^2} \). For \( \gamma \geq 2 \), there exists a constant \( d_0(\gamma) \) such that for \( d \geq d_0 \), we have \( \Pr[\Omega] \geq 1 - \gamma ne^{-n/d^3} \).

**Proof.** For \( d \) larger than some constant \( d_0(\gamma) \), we have \( \frac{n}{d^3} \leq \frac{n}{18 \gamma d^2} \), hence it suffices to prove the result for all \( k \in [n/d^3, n] \). Fix \( k \) and let us define \( \eta_k \) to be the ratio \( k/n \). Choose a subset \( S \) of size \( k \). Furthermore, fix \( j \leq \frac{\gamma}{2} n \). We proceed by applying the following tail bound on Hypergeometric distributions to \( e(S, j) \) which is clear from [4].
Lemma 28 (Hypergeometric Tail Bound from [6]). Let $X$ be given by a hypergeometric distribution with parameters $A, B, c$. Then for all $t < \frac{B}{A}$,

\[
\Pr [X \leq tc] \leq e^{-cD_{KL}(t||B/A)}.
\]  

(65)

Similarly, letting $Y$ be given by a hypergeometric distribution with parameters $A, A - B, c$, for $t > \frac{B}{A}$, we have

\[
\Pr [X \geq tc] = \Pr [Y \leq (1 - t)c] \leq e^{-cD_{KL}(t||1 - B/A)} = e^{-cD_{KL}(t||B/A)}.
\]  

(66)

Using this bound, we have

\[
P[e(S, j) \leq (\eta_k - t)(\gamma dn - dj)] \leq e^{-D_{KL}(\eta_k - t||\eta_k)(\gamma dn - dj)} \leq e^{-D_{KL}(\eta_k - t||\eta_k)^{\frac{2dn}{4}}}
\]

for all $t \in (0, \eta_k)$. The following claims expands this KL-divergences for $t = \eta_k/2$.

Claim 8.

\[
D_{KL}(\eta_k - \eta_k/2||\eta_k) \geq \frac{\eta_k}{12}.
\]  

(67)

Proof.

\[
D_{KL}(\eta_k/2||\eta_k) = \frac{\eta_k}{2} \ln(1/2) + \left(1 - \frac{\eta_k}{2}\right) \ln \left(\frac{1 - \eta_k/2}{1 - \eta_k}\right)
\]

\[
\geq \frac{\eta_k}{2} \ln(1/2) + \left(1 - \frac{\eta_k}{2}\right) \left(\frac{\eta_k}{2} + \frac{3\eta_k^2}{8}\right)
\]

\[
= \frac{\eta_k}{2} \left(\ln(1/2) + \left(1 - \frac{\eta_k}{2}\right) \left(1 + \frac{3\eta_k}{4}\right)\right)
\]

\[
\geq \frac{\eta_k}{12}.
\]  

(68)

Here the first inequality follows by using the Taylor expansion for $\ln(1 - x)$ and the final inequality follows by noting that the quadratic $\ln(1/2) + \left(1 - \frac{\eta_k}{2}\right) \left(1 + \frac{3\eta_k}{4}\right)$ achieves its minimum over $\eta_k \in [0, 1]$ at $\eta_k \in \{0, 1\}$. □

Union bounding over all $\binom{n}{k}$ sets $S$ of size $k$, we have

\[
\Pr \left[ \exists S, k \mid |S| \in [n/d^3, n] : e(S, j) \leq \frac{\eta_k}{2}(\gamma dn - dj) \right] \leq \sum_{k=n/d^3}^{n} \binom{n}{k} e^{-\frac{\eta_k \gamma dn}{24}}
\]

\[
\leq \sum_{k=n/d^3}^{n} \left(\frac{en}{k}\right)^{k} e^{-\frac{k\gamma d}{24}}
\]

\[
= \sum_{k=n/d^3}^{n} e^{-k\left(\frac{\gamma d}{24} - \ln(1/\eta_k) - 1\right)}
\]

\[
\leq \sum_{k=n/d^3}^{n} e^{-k\left(\frac{\gamma d}{24} - \ln(d^2) - 1\right)}
\]

\[
\leq \sum_{k=n/d^3}^{n} e^{-k}
\]

\[
\leq 2e^{-n/d^3}.
\]

We take a union bound over $j \in [1, \gamma n/2]$ to achieve

\[
\Pr \left[ \exists S, k \mid |S| \in [n/d^3, n] : e(S, j) \leq \frac{\eta_k}{2}(\gamma dn - dj) \right] \leq \gamma ne^{-n/d^3}.
\]  

\[\square\]
Conditioned on Ω, we can use our anti-concentration bound, Lemma \ref{lem:anti-concentration} to show that for \( j \leq \gamma n/2 \), \( \mu_j \) often decreases.

**Lemma 29.** For \( j \leq \gamma n/2 \), conditioned on Ω, if \( R_j \geq 1 \), then with probability at least

\[
\mu_k := \frac{1}{2} \left( 1 - e^{-\frac{\theta}{3k}} \right) - \frac{3\gamma d^3}{k},
\]

we have

\[
R_j \leq R_{j-1} - 1.
\]

**Proof.** Suppose \( R_{j-1} \geq 1 \), and let \( v \) be any vector in \((\mathbb{R}\setminus\{0\})^k \) such that \( A_{S,j}^{-1}v = 0 \). Let \( v' \in \mathbb{R}^{\gamma dk} \) be the vector which repeats each coordinate of \( v \) \( \gamma d \) times: that is, \( v'_i = v(\frac{i}{\gamma d}) \).

Let \( h_j \in \{0,1\}^{\gamma dn} \) be \( j \)th column of \( H \), which has exactly \( d \) ones indicating the half-edges matched to the \( d \) half-edges from the column-node \( j \). Let \( h_j(S) \in \{0,1\}^{\gamma dk} \) be the restriction of \( h_j \) to the entries corresponding to half-edges from row-nodes in \( S \).

**Claim 9.** If at most one half-edge is matched from column-node \( j \) to a single row-node in \( S \) and \( h_j(S) \cdot v' \neq 0 \), then \( R_j \leq R_{j-1} - 1 \).

**Proof.** If at most one half-edge is matched from column-node \( j \) to a single row-node in \( S \), then

\[
v \cdot (A_S)_j = \sum_{i \in S} v_i A_{ij} = \sum_{i \in S} v_i(\exists \ell, \gamma d(i-1) < \ell \leq \gamma di, H_{ij} = 1) = \sum_{i \in S} v_i \sum_{\ell = \gamma di}^{\gamma di + \gamma d} H_{ij} = v' \cdot h_j(S) \neq 0.
\]

Hence \( v \notin D(A_{S,j}^2) \), so the rank of \( D(A_{S,j}^2) \) is strictly smaller than that of \( D(A_{S,j}^{j-1}) \).

**Claim 10.**

\[
\Pr[h_j(S) \cdot v' = 0 | \Omega] \leq \frac{1}{2} \left( 1 + e^{-\frac{\theta}{2k}} \right) + \frac{d}{\gamma k}. \tag{69}
\]

**Proof.** We condition on the number of non-zero entries in \( h_j(S) \), which we denote \( s \). Observe that conditioned on \( s \), the vector \( h_j(S) \) is a uniformly random vector from the set of all vectors in \( \{0,1\}^{\gamma dk} \) with \( s \) 1s. This holds even when we conditioned on Ω, because this event says nothing about which half-edges among the nodes in \( S \) have been paired. By Lemma \ref{lem:anti-concentration}, since \( v' \) contains no zeros (and hence its most common element is not zero), we have

\[
\Pr[h_j(S) \cdot v' = 0 | s, \Omega] \leq \frac{1}{2} + \frac{d}{\gamma k}. \tag{70}
\]

It remains to consider the probability that \( s = 0 \), since in this case, we always have \( h_j(S) \cdot v' = 0 \). We know that \( s \) is distributed like a hypergeometric random variable \( \text{HyperGeom}(\gamma dn - d j, d, e(S,j)) \). Indeed, there are \( d \) half-edges that are paired with the addition of the \( (j+1) \)-th column, there are \( e(S,j) \) unpaired half-edges among the nodes in \( S \), and there are \( \gamma dn - dj \) total unpaired half-edges among the row nodes. Since we have conditioned on Ω, we know that \( \frac{s(S,j)}{\gamma dn - dj} \geq \frac{k}{2n} \). Hence we can compute

\[
\Pr[s = 0 | \Omega] \leq \left( 1 - \frac{k}{2n} \right)^d \leq e^{-\frac{dk}{2n}}. \tag{71}
\]

Combining Equations \ref{eq:claim9} and \ref{eq:claim10} it follows that

\[
\Pr[h_j(S) \cdot v' = 0 | \Omega] = \Pr[s = 0 | \Omega] + (1 - \Pr[s = 0 | \Omega]) \Pr[h_j(S) \cdot v' = 0 | s \geq 1, \Omega]
\]

\[
\leq \Pr[s = 0 | \Omega] + (1 - \Pr[s = 0 | \Omega]) \left( \frac{1}{2} + \frac{d}{\gamma k} \right)
\]

\[
\leq \Pr[s = 0 | \Omega] + (1 - \Pr[s = 0 | \Omega]) \left( \frac{1}{2} + \frac{d}{\gamma k} \right)
\]

\[
= \frac{1}{2} \left( 1 + \Pr[s = 0 | \Omega] \right) + \frac{d}{\gamma k} \leq \frac{1}{2} \left( 1 + e^{-\frac{\theta}{2k}} \right) + \frac{d}{\gamma k}.
\]
Claim 11. The probability that more than one half-edge is matched from a column-node \( j \) to a single row-node in \( S \) is at most \( \frac{2\gamma d^3}{k} \).

Proof. Conditioned on \( \Omega \), we have at each step \( j \), there are at least \( \frac{k(\gamma d n - d j)}{2 n \gamma d} \) open half-edges out of \( S \), and hence at least \( \frac{k(\gamma d n - d j)}{2 n \gamma d} \) nodes in \( S \) with at least one open half-edge. Since each node in \( S \) has at most \( \gamma d \) open half-edges, each pair of half-edges from the \( j \)th column collide with a row-node with probability at most \( \frac{\gamma d}{k/4} \). Hence by a union bound, the probability of collision is at most

\[
\binom{d}{2} \left( \frac{\gamma d}{k/4} \right) \leq \frac{2\gamma d^3}{k}.
\]

It follows from the previous two claims that the probability that \( h_j(S) \cdot v' \neq 0 \) and at most one half-edge is match from a column-node \( j \) to a single row-node in \( S \) is at least

\[
1 - \Pr[\text{Event in Claim 11 occurs}] - \Pr[h_j(S) \cdot v' = 0|\Omega] \geq 1 - \frac{2\gamma d^3}{k} - \frac{d}{\gamma k} - \frac{1}{2} \left( 1 + e^{-\frac{dk}{2n}} \right)
= \frac{1}{2} \left( 1 - e^{-\frac{dk}{2n}} \right) - \frac{2\gamma d^3}{k} - \frac{d}{\gamma k}
\geq \frac{1}{2} \left( 1 - e^{-\frac{dk}{2n}} \right) - \frac{3\gamma d^3}{k}.
\]

(73)

Using Claim 9, this proves the lemma.

We are now ready to prove Lemma 25.

Proof of Lemma 25. Recall that our goal is to show that with high probability, for all \( k \geq \frac{n}{18e\gamma d^2} \), for all sets \( S \) of size \( k \), we have \( R_{\gamma n/2} = 0 \). Throughout the rest of the proof, we assume that we have conditioned on \( \Omega \), since \( \Pr[\Omega] = 1 - o(1) \).

For a fixed set \( S \) of size \( k \), by Lemma 29, conditioned on each term being positive, the random process \( R_1, R_2, \cdots, R_{\gamma n/2} \) is stochastically dominated by the random process \( Y_1, Y_2, \ldots, Y_{\gamma n/2} \), where \( Y_{i+1} = Y_i - \text{Ber}(\mu_k) \) and

\[
\mu_k = \frac{1}{2} \left( 1 - e^{-\frac{dk}{2n}} \right) - \frac{3\gamma d^3}{k}.
\]

Hence

\[
\Pr[R_{\gamma n/2} > 0] \leq \Pr[Y_{\gamma n/2} > 0],
\]

By a Chernoff Bound, since \( Y_0 = k \), for any \( \mu \leq \mu_k \), we have

\[
\Pr[Y_{\gamma n/2} > 0] = \Pr[Y_0 - Y_{\gamma n/2} < k] \leq e^{-\gamma n \mu/2} \left( \frac{e^{\gamma n \mu}}{2k} \right)^k.
\]

(74)

Define \( \eta_k := \frac{k}{n} \). Taking a union bound over all sets \( S \) of size \( k \), conditioned on \( \Omega \), the probability that at least one set \( S \) has \( R_{\gamma n/2} > 0 \) is at most

\[
\binom{n}{k} e^{-\gamma n \mu/2} \left( \frac{e^{\gamma n \mu}}{2k} \right)^k \leq \left( \frac{e^{\gamma n \mu/2}}{k} \right)^k \left( \frac{e^{\gamma n \mu/2}}{2k} \right)^k
= e^{-\gamma n \mu/2} \left( \frac{e^{2 \gamma n^2 \mu / 2k}}{2k} \right)^k
= e^{-\gamma n \mu / 2 + k \log \left( \frac{2^{2 \gamma n^2 / 2k}}{k} \right)}
= e^{k \left( -\frac{2k}{2k} + \log \left( \frac{2k}{2k} \right) \right)}
\]

(75)
We consider two cases:

- Case 1: \( \frac{n}{18\gamma d^2} \leq k \leq \frac{5n}{d} \).
- Case 2: \( k \geq \frac{5n}{d} \).

In the first case, since \( 1 - e^{-x} \geq x/2 \) for \( 0 < x < 1 \), for \( n \) large enough, we have

\[
\mu_k \geq 1/2 \left( \frac{dn_k}{4} \right) - 3\gamma d^3 \left( \frac{dn_k}{k} \right) \geq \frac{dn_k}{10}.
\]

Hence for \( \mu = \frac{dn_k}{10} \), we have

\[
- \frac{\gamma \mu}{2\eta k} + \log \left( \frac{e^2 \gamma \mu}{2\eta_k^2} \right) \leq - \frac{\gamma d}{20} + \log \left( \frac{e^2 \gamma d}{20\eta_k^2} \right) \leq - \frac{\gamma d}{20} + \log \left( \frac{18^2 e^3 \gamma^3 d^5}{20} \right) \leq -0.25 \tag{76}
\]

for \( d \) larger than some constant \( d_0 \).

In the second case, we have \( \mu_k \geq 0.45 \), and hence for \( \mu = 0.45 \) and \( \gamma \geq 16 \), we have

\[
- \frac{\gamma \mu_k}{2\eta_k} + \log \left( \frac{e^2 \gamma \mu_k}{2\eta_k^2} \right) \leq \max_{\eta} \left[ - \frac{3.6}{\eta} + \log \left( \frac{3.6e^2}{\eta^2} \right) \right] \leq -0.25. \tag{77}
\]

Returning to Equation 75 conditioned on \( \Omega \), the probability that there is a set \( S \) of size \( k \) for which \( R_{\gamma n/2} > 0 \) is at most \( e^{-k/4} \). Summing over all \( k \geq \frac{n}{18\gamma d^2} \) yield a probability of failure among any \( k \) of at most \( e^{-n/64 e\gamma d^2} \).

Unioning with the probability that \( \Omega \) doesn’t occur from Lemma 27 yields Lemma 25.

8.3 Proof of Theorem 3

We are now ready to prove Theorem 3, which we restate here for the reader’s convenience.

Theorem 3 (Characterization ABC). There exist universal constants \( c \) and \( \gamma_0 \) such that for any constant \( p < 1/2 \), \( \gamma > \gamma_0 \), and \( d \geq d_0(\gamma) \), for \( A \sim \text{ABC}_p(n, \gamma, d) \), with probability \( 1 - o(1) \):

1. All minimal dependencies of \( k \) rows of \( A \) are in \( \mathcal{T}_k \cup \mathcal{T}_k^+ \cup \mathcal{T}_k^C \).

2. The number of rows involved in a linear dependency of \( A \) is at most \( np^{\gamma d - c \log(\gamma d)} \), that is

\[
\left| \bigcup_{x: x^T A = 0} \text{supp}(x) \right| \leq np^{\gamma d - c \log(\gamma d)}.
\]

Proof of Theorem 3. By combining the result of the small case, Lemma 24, with the result of the large case, Lemma 25, we observe that \( 1 - o(1) \), there are no minimal dependencies in \( A \) that are not in

\[
\bigcup_{k \leq \log(n)} \mathcal{T}_k \cup \mathcal{T}_k^+ \cup \mathcal{T}_k^C.
\]

This proves the first statement in the theorem.

Next we bound the size of \( D \), the set of rows involved in linear dependencies:

\[
D := \bigcup_{x: x^T A = 0} \text{supp}(x).
\]

Define

\[
\mathcal{T}_k := \mathcal{T}_k \cup \mathcal{T}_k^+ \cup \mathcal{T}_k^C.
\]
For a set $S \subseteq [n]$, let $X_S$ be the indicator of the event that $A_S \in \tilde{T}_k$.

Notice that with probability $1 - o(1)$, $|D|$ is at most

$$X := \sum_{S \subseteq [n] : |S| \leq \log(n)} |S| X_S.$$  \hfill (78)

We will bound $X$ with high probability via the second moment method with Lemma 24 as our main tool.

First, we compute the expectation of $X$.

Let $c$ be a permutation defining the mapping of half-edges used to generate $\gamma_d k$. Proof of Claim 12. Let $\gamma_d k, \gamma_d n$ be the set of injections from $[\gamma_d k]$ to $[\gamma_d n]$. For $\pi \in P_{\gamma_d k, \gamma_d n}$, recall that $A(\pi) \in \{0, 1\}^{k \times \gamma_d n(1-\rho)}$ is the first $k$ rows of the ABC$_p$ matrix which is generated from mapping half-edges according to $\pi$ and then dropping the last $\gamma m$ columns.

Let $R_k = \{ \pi \in P_{\gamma_d k, \gamma_d n} : A(\pi) \in \tilde{T}_k \}$.

Since each element of $P_{\gamma_d k, \gamma_d n}$ is equally likely to be the restriction of $\rho$ to $[\gamma_d k]$, we have

$$E[X_S] = \frac{|R_k|}{|P_{\gamma_d k, \gamma_d n}|}.$$  \hfill (79)

Our goal is to compute

$$E[X_S X_T] - E[X_S]E[X_T] = \sum_{\rho \sim S, \rho \sim T} \Pr[\rho_S \in R_k \land \rho_T \in R_j] - \Pr[\rho_S \in R_k] \Pr[\rho_T \in R_j]$$

$$= \sum_{\pi_1 \in R_k, \pi_2 \in R_j} \left( \Pr[\rho_S = \pi_1 | \rho_T = \pi_2] - \Pr[\rho_S = \pi_1] \Pr[\rho_T = \pi_2] \right).$$  \hfill (80)

Here we have abused notation to interpret the restriction of $\rho$ to a set $U$ as being an element of $P_{\gamma_d |U|, \gamma_d n}$.

If the images of $\pi_1$ and $\pi_2$ intersect, then $\Pr[\rho | \rho_S = \pi_1 | \rho_S = \pi_2] = 0$.

Otherwise, if the images do not intersect, we have

$$\Pr[\rho_S = \pi_1 | \rho_S = \pi_2] = \frac{1}{|P_{\gamma_d k, \gamma_d n - \gamma_d k}|} = \Pr[\rho_S = \pi_1] \frac{|P_{\gamma_d k, \gamma_d n - \gamma_d k}|}{|P_{\gamma_d k, \gamma_d n}|}. $$  \hfill (81)

Therefore, the expectation of $X_S X_T$ is

$$E[X_S X_T] = \sum_{\rho \sim S, \rho \sim T} \Pr[\rho_S \in R_k \land \rho_T \in R_j] - \Pr[\rho_S \in R_k] \Pr[\rho_T \in R_j]$$

$$= \sum_{\pi_1 \in R_k, \pi_2 \in R_j} \left( \Pr[\rho_S = \pi_1 | \rho_T = \pi_2] - \Pr[\rho_S = \pi_1] \Pr[\rho_T = \pi_2] \right).$$  \hfill (82)

In the next claim, we show that if $T \cap S = \emptyset$, then the events $X_S$ and $X_T$ are almost uncorrelated.

**Claim 12.** Let $T, S \subseteq [n]$ and $|T|, |S| \leq \log(n)$. If $T \cap S = \emptyset$, then

$$E[X_S X_T] \leq \left( 1 + \frac{6\gamma d \min(|S|, |T|)}{n} \right) E[X_S]E[X_T].$$  \hfill (83)
It follows that
\[
\mathbb{E}[X_S X_T] - \mathbb{E}[X_S] \mathbb{E}[X_T] = \sum_{\pi_1 \in \mathcal{R}_1, \pi_2 \in \mathcal{R}_2} (\Pr[\rho_S = \pi_1 | \rho_T = \pi_2] - \Pr[\rho_S = \pi_1]) (\Pr[\rho_T = \pi_2]) \\
= \sum_{\pi_1 \in \mathcal{R}_1, \pi_2 \in \mathcal{R}_2} \left( \frac{\left| \mathcal{P}_{\rho_T = \pi_2} \right|}{\left| \mathcal{P}_{\rho_T = \pi_2, \rho_S = \pi_1} \right|} - 1 \right) \Pr[\rho_T = \pi_2] (\Pr[\rho_S = \pi_1]) \\
= \left( \frac{(\gamma d)!((\gamma d) + 1)!}{(\gamma d)!(\gamma d - 1)!} - 1 \right) \mathbb{E}[X_S] \mathbb{E}[X_T] \\
= \left( \prod_{i=1}^{\gamma d} \frac{\gamma d - i + 1}{\gamma d - \gamma d - i + 1} - 1 \right) \mathbb{E}[X_S] \mathbb{E}[X_T] \\
\leq \left( 1 + \frac{1}{\gamma d} \right)^{\gamma d} - 1 \mathbb{E}[X_S] \mathbb{E}[X_T] \\
\leq \left( 1 + \frac{3k}{\gamma d} \right)^{\gamma d} - 1 \mathbb{E}[X_S] \mathbb{E}[X_T] \\
\leq \left( 1 + \frac{6k^2}{\gamma d} \right) \mathbb{E}[X_S] \mathbb{E}[X_T].
\]
(84)

where the first inequality follows from the fact that \( \frac{1}{1+x} \leq 1 + 2x \) for \( x \leq \frac{1}{4} \).

We can now bound the variance of \( X \):
\[
\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
= \sum_{S: |S| \leq \log(n)} \sum_{|T| \leq \log(n)} |S||T| (\mathbb{E}[X_S X_T] - \mathbb{E}[X_S] \mathbb{E}[X_T]) \\
\leq \sum_{S: |S| \leq \log(n)} \sum_{T: |T| \neq \emptyset} |S||T| \sum_{T: T \cap S \neq \emptyset} (\mathbb{E}[X_S X_T] - \mathbb{E}[X_S] \mathbb{E}[X_T]) + \sum_{S: |S| \leq \log(n)} \sum_{T: T \cap S \neq \emptyset} |S||T| \mathbb{E}[X_S X_T] \\
\leq \frac{6\gamma d \log(n)^2}{n} \mathbb{E}[X^2] + \sum_{S: |S| \leq \log(n)} \sum_{T: T \cap S \neq \emptyset} |S||T| \left( \frac{|S \cup T|}{n} \right)^{\frac{|S \cup T|}{n} - 1} \rho^{\gamma d |S \cup T|} \mathbb{E}[X^2] \\
\leq \frac{6\gamma d \log(n)^2}{n} \mathbb{E}[X^2] + \sum_{R: 1 \leq |R| \leq 2 \log(n)} \sum_{S: T \leq R} |R|^2 \left( \frac{|R|}{n} \right)^{|R| - 1} \rho^{\gamma d |R| - |R| \log(n)} \\
\leq \frac{6\gamma d \log(n)^2}{n} \mathbb{E}[X^2] + \sum_{R: 1 \leq |R| \leq 2 \log(n)} 2^{2|R|} |R|^2 \left( \frac{|R|}{n} \right)^{|R| - 1} \rho^{\gamma d |R| - |R| \log(n)} \\
\leq \frac{6\gamma d \log(n)^2}{n} \mathbb{E}[X^2] + n \rho^{\gamma d - c_2 \log(n)} \\
\leq \frac{6\gamma d \log(n)^2}{n} \mathbb{E}[X^2] + np^{\gamma d - c_2 \log(n)} + np^{\gamma d - c_2 \log(n)}. \tag{85}
\]

for some constant \( c_2 \) (the second to last line follows from a similar calculation as used when computing \( \mathbb{E}[X] \)).

It follows by Markov’s law that for any \( t \),
\[
\Pr[X \geq \mathbb{E}[X] + t] \leq \frac{\text{Var}(X)}{t^2} \tag{86}
\]

38
Plugging in $t = (np^d)^{3/4}$, we have

$$\Pr[X \geq \mathbb{E}[X] + n^{3/4}p^{-\gamma d}] \leq \frac{\text{Var}(X)}{(np^d)^{3/2}} \leq \frac{6\gamma d \log(n)^2 np^{2\gamma d - 2c_1 \log(\gamma d)} + np^{\gamma d - c_2 \log(\gamma d)}}{(np^d)^{3/2}} \leq b^{-1/2} \left(6\gamma d \log(n)^2 p^{\gamma d/2 - 2c_1 \log(\gamma d)} + p^{-\gamma d/2 - c_2 \log(\gamma d)}\right) \leq n^{-1/2} p^{-\gamma d/2 - c_2 \log(\gamma d)} \log(n)^2$$

It follows that with probability at most $n^{-1/2} p^{-\gamma d/2 - c_2 \log(\gamma d)} \log(n)^2 = o(1)$, we have

$$X \leq np^{\gamma d - c_1 \log(\gamma d)} + \sqrt{np^{\gamma d} \log(n)^2 p^{-c_2 \log(\gamma d)}} \leq np^{\gamma d - c \log(\gamma d)},$$

for some constant $c$. This proves the theorem.

9 Applications to Gradient Coding

In the following section we will address the motivating application of our work: the design of gradient codes with small decoding error. The literature on gradient codes largely focuses on the special case where the assignment matrix is a square. Hence, to align our results with this standard, we introduce the following stacked ABC construction, which will us to apply our results on wide matrices to square $n \times n$ assignment matrices.

**Definition 7.** For $\gamma, d, n \in \mathbb{Z}^+$ such that $\gamma \mid d$ and $\gamma \mid n$, we define the $\gamma$-stacked Augmented Biregular Code $B$ to be an $n \times n$ matrix formed by sampling $A_0 \sim \text{ABC}(n/\gamma, \gamma, d/\gamma)$ and stacking $\gamma$ identical copies of $A_0$. We will denote the distribution of such matrices as $\text{ABC}_{\gamma \text{-stacked}}(n, d)$ (see Figure 3).

**Remark 6.** We note that these stacked designs can be viewed as a natural generalization of the Fractional Repetition Code (FRC) of [21]. In particular, any $B \sim \text{ABC}_{\text{stacked}}(N, d, d)$ is the $N \times N$ FRC matrix with $d$ ones in each column, up to a permutation of the rows and columns. The key improvement is that our generalization allows the stacking variable $\gamma$ — which increases the adversarial decoding error — to stay constant for arbitrarily large $d$.

Using these stacked ABC matrices, we will prove the following theorem:

**Theorem 7.** Let $c, \gamma_0, d_0$ be the universal constants from Theorem 6. Choose any $\gamma, d \in \mathbb{Z}^+$ such that $\gamma \geq \gamma_0$, $\gamma \mid d$ and $\frac{d}{\gamma} \geq d_0$. For any sufficiently large $n$ divisible by $\gamma$, let $B \sim \text{ABC}_{\text{stacked}}(n, \gamma, d)$. Then with constant probability over the choice of $B$:

$$\frac{1}{n} \mathbb{E}_{S \sim \text{bin}(n, p)} \text{err}(B, S) \leq p^{d - c \log(d)} + o(1),$$

(8)
and

\[
\frac{1}{n} \max_{S \in \binom{[n]}{\left\lfloor n/\gamma \right\rfloor}} (\text{err}(B, S)) \leq \left( \frac{8\gamma^3 p}{d} \right) + o(1). \tag{9}
\]

While the first bound will be a direct corollary of Theorem \[6\], the second bound relies on some external lemmas that will be introduced below. Thus, for the sake of clarity, we will split the proof of Theorem \[7\] into the following two lemmas.

Lemma 30. Let \(c, \gamma_0, d_0\) be the universal constants from Theorem \[6\]. Choose any \(\gamma, d \in \mathbb{Z}^+\) such that \(\gamma \geq \gamma_0\), \(\gamma \mid d\), and \(d \geq d_0\). For sufficiently large \(N\) such that \(\gamma \mid n\), let \(B \sim \text{ABC}_{\text{stacked}}(n, \gamma, d)\). Then with probability \(1 - o(1)\) over the choice of \(B\), we have that for any \(p < \frac{1}{2}\):

\[
\frac{1}{n} E_{S \sim \binom{[n]}{\lfloor n/\gamma \rfloor}} (\text{err}(B, S)) \leq p^{d+c\log(d)} + o(1) \tag{88}
\]

Lemma 31. Choose any \(\gamma, d, N \in \mathbb{Z}^+\) such that \(\gamma \mid d\) and \(\gamma \mid n\). Let \(B \sim \text{ABC}_{\text{stacked}}(n, \gamma, d)\). With constant probability, we have:

\[
\frac{1}{n} \max_{S \in \binom{[n]}{\left\lfloor n/\gamma \right\rfloor}} (\text{err}(B, S)) \leq \left( \frac{8\gamma^3 p}{d} \right) + o(1). \tag{89}
\]

Temporarily assuming these lemmas, we note that Theorem \[7\] follows immediately.

We will now prove Lemma \[30\].

Proof of Lemma \[30\]. Let \(A_0 \sim \text{ABC}(n/\gamma, \gamma, d/\gamma)\) denote the ABC matrix which is stacked \(\gamma\) times to generate \(B\), then it follows that:

\[
\frac{1}{n} E_{S \sim \binom{[n]}{\lfloor n/\gamma \rfloor}} (\text{err}(B, S)) = \frac{1}{n} E_{S \sim \binom{[n]}{\lfloor n/\gamma \rfloor}} \min_{w : w_j = 0 \forall j \in S} |Bw - \mathbf{1}|_2^2 \\
= \frac{1}{n} E_{S \sim \binom{[n]}{\lfloor n/\gamma \rfloor}} \min_{w : w_j = 0 \forall j \in S} |A_0 w - \mathbf{1}|_2^2 \\
= \frac{\gamma}{n} E \min_w |A w - \mathbf{1}|_2^2
\]

where \(A \sim \text{ABC}_p(n/\gamma, \gamma, d/\gamma)\). First observe trivially that \(\frac{\gamma}{n} \min_w |A w - \mathbf{1}|_2^2\) is always at most 1. This can be seen by taking \(w\) to be the vector of all zeros. By Theorem \[4\] there is a \(1 - o(1)\) chance that \(A_0\) has the property that the following holds with probability \(1 - o(1)\) over the choice of \(S\):

\[
\frac{\gamma}{n} \min_w |A w - \mathbf{1}|_2^2 \leq p^{d+c\log(d)} + o(1). \tag{90}
\]

For \(A_0\) where this holds, we can calculate the expectation as:

\[
\gamma E \left( \min_w |A w - \mathbf{1}|_2^2 \right) \leq (1 - o(1)) \left( p^{d+c\log(\gamma d)} + o(1) \right) n + o(n). 
\]

This gives us the desired result:

\[
\frac{1}{n} E_{S \sim \binom{[n]}{\lfloor n/\gamma \rfloor}} (\text{err}(B, S)) \leq p^{d+c\log(d)} + o(1) \tag{92}
\]

for a matrix \(B \sim \text{ABC}_{\text{stacked}}(n, \gamma, d)\) with probability \(1 - o(1)\). \(\square\)

Lemma \[31\] requires a bit more machinery. We are going employ the following lemma from \[13\] which allows us to bound the adversarial error of a gradient code as a function of the second largest singular value. Formally, we have the following lemma.

Lemma 32 (Proposition 4.1 of \[13\]). Let \(A \in \{0, 1\}^{N \times M}\) be an assignment matrix such that each row has exactly \(D\) ones. Let \(\sigma_2\) be the largest singular value of \(A\). Then for any set of stragglers \(S\) such that \(|S| = s\), we have:

\[
\frac{1}{N} \text{err}(A, S) \leq \frac{1}{N} \left( \frac{\sigma_2}{D} \right)^2 \frac{sM}{M-s} \tag{93}
\]
To calculate an upper bound on the second largest singular value of an ABC matrix, we first need a result from [2] which states that with constant probability, the configuration model we use to generate our ABC matrix has no rows which map to the same column node more than once i.e., the bipartite graph produced is simple.

Let $\mathcal{G}(n, \gamma n, d, \gamma d)$ denote the uniform distribution on simple $(d, \gamma d)$-biregular bipartite graphs with $n$ left nodes and $\gamma n$ right nodes.

**Lemma 33** ([2]). Let $A_0 \sim ABC(n, \gamma, d)$, then the probability that $\begin{pmatrix} 0 & A_0 \\ A_0^T & 0 \end{pmatrix}$ is the adjacency matrix of a bipartite, biregular random graph $G$ is at least $\varepsilon(d) > 0$. Furthermore, if we condition on this event occurring, then $G \sim \mathcal{G}(n, \gamma n, d, \gamma d)$.

The previous lemma allows us to apply the following result of [3] to bound the second largest singular value of these well behaved ABC matrices.

**Lemma 34** (Theorem 4 of [3]). Let $A = \begin{pmatrix} 0 & X \\ X^T & 0 \end{pmatrix}$ be the adjacency matrix of a bipartite, biregular random graph $G \sim \mathcal{G}(n, \gamma n, d, \gamma d)$. Then, with probability $1 - o(1)$, $A$’s second largest eigenvalue $\lambda_2$ satisfies

$$\lambda_2 \leq \sqrt{d_1 - 1} + \sqrt{d_2 - 1} + o(1).$$

We are now ready to prove Lemma 31.

**Proof of Lemma 31**. Let $A_0 \sim ABC(n/\gamma, \gamma, d/\gamma)$ denote the $ABC$ matrix which is stacked $\gamma$ times to generate $B$. It follows that:

$$\frac{1}{n} \max_{S \in \binom{[n]}{m}} \text{err}(B, S) = \frac{1}{n} \max_{S \in \binom{[n]}{m}} \min_{w : w_j = 0, \forall j \in S} |Bw - 1|_2^2$$

$$= \frac{\gamma}{n} \max_{S \in \binom{[n]}{m}} \min_{w : w_j = 0, \forall j \in S} |A_0w - 1|_2^2$$

$$= \frac{\gamma}{n} \max_{S \in \binom{[n]}{m}} \text{err}(A_0, S)$$

By Lemma 33, with constant probability, $A' = \begin{pmatrix} 0 & A_0 \\ A_0^T & 0 \end{pmatrix}$ is the adjacency matrix of a bipartite, biregular random graph $G$. Condition on this event occurring. Then $G$ is uniformly sampled from $\mathcal{G}(n/\gamma, N, d/\gamma, d)$. By Lemma 34, with probability $1 - o(1)$, the second largest eigenvalue of $A'$ is $\lambda_2 \leq \sqrt{d_1 - 1} + \sqrt{d_2 - 1} + o(1)$. Thus, the second largest singular value of $A_0$ is $\sigma_2 \leq \sqrt{d_1 - 1} + \sqrt{d_2 - 1} + o(1)$.

Hence, by Lemma 32 we have:

$$\frac{\gamma}{n} \max_{S \in \binom{[n]}{m}} \text{err}(A_0, S) \leq \frac{\gamma}{n} \left( \frac{\gamma \sigma_2}{d} \right)^2 \frac{pm^2}{(1 - p)n}$$

$$\leq \gamma^3 \left( \frac{2\sqrt{d} + o(1)}{d} \right)^2 \frac{p}{(1 - p)}$$

$$\leq \left( \frac{4\gamma^3 p}{d(1 - p)} \right) + o(1)$$

$$\leq \left( \frac{8\gamma^3 p}{d} \right) + o(1)$$

as desired.

**Acknowledgements**

The authors thank Mary Wootters for helpful comments on this manuscript.
Table 2: Comparison of Related Work. We have normalized the decoding error by $1/N$.

| Assignment Matrix Design | Expected Decoding Error | Adversarial Decoding Error |
|--------------------------|-------------------------|-----------------------------|
| Expander Code (Cor. 23 [19]) | $-$ | $< \frac{3p}{d(1-p)}$ |
| Pairwise Balanced (11) | $\geq \frac{p}{d(1-p)}$ | $O\left(\frac{1}{\sqrt{m}}\right)$ |
| BIBD ( Const. 1 [17]) | $-$ | $d = \Omega(\sqrt{m})$ |
| BRC (25) | $d = \Omega\left(\frac{\log\log(m)}{\log(p)}\right)$ | $-$ |
| rBGC (5) | $< \frac{1}{(1-p)d}$ | $p$ |
| FRC of [21] (and [24]) | $p^{d-o(d)}$ | $\frac{(1+o(1))p}{d(1-p)}$ |
| Graph-based Assignment [13] | $p^{d-o(d)}$ | $\Theta(p/d)$ |

A Proof of Lemmas 6 and 7 for classification of minimal dependencies

Recall Lemma 6

**Lemma 6 (Classification of Dependencies).** We have the following three equivalences:

1. $S_{2k-2,k} = T_k$.
2. $S_{2k-1,k} = T_k^+.$
3. $S'_{2k,k} = T_k^C$.

We break the proof of Lemma 6 into three lemmas, which we prove independently.

**Lemma 35.** $S_{2k-2,k} = T_k$. Further, for $B \in T_k = S_{2k-2,k}$, there is a unique (up to constant multiple) non-zero vector $v$ satisfying $B^Tv = 0$, where $v_i = (-1)^{d_G(i,j)}v_j$, and $d_G(i,j)$ is the path length from vertex $i$ to $j$ in the graph $G$ encoded by $B$.

**Proof.** We prove this by showing both inclusions in the following two claims.

**Claim 13.** $T_k \subseteq S_{2k-2,k}$.

**Proof.** Let $B \in T_k$. Since $B$ has $k-1$ non-zero rows, it has rank at most $k-1$, and so the nullspace of $B^T$ has dimension at least 1. Let $v$ be any vector such that $B^Tv = 0$. Suppose $(i,j)$ is an edge in the tree $G$ encoded by $T_k$, then there is a column $C$ of $B$ such that $B$ has 1’s exactly at coordinates $i$ and $j$. We have $C \cdot v = 0$ thus $v_j = -v_i$. Applying this to paths of multiple edges gives that $v_i = (-1)^{d_G(i,j)}v_j$, where $d_G$ is the path length from $i$ to $j$ in $G$ (which is unique and well-defined because $G$ is a tree). By this formula, if any $v_i$ is non-zero, then all other entries of $v$ are non-zero, and are uniquely determined by $v_i$. Thus $B^T$ (and $B$) have rank $k-1$. We also know that $B$ has $k-1$ non-zero columns, each with 2 1’s, so it has $2k-2$ non-zero entries. Thus $B \in S_{2k-2,k}$. In addition, by the description of $v$, this establishes the second statement of the lemma.

**Claim 14.** $S_{2k-2,k} \subseteq T_k$.

**Proof.** Suppose $B \in S_{2k-2,k}$. Because $B$ has rank $k-1$ it must have at least $k-1$ non-zero columns. Recall that by Observation 11, no column of $B$ can have exactly one 1. Thus, $B$ must have exactly $k-1$ non-zero columns, each with exactly 2 1’s. Let $G$ be the graph on vertices $[1, \ldots, k]$ encoded by incidence matrix given by the non-zero columns of $B$. We claim $G$ is connected. Suppose otherwise, and without loss of generality let $[1, \ldots, L], [L+1, \ldots, k]$ be two disconnected components of $G$ for $1 \leq L \leq k-1$. By definition of $S_{2k-2,k}$ there is non-zero $v \in \mathbb{R}^k$ with $B^Tv = 0$, and $v$ has all non-zero entries. Let $v^{(L)}, v^{(R)} \in \mathbb{R}^k$ with $v^{(L)} = (v_1, v_2, \ldots, v_L, 0, 0, \ldots, 0)$ and $v^{(R)} = (0, 0, \ldots, 0, v_{L+1}, v_{L+2}, \ldots, v_k)$. $B^T(v^{(L)} + v^{(R)}) = B^Tv = 0$. 

42
By the disconnectedness of \([1, \ldots, L]\) and \([L + 1, \ldots, k]\), the set of columns of \(B\) having non-zero entries in rows \([1, \ldots, L]\) is disjoint from the set of columns of \(B\) having non-zero entries in rows \([L + 1, \ldots, k]\). Thus \(B^T v^{(L)}\) and \(B^T v^{(R)}\) are non-zero in disjoint coordinates. Thus \(B^T (v^{(L)} + v^{(R)}) = 0 \iff B^T v^{(L)} = B^T v^{(R)} = 0\). This implies the null space of \(B^T\) has rank at least 2, which contradicts that \(B\) has rank \(k - 1\). Thus \(G\) is a connected graph. Because \(G\) is a connected graph on \(k\) vertices with at most \(k - 1\) edges, \(B\) must encode a tree, and all its \(k - 1\) non-zero columns encode distinct edges. Thus \(B \in T_k\).

Call a column of a \(\{0,1\}\)-matrix an \(n\)-column if it has exactly \(n\) non-zero entries.

**Lemma 36.** \(T_k^+ = S_{2k-1,k}\)

**Proof.** We prove this by showing both inclusions in the following two claims.

**Claim 15.** \(T_k^+ \subset S_{2k-1,k}\).

**Proof.** Suppose \(B \in T_k^+\). \(B\) is the incidence matrix of a forest with two trees connected by a 3-hyperedge; without loss of generality we may assume the first column of \(B\) encodes this hyper-edge. Further, by relabeling rows appropriately, we may assume \(B_{11} = B_{21} = B_{k1} = 1\), with vertices 1 and 2 in the same tree and connected by an even path, and vertex \(k\) in the other tree. Then we may further relabel rows such that rows 1, 2, \ldots, \(L\) correspond to one tree, and \(L + 1, \ldots, k\) correspond to the other, for some \(L\) with \(2 \leq L \leq k - 1\).

Let \(B''\) be \(B\) without its first column. The entries of \(B''\), restricted to rows 1, \ldots, \(L\), define a tree in \(T_L\). So by the second statement of Lemma 35 there is a unique (up to constant multiple) non-zero vector \(v_\ell \in \mathbb{R}^L\) satisfying \((B''_{[L]})^T v_\ell = 0\). Similarly, there is a unique (up to constant multiple) non-zero vector \(v_r \in \mathbb{R}^{k-L}\) satisfying \((B''_{[k\setminus[L]]})^T v_r = 0\). Because the non-zero columns of rows 1, \ldots, \(L\) are disjoint from the non-zero columns of rows \(L + 1, \ldots, k\), the null space of \((B'')^T\) is the direct sum of the null spaces of \((B''_{[L]})^T\) and \((B''_{[k\setminus[L]]})^T\). Concretely, let \(v^{(L)}, v^{(R)} \in \mathbb{R}^k\) be defined \(v^{(L)} = v_\ell || 0^{k-L}\) and \(v^{(R)} = 0^L || v_r\), where \(||\) means concatenation. Then the null space of \(B''\) are precisely the vectors \(\alpha v^{(L)} + \beta v^{(R)}\) for \(\alpha, \beta \in \mathbb{R}\). Recall \(B\) is \(B''\), with the additional first column with three 1’s in \(B_{11}, B_{21}, B_{k1}\). Thus the null space of \(B\) is:

\[
\alpha v^{(L)} + \beta v^{(R)}: \alpha, \beta \in \mathbb{R} \text{ subject to } \alpha (v_1^{(L)} + v_2^{(L)}) + \beta v_k^{(R)} = 0
\]

From the second statement of Lemma 35 for \(B_L \in T_L\), a non-zero vector \(v\) satisfying \(B_L^T v = 0\) has \(v_a = (-1)^{d(a,b)} v_b\) relative to the encoded graph \(G_L\). We have \(B''_{[L]} \in T_L\). Vertices 1 and 2 are connected by an even path in the relevant graph, which establishes \(v_1^{(L)} = v_2^{(L)}\), so in particular \((v_1^{(L)} + v_2^{(L)}) \neq 0\). Thus, the null space of \(B\) is precisely:

\[
t (v_1^{(L)} v_r - (v_1^{(L)} + v_2^{(L)}) v_r) , t \in \mathbb{R}
\]

Observe that \((v_1^{(R)} v^{(L)} - (v_1^{(L)} + v_2^{(L)}) v^{(R)}) \in \mathbb{R}^k\) has all non-zero entries because the coefficients \(v_1^{(R)}\) and \((v_1^{(L)} + v_2^{(L)})\) are non-zero, and vectors \(v^{(L)}\) and \(v^{(R)}\) have disjoint support. Thus \(B \in S_{2k-1,k}\).

**Claim 16.** \(S_{2k-1,k} \subset T_k^+\).

**Proof.** Suppose \(B \in S_{2k-1,k}\). Since \(B\) is a minimal dependency, none of its columns have a single 1. Further, \(B\) must have at least \(k - 1\) non-zero columns since \(B\) must have rank \(k - 1\). Hence \(B\) has \(k - 2\) columns with two 1’s and one column with three 1’s. Without loss of generality we suppose \(B\)’s first column is the one with 3 1’s.

Again define \(B''\) as \(B\) omitting its first column. Because \(B\)’s non-zero columns are linearly independent, \(B''\) has rank \(k - 2\). We show that \(B''\) encodes two disjoint trees by proving that \(B''\) must encode a graph with no more than two connected components. Suppose for sake of contradiction that \(B''\) encodes a graph \(G''\) having connected components \(X, Y, Z\), where \(X, Y, Z\) are disjoint non-empty subsets of \([1, \ldots, k]\). Let \(v \in \mathbb{R}^k\) be an all-non-zero vector such that \(B''^T v = 0\). For a vertex set \(S \subset [1, \ldots, k]\) write \(v_S\) for the vector with \((v_S)_i = v_i\) for \(i \in S\), \((v_S)_i = 0\) otherwise. So, \(v = v_X + v_Y + v_Z\), and \(v_X, v_Y, v_Z\) have non-zero entries
in disjoint locations. We have \( B^T(v_X + v_Y + v_Z) = 0 \in \mathbb{R}^m \), where \( m \) is the number of columns in \( B \), so we have \( (B'')^T(v_X + v_Y + v_Z) = 0 \in \mathbb{R}^{m-1} \). But because \( X,Y,Z \) are disjoint in \( G'' \), the non-zero entries of \((B'')^T v_X, (B'')^T v_Y, (B'')^T v_Z \) must be pairwise disjoint. Thus, \((B'')^T(v_X + v_Y + v_Z) = 0 \rightarrow (B'')^T v_X = (B'')^T v_Y = (B'')^T v_Z = 0 \). Because \( v_X, v_Y, v_Z \) are linearly independent, the nullity of \((B'')^T\) is at least 3, so \( \text{rank}(B'') = \text{rank}((B'')^T) < k-3 \). This contradicts our result that \( \text{rank}(B'') = k-2 \). Thus, \( G'' \) has at most two connected components. Because \( B'' \) has \( k-2 \) non-zero columns, \( G'' \) has \( k-2 \) edges. Since \( G'' \) has \( k \) vertices and no more than two connected components, it must be two disjoint trees.

We return to the first column of \( B \). Without loss of generality let vertices \([1, \ldots, L]\) correspond to one of the disjoint trees in \( G'' \), and \([L+1, \ldots, k]\) the others, with \( 1 \leq L < k-1 \). We want to reason about \( a, b, c \), the three coordinates with \( B_{a1} = B_{b1} = B_{c1} = 1 \). Applying Lemma 33 again to the two disjoint trees gives that there is a vector \( v^{(l)} \in \mathbb{R}^k \) non-zero on exactly the coordinates \( 1, 2, \ldots, L \) with \((B'')^T v^{(l)} = 0\), and a vector \( v^{(r)} \in \mathbb{R}^k \) non-zero on exactly the coordinates \( L+1, \ldots, k \) with \((B'')^T v^{(r)} = 0\). We can rule out that all of \( a, b, c \in [1, \ldots, L] \): Suppose for sake of contradiction this was the case, then observe that \( u = v^{(r)} \) would satisfy \((B^T u)_1 = 1v_a^{(r)} + 1v_b^{(r)} + 1v_c^{(r)} = 0 + 0 + 0 = 0\), and \((B'')^T v^{(r)} = 0\), so \( B^T u = 0 \): u is non-zero but not all of its entries are non-zero, which contradicts \( B \in S_{2k-2,k} \). Symmetrically, we cannot have all of \( a, b, c \in [L+1, \ldots, k] \). Then without loss of generality we have \( a, b \in [1, \ldots, L], c \in [L+1, \ldots, k] \). Write \( G' \) for the subgraph of \( G'' \) induced by vertices \([1, \ldots, L]\): this is a tree containing \( a \) and \( b \). Suppose for sake of contradiction that \( a \) and \( b \) are connected by an odd length path. Then \( v^{(l)}_a = -v^{(r)}_b \). Then let \( u = 1v^{(l)} \): \((B'')^T u = 0\) as before, and \((B^T u)_1 = 1v_a^{(l)} + 1v_b^{(l)} + 1v_c^{(l)} = -v^{(l)}_b + v^{(r)}_b + 0 = 0\). Again we have non-zero \( u \) with not all entries 0 satisfying \( B^T u = 0 \) --- a contradiction. Thus we must have that \( a, b \) are connected by a path of even length in \( G' \).

It follows that \( B \) is the vertex-hyperedge incidence matrix of a forest with two trees connected by a 3-hyperedge where the two vertices of the 3-hyperedge in the same tree are connected by an even length path. Thus \( B \in \mathcal{T}_k^C \).

\[ \square \]

Lemma 37. \( S'_{2k,k} = \mathcal{T}_k^C \).

Proof. We prove this by showing both inclusions in the following two claims.

Claim 17. \( \mathcal{T}_k^C \subset S'_{2k,k} \).

Proof. Suppose \( B \in \mathcal{T}_k^C \). Then \( B \) has \( k \) columns with two 1’s each. Without loss of generality we may let column 1 of \( B \) be an edge in the cycle; then let \( B' \) be \( B \) with the first column set to zero. Then \( B' \in \mathcal{T}_k \), so \( B' \in S_{2k-2,k} \) by Lemma 33. Hence there is a unique (up to constant multiple) non-zero \( v \) with \((B')^T v = 0\), and \( v \) has all non-zero entries. Write the first column of \( B \) as \( e_x + e_y \), for \( 1 \leq x < y \leq k \). We claim that \( e_x + e_y \) is in the span of \( B' \)'s columns. By definition, the edge \((x,y)\) is part of an even length cycle in the graph encoded by \( B \). We encode this cycle. For odd \( n \geq 1 \), there are ordered pairs \((a_i, b_i)\), \( 1 \leq i \leq n \) such that each sum of basis vectors \( e_{a_i} + e_{b_i} \) is a column of \( B' \), \( b_i = a_{i+1} \) for \( 1 \leq n-1 \), and \( a_1 = x \) and \( b_n = y \). Then observe that

\[
(e_{a_1} + e_{b_1}) - (e_{a_2} + e_{b_2}) + \ldots + (-1)^{n+1}(e_{a_n} + e_{b_n}) = e_{a_1} + e_{b_n} = e_x + e_y.
\]

Thus column 1 of \( B \) is in the span of the rest of the columns of \( B \). Thus \( B^T v = 0 \) as well, and because \( B' \) has rank \( k-1 \), \( B \) must have rank \( k-1 \) as well. So \( B \in S_{2k,k} \). Moreover, \( B \in S'_{2k,k} \) because it has \( k \) non-zero columns each with two 1’s.

\[ \square \]

Claim 18. \( S'_{2k,k} \subset \mathcal{T}_k^C \).

Proof. Let \( B \in S'_{2k,k} \). Because \( B \) has rank \( k-1 \), and \( k \) non-zero columns, there is a subset of \( k-1 \) columns of \( B \) with rank \( k-1 \). Then without loss of generality assume the first column of \( B \) is in the span of the other columns, and let \( B' \) be \( B \), but with the first column set to 0. There is a unique (up to constant multiple) non-zero vector \( v \) such that \( B^T v = 0 \), and \( v \) has all non-zero entries. Then by construction \( B'^T v = 0 \). Further because \( B' \) has rank \( k-1 \), so does \( B'^T \), and the rank of the nullspace of \( B'^T \) is 1. Thus \( v \) must be the unique (up to constant multiple) non-zero vector satisfying \( B'^T v = 0 \). So \( B' \in S_{2k-2,k} \). Thus \( B' \in \mathcal{T}_k \).

\[ \square \]
Write the first column of $B$ as $e_x + e_y$, $1 \leq x < y \leq k$. To show $B \in \mathcal{T}^C_k$, we must show that $x$ and $y$ are connected by an odd length path in the tree encoded by $B'$. We know that $e_x + e_y$ is in the span of the columns of $B'$, so we can write
\[
e_x + e_y = \sum_{i=2}^{m} \alpha_i(e_{a_i} + e_{b_i}),
\]
where $m$ is the width of $B'$, and for each $i$ where $\alpha \neq 0$ and the column $B_i$ is non-zero, we have $B_i = e_{a_i} + e_{b_i}$. Let $S = \{i : \alpha_i \neq 0 \land B_i \neq 0\}$. If some $1 \leq j \leq k$ appears in only one pair in $\{(a_i, b_i)\}_{i \in S}$, it must be the case that $\sum_{i=1}^{m} \alpha_i(e_{a_i} + e_{b_i})$ has non-zero $e_j$ coefficient. So $\{(a_i, b_i)\}_{i \in S}$, is a collection of edges of a tree such that at most two vertices $(x, y)$ belong to only one edge. Then $\{(a_i, b_i)\}_{i \in S}$, must be a path, with $x$ and $y$ as the end points. Without loss of generality let $a_2 = x, b_{|S|+1} = y$ and $a_i = b_{i+1}$ for $2 \leq i \leq |S|$, because $e_{b_i} \perp e_x + e_y$, we must have that $\alpha_{i+1} = -\alpha_i$ so that the sum will cancel in coordinate $b_i$. It follows that
\[
e_x + e_y = \sum_{i=2}^{|S|+1} (-1)^i\alpha_2(e_{a_i} + e_{b_i}) = \alpha_2 e_x + (-1)^{|S|+1} \alpha_2 e_y.
\]
It is clear we must choose $\alpha_2 = 1$ and $|S|$ must be odd. Thus $e_x + e_y$ fulfills the conditions of the additional edge forming an even cycle in $\mathcal{T}_k^C$. We thus have $B \in \mathcal{T}_k^C$.

We now prove Lemma 38 which we restate here.

**Lemma 38.** Suppose we have two sets $S$ and $T$ with $S \cap T \neq \emptyset$ where $A_S \in \mathcal{M}_{|S|}$ and $A_T \in \mathcal{M}_{|T|}$. Let $\ell$ be the number of non-zero entries in $A_{S \cup T}$. Then there are at least $\max(|S \cup T| - 1, \frac{\ell}{2})$ non-zero entries in $A_{S \cap T}$ that are not the first (top) non-zero entry in their column.

**Proof.** Let $R := S \cup T$. By Observation 1 since $A_S$ and $A_T$ are minimal dependencies, there cannot be a column in $A_S$ or $A_T$ with a single one. Hence no column of $A_R$ has a single one. It follows immediately that there are at least $\frac{\ell}{2}$ non-zero entries that are not the first non-zero entry in their column in $A_R$.

Next we show that there are at least $|R| - 1$ non-zero entries in $A_R$. Let $G = (R, E)$ be the hypergraph given by the vertex-hyperedge incidence matrix $A_R$. We claim that $G$ must be connected. Indeed, each of the hypergraphs given by the incidence matrix $A_S$ and $A_T$ are connected on their own. Since $S \cap T \neq \emptyset$, the hypergraph given by $A_R$ must be connected. Let $F$ be an arbitrary spanning tree of $G$ whose edges are contained in the hyperedges of $G$. Let $v \in R$ be an arbitrary node in $G$, which we call the root. For each $u \in R \setminus \{v\}$, consider the entry $(u, e(u))$ of $A$, where $e(u)$ is the index of the hyperedge containing the edge from $u$ towards $v$ in $F$. Note that $A_{u, e(u)} = 1$, and further, each hyperedge $e$ in $G$ contains at least one node $x(e)$ (which can be chosen arbitrarily) for which $e(x(e)) \neq e$. Consider $x(e)$ to be the “first” non-zero entry in the column $e$. It follows that the entries $\{(u, e(u))\}_{u \in R \setminus \{x\}}$ are not the “first” in their column. We show this argument pictorially in Figure 3 where we draw an arrow from each $u$ to $e(u)$, and we circle the entry $A_{u, e(u)}$.

\[\square\]

**B Proof of Lemmas 2, 17 and 18**

We prove Lemmas 17 and 18. Lemma 2 is a corollary of Lemma 18.

**Lemma 2.** Let $A \in \mathbb{R}^{n \times m}$, and let $D = \bigcup_{x \in \text{supp}(x)} \text{supp}(x)$. Then for any $i \notin D$ we have $e_i \in \text{Span}(A)$.

**Lemma 17.** Let $A$ be a matrix with columns $A_i$ for $i \in [n]$. Let $H_i$ be the space spanned by the column vectors $A_1, A_2, \cdots A_{i-1}, A_{i+1}, \cdots A_n$. Let $S$ be the set of all $i$ such that $A_i \in H_i$. Then there exists some $y$ with $\text{supp}(y) = S$ such that $Ay = 0$.
Figure 4: Illustration of the proof that there are at least \(|R| - 1\) non-zero entries that are not first in their column. Each hyperedge contains at least one vertex that is not pointed at, otherwise, the arrows would create a cycle.

**Proof.** We prove this by linearly combining the dependencies. Without loss of generality, let \(S = [k]\) for some \(k\). For all \(i \in [k]\), since \(A_i \in H_i\), there exists some \(x^{(i)}\) such that \(x^{(i)}_i \neq 0\) and \(Ax^{(i)} = 0\). Observe that for each \(i \in [k]\), we have \(\text{supp}(x^{(i)}) \subseteq S\) - otherwise it would imply that some column \(A_j\) for \(j \notin S\) is spanned by \(H_j\). Choose random coefficients \(c_i\) for \(i \in [k]\) from any continuous distribution and let \(y = \sum_i c_i x^{(i)}\). Then with probability 1, \(y\) is non-zero on all \(i \in S\).

**Lemma 18.** With the terminology of the previous lemma, \(e_i \in \text{Span}(A^T)\), if and only if \(A_i \notin H_i\).

**Proof.** For the first part, let \(v\) be such that \(A^Tv = e_i\). Then for any \(w\) with \(i \in \text{supp}(w)\), we have \(w^TA^Tv = w_i \neq 0\), so it is impossible that \(Aw = 0\).

For the converse, suppose \(A_i \notin H_i\). Then there must exist some \(w_i\) such that \(\langle w_i, A_i \rangle \neq 0\), but \(w_i \perp A_j\) for all \(j \neq i\). However, this implies that \(Aw_i = \langle w_i, A_i \rangle e_i\), so \(e_i \in \text{Span}(A^T)\), which is a contradiction.

**C Proofs of Lemmas 8 and 9, and 10 on bounds related to Binomial distributions**

**Lemma 8** (Tail Bound on Binomial). If \(t \geq 2np\), then

\[
\Pr[\text{Bin}(n, p) \geq t] \leq 2 \left(\frac{enp}{t}\right)^t. \tag{13}
\]
Proof.

\[
\Pr[\text{Bin}(n, p) \geq \ell] \leq \sum_{i=\ell}^{n} \binom{n}{i} p^i
\]

\[
\leq \sum_{i=\ell}^{n} \left(\frac{n}{i} \right) p^i \prod_{j=\ell+1}^{i} \left(\frac{n-j+1}{j} \right)
\]

\[
\leq \sum_{i=\ell}^{n} \left(\frac{n}{i} \right) p^i \prod_{j=\ell+1}^{i} \left(\frac{n}{\ell} \right)
\]

\[
= \sum_{i=\ell}^{n} \left(\frac{n}{i} \right) p^i \left(\frac{\ell}{n} \right)^{i-\ell}
\]

\[
\leq \left(\frac{n}{\ell} \right) p^\ell \sum_{j=0}^{\infty} \left(\frac{\ell}{n} \right)^j
\]

\[
= \left(\frac{n}{\ell} \right) p^\ell \frac{1}{1 - \frac{\ell}{n}}.
\]

For \( t \geq 2np \), plugging in Sterling’s formula yields the lemma.

\[\square\]

**Lemma 9 (Small Case Binomial Calculation).** For constants \( \gamma, d > 0 \), for \( k \leq \frac{n}{8e^4 \gamma d^2} \), there exists a constant \( c_9 \) such that for any \( j \in \{k - 1, k, k + 1\} \) and \( \gamma \geq 1/2 \), we have

\[
\sum_{\ell \geq 1}^\infty \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \max \left( j, \frac{\ell}{2} \right) \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell] \leq e^{-\gamma dk + 4k \log(\gamma d)} \left( \frac{k}{n} \right)^j.
\]

**Proof.** We break down this sum as follows.

\[
\sum_{\ell \geq 1}^\infty \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \max \left( j, \frac{\ell}{2} \right) \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell]
\]

\[
\leq 2^j \sum_{\ell = j}^\infty \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq j \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell]
\]

\[
+ \sum_{\ell \geq 2j+1}^{n/3} \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \frac{\ell}{2} \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell]
\]

\[
+ \sum_{\ell \geq n/3}^\infty \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \frac{\ell}{2} \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell]
\]

We bound the first term in the following claim.

**Claim 19.**

\[
\sum_{\ell = j}^{2j} \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq j \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell] \leq 8k \left( \frac{20e^k}{\gamma n} \right)^j \left( 2e \gamma d \right)^{4k} e^{-\gamma dk}.
\]
Proof.

\[
\sum_{\ell=j}^{2j} \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq j \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell] \\
\leq 2j \Pr \left[ \text{Bin} \left( 2j, \frac{2j + k}{\gamma n} \right) \geq j \right] \Pr[\text{Bin}(\gamma nk, d/n) \leq 2j] \\
\leq 4j \left( \frac{e^{2j(2j + k)}}{\gamma n^j} \right)^j \Pr[\text{Bin}(\gamma nk, d/n) \leq 2j] \\
\leq 8k \left( \frac{20e k}{\gamma n} \right)^j \Pr[\text{Bin}(\gamma nk, d/n) \leq 2j].
\]

(97)

Here the first inequality follows from the fact that the summand is highest for \( \ell = 2j \), the second inequality follows from the tail bound in Lemma 8 in Section C, and then third inequality follows from the fact that \( j \leq k + 1 \leq 2k \).

Now

\[
\Pr[\text{Bin}(\gamma nk, d/n) \leq 2j] \leq \Pr[\text{Bin}(\gamma nk, d/n) \leq 2j] \\
\leq \left( \frac{\gamma nk}{2j} \right) \left( \frac{d}{n} \right)^{2j} \left( 1 - \frac{d}{n} \right)^{\gamma nk - 2j} \\
\leq \left( \frac{e^{\gamma nk}}{2j} \right)^{2j} \left( \frac{d}{n - d} \right)^{2j} \left( 1 - \frac{d}{n} \right)^{\gamma nk} \\
\leq \left( \frac{e^{\gamma dk}}{j} \right)^{2j} e^{-\gamma dk} \\
\leq (2e^{\gamma d})^k e^{-\gamma dk}
\]

(98)

Combining this with Equation (97) yields the claim.

We bound the second term in Equation (95) in the following claim.

Claim 20.

\[
\sum_{\ell \geq 2j+1}^{n/3} \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \frac{\ell}{2} \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell] \leq 4e^{\gamma dk} \left( \frac{8e^3 \gamma d^2 k}{n} \right)^{j+1}.
\]

(99)

Proof. For \( \ell \leq n/3 \), using Lemma 8 andSterling’s formla, we have

\[
\Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \frac{\ell}{2} \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell] \\
\leq 2 \left( \frac{e^{\ell(\ell + k)}}{\gamma n^{\ell/2}} \right) \Pr[\text{Bin}(\gamma nk, d/n) = \ell] \\
= 2 \left( \frac{e^{\ell(\ell + k)}}{\gamma n^{\ell/2}} \right) \Pr[\text{Bin}(\gamma nk, d/n) = \ell] \\
= 2 \left( \frac{2e^{\ell(\ell + k)}}{\gamma n^{\ell/2}} \right) \Pr[\text{Bin}(\gamma nk, d/n) = \ell] \\
\leq 2 \left( \frac{2e^{\ell(\ell + k)}}{\gamma n^{\ell/2}} \right) \left( \frac{e^{\gamma nk}}{\ell} \right) \left( \frac{d}{n - d} \right)^{\ell} \left( 1 - \frac{d}{n} \right)^{\gamma nk - \ell} \\
\leq 2 \left( \frac{2e^{\ell(\ell + k)}}{\gamma n^{\ell/2}} \right) \left( \frac{e^{\gamma nk}}{\ell} \right) \left( \frac{d}{n - d} \right)^{\ell} e^{-\gamma dk} \\
\leq 2 \left( \frac{4e^{\ell(\ell + k)}}{\gamma n^{\ell/2}} \right) \left( \frac{2e^{\gamma dk}}{\ell} \right)^{\ell} e^{-\gamma dk}.
\]

(100)
We do casework on the parity of $\ell$. Let $\ell = 2a + b$, where $b \in \{0, 1\}$. If $b = 1$, then

$$
\left( \frac{4e\ell}{\gamma n} \right)^{\lfloor \ell/2 \rfloor} \left( \frac{2e\gamma dk}{\ell} \right)^\ell = \left( \frac{4e(2a + 1)}{\gamma n} \right)^{a+1} \left( \frac{2e\gamma dk}{2a + 1} \right)^{2a+1}
$$

$$
= \left( \frac{16e^3d^2k^2(2a + 1)}{\gamma n(2a + 1)^2} \right)^a \left( \frac{8e^2\gamma dk(2a + 1)}{\gamma n(2a + 1)} \right)
$$

$$
\leq \left( \frac{8e^3\gamma d^2k^2}{na} \right)^a \left( \frac{8e^2dk}{n} \right)^{a+1}.
$$

(101)

Now since the maximum over $x$ of $f(x) = \left( \frac{y}{x} \right)^x$ is achieved at $x = y/e$, and above this value of $x$, the $f(x)$ is decreasing, since $j \geq k - 1 \geq \frac{e}{2\gamma d^2k^2}$, we have for all $a \geq j$,

$$
\left( \frac{8e^3\gamma d^2k^2}{na} \right)^a \left( \frac{8e^2dk}{n} \right) \leq \left( \frac{8e^3\gamma d^2k^2}{n} \right)^a \left( \frac{8e^2dk}{n} \right)
$$

$$
\leq \left( \frac{8e^3\gamma d^2k^2}{n(j+1)} \right)^a \left( \frac{8e^2dk}{n} \right)
$$

$$
\leq \left( \frac{8e^3\gamma d^2k^2}{n} \right)^{a+1}.
$$

(102)

If $b = 0$, then

$$
\left( \frac{4e\ell}{\gamma n} \right)^{\lfloor \ell/2 \rfloor} \left( \frac{2e\gamma dk}{\ell} \right)^\ell = \left( \frac{8ea}{\gamma n} \right)^a \left( \frac{2e\gamma dk}{2a} \right)^{2a}
$$

$$
= \left( \frac{8e^3\gamma d^2k^2}{na} \right)^a.
$$

(103)

By the same reasoning as before, we have for all $a \geq j + 1$,

$$
\left( \frac{8e^3\gamma d^2k^2}{na} \right)^a \leq \left( \frac{8e^3\gamma d^2k^2}{n(j+1)} \right)^a \leq \left( \frac{8e^3\gamma d^2k^2}{n} \right)^a.
$$

(104)

Combining these two cases back into Equation 100, we have for all $\ell \geq 2j + 1$,

$$
\Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \frac{\ell}{2} \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell] \leq 2e^{-\gamma dk} \left( \frac{8e^3\gamma d^2k^2}{n} \right)^{\lfloor \frac{\ell}{2} \rfloor}.
$$

(105)

Summing over all $\ell \geq 2j + 1$, we have

$$
\sum_{\ell \geq 2j+1} \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \frac{\ell}{2} \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell] \leq 4e^{-\gamma dk} \left( \frac{8e^3\gamma d^2k^2}{n} \right)^{j+1}.
$$

(106)

Finally, we bound the third term in Equation 95 in the following claim.

Claim 21.

$$
\sum_{\ell \geq n/3} \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \frac{\ell}{2} \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell] \leq 2 \left( \frac{3e\gamma dk}{n} \right)^{n/3}.
$$

(107)

Proof. It suffices to bound the probability

$$
\Pr[\text{Bin}(\gamma nk, d/n) \geq n/3].
$$

Again employing Lemma 8 we have

$$
\Pr[\text{Bin}(\gamma nk, d/n) \geq n/3] \leq 2 \left( \frac{3e\gamma dk}{n} \right)^{n/3}.
$$

(108)
Combining claims \[19, 20 \text{ and } 21\], we have
\[
\sum_{\ell \geq 1} \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \max \left( j, \frac{\ell}{2} \right) \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell] \\
\leq 8k \left( \frac{20ek}{\gamma n} \right)^j (2e\gamma d)^{4k} e^{-\gamma dk} \\
+ 4e^{-\gamma dk} \left( \frac{8e^3d^2k}{n} \right)^{j+1} \\
+ 2 \left( \frac{3e\gamma dk}{n} \right)^{n/3}.
\] (109)

It is easy to check that this sum is dominated by the first term, and hence for some universal constant \(c\),
\[
\sum_{\ell \geq 1} \Pr \left[ \text{Bin} \left( \ell, \frac{\ell + k}{\gamma n} \right) \geq \max \left( j, \ell \right) \right] \Pr[\text{Bin}(\gamma nk, d/n) = \ell] \\
\leq 16k \left( \frac{20ek}{\gamma n} \right)^j (2e\gamma d)^{4k} e^{-\gamma dk} \\
\leq \left( \frac{k}{n} \right)^j e^{-\gamma dk + e^{k\log(\gamma d)}}.
\] (110)

This proves the lemma.

\[\square\]

**Lemma 10 (Large Case Binomial Calculation).** There exists constants \(C, d_0\), and \(c\) such that for all \(d \geq d_0\), for any positive integer \(\frac{d}{n} \leq k \leq n/C\), we have
\[
\sum_{k=\frac{n}{2C}}^{\frac{n}{C}} \binom{n}{k} \Pr \left[ \text{Bin} \left( n - k - 1, 1 - \frac{1}{\sqrt{kd/n}} \right) \leq k \right] \leq e^{-\Theta(n)}.
\] (15)

**Proof.** Because \(n/2 < n - k - 1\), we may bound the probability by:
\[
\Pr \left[ \text{Bin} \left( n - k - 1, 1 - \frac{1}{\sqrt{kd/n}} \right) < k \right] \leq \Pr \left[ \text{Bin} \left( \frac{n}{2}, 1 - \frac{1}{\sqrt{kd/n}} \right) < k \right].
\] (111)

We use the Chernoff bound \(\Pr[X \leq (1 - \delta)\mu] \leq e^{-\mu\delta^2/2}\) for \(\mu = \text{E}[X]\), plugging in \(\mu = \frac{n}{2} \left( 1 - \frac{1}{\sqrt{kd/n}} \right)\) and \(\delta = (1 - \frac{1}{\sqrt{kd/n}})^{-1} \left( 1 - \frac{1}{\sqrt{kd/n}} - \frac{2k}{n} \right)\). This gives
\[
\Pr \left[ \text{Bin} \left( n - k - 1, 1 - \frac{1}{\sqrt{kd/n}} \right) < k \right] \leq \Pr \left[ \text{Bin} \left( \frac{n}{2}, 1 - \frac{1}{\sqrt{kd/n}} \right) < k \right] \\
\leq e^{-\frac{n}{4} \left( 1 - \frac{1}{\sqrt{kd/n}} \right) \left( 1 - \frac{1}{\sqrt{kd/n}} - \frac{2k}{n} \right)^2} \\
\leq e^{-n\epsilon},
\] (112)

where \(\epsilon \geq 1/36\). To achieve this value of \(\epsilon\), we plugged in \(k < \frac{n}{12}\) in the last inequality.

Now we compute the sum over \(k\):
\[
\sum_{k=\frac{n}{2C}}^{\frac{n}{C}} \binom{n}{k} e^{-\epsilon n} \leq n \left( \frac{n}{n/C} \right) e^{-\epsilon n} \leq n(eC)^{n/C} e^{-\epsilon n} = e^{n \left( \log n \frac{1}{n} + \frac{1 + \log C}{2} \right)} e^{-\epsilon n}
\] (113)

which for constant \(C\) large enough, is \(e^{-\Theta(n)}\).
In the following Appendix, we will prove the claims stated in the proof of Lemma 24. Before proving these claims, we will need the following additional lemma: The following bound gives a general approximation for the probability mass function of a HyperGeometric Distribution. It also provides a second bound under the additional assumption that the number of draws in our distribution is not too large relative to the population size.

**Lemma 39.** Let \( X \sim \text{HyperGeom}(A, B, n) \). Furthermore, let us define \((1 - q) = \frac{B}{A}\). Then

\[
\Pr[X = k] \leq \binom{n}{k}(1 - q)^k \left( q + \frac{k}{A - n} \right)^{n-k}.
\]

Furthermore, assuming \( n \leq \frac{3}{2}qA \) and \( q \leq 1/2 \), we have:

\[
\Pr[X = k] \leq \left( \frac{en}{k} \right)^k (1 - q)^k q^{n-k} \left( e^{(6e)k} \right) \leq \left( \frac{en}{k} \right)^k (1 - q)^k q^{n-k} \left( e^{(6e)k} \right)
\]

\[
\leq \left( \frac{en}{k} \right)^k (1 - q)^k q^{n-k} \left( e^{(6e)k} \right)
\]

\[
= (114)
\]

**Proof.** Recall that, by definition of the hypergeometric distribution, we have:

\[
\Pr[X = k] = \frac{\binom{B}{k}(A - B)}{\binom{n}{k}}
\]

We can expand out the binomial terms into their factorial representations to see:

\[
\Pr[X = k] = \frac{B!}{(B-k)!} \cdot \frac{(A-B)!}{(A-B-n+k)!(n-k)!} \cdot \frac{(A-n)!}{A!}
\]

\[
= \binom{n}{k} \left( \binom{B}{k} \cdot \frac{(A-B)!}{(A-B-n+k)!(n-k)!} \cdot \frac{(A-n)!}{A!} \right)
\]

\[
= \binom{n}{k} \prod_{i=1}^{k} \left( \binom{B}{k} \cdot \frac{(A-B)!}{(A-B-n+k)!(n-k)!} \cdot \frac{(A-n)!}{A!} \right)
\]

\[
= \binom{n}{k} \prod_{i=1}^{k} \left( \binom{B}{k} \cdot \frac{(A-B)!}{(A-B-n+k)!(n-k)!} \cdot \frac{(A-n)!}{A!} \right)
\]

\[
= \binom{n}{k} \prod_{i=1}^{k} \left( \binom{B}{k} \cdot \frac{(A-B)!}{(A-B-n+k)!(n-k)!} \cdot \frac{(A-n)!}{A!} \right)
\]

\[
\leq \binom{n}{k} \prod_{i=1}^{k} \left( \frac{B}{A - k + i} \cdot \frac{n-k}{A - n + i} \right)
\]

\[
\leq \binom{n}{k} \prod_{i=1}^{k} \left( 1 - \frac{B}{A - n + i} + \frac{k}{A - n + i} \right)
\]

By definition of the hypergeometric distribution, \( k \leq B \). Thus, it follows \( k - i < B < A \) for all \( i \in [k] \). This implies:

\[
\frac{B - k + i}{A - k + i} \leq \frac{B}{A} = (1 - q)
\]

This gives us the first inequality of the lemma:

\[
P[X = k] \leq \binom{n}{k} \prod_{i=1}^{k} \left( 1 - \frac{B}{A - n + i} + \frac{k}{A - n + i} \right)
\]

\[
= \binom{n}{k} \prod_{i=1}^{k} \left( q + \frac{k}{A - n} \right)^{n-k}
\]

\[
= \binom{n}{k} (1 - q)^k \left( q + \frac{k}{A - n} \right)^{n-k}
\]
We will now proceed under the assumption that \( n \leq \frac{3}{2} q A \) to achieve the second bound. We write:

\[
(q + \frac{k}{A - n})^{n-k} \leq (q + \frac{k}{(1 - \frac{3}{2}q)A})^{n-k}
\]

\[
= \sum_{i=0}^{n-k} \binom{n-k}{i} q^{n-k-i} \left( \frac{k}{(1 - \frac{3}{2}q)A} \right)^i
\]

\[
\leq q^{n-k} \sum_{i=0}^{n-k} \left( \frac{\frac{3}{2}eqA}{i} \right)^i \left( \frac{1}{q} \right)^i \left( \frac{k}{(1 - \frac{3}{2}q)A} \right)^i
\]

\[
\leq q^{n-k} \sum_{i=0}^{n-k} \left( \frac{\frac{3}{2}ek}{i(1 - \frac{3}{2}q)} \right)^i
\]

\[
\leq q^{n-k} \sum_{i=0}^{n-k} \left( \frac{1}{i} \right) \left( \frac{\frac{3}{2}ek}{1 - \frac{3}{2}q} \right)^i
\]

\[
\leq q^{n-k} \left( e^{\frac{3}{2}ek} \right)
\]

\[
\leq q^{n-k} \left( e^{\frac{6}{3}ek} \right)
\]

Applying this to the first inequality gives:

\[
P[X = k] \leq \binom{n}{k} (1 - q)^k q^{n-k} e^{(6ek)}
\]

as claimed.

We are now ready to prove claims 5 and 6.

**Claim 5.** Let \( p < \frac{1}{2} \). Let \( K \leq \frac{3}{2}pN \). There exists constants \( c_5 \) and \( d_0 \) such that for all \( \gamma > 1 \) and \( d > d_0 \), the following two bounds hold.

For \( \ell \in \{2K - 2, 2K - 1, 2K\} \), and \( j \leq \left\lfloor \frac{\ell}{2} \right\rfloor \), we have:

\[
\Pr \left[ \text{Bin} \left( \ell, 1 - \frac{\ell}{\gamma(1-p)N} \right) \leq j \right] \cdot \Pr[\text{HyperGeom}(\gamma dN, \gamma d(1-p)N, \gamma dK) = \ell] \leq \left( p^{\gamma d - \log(\gamma d)} \right)^K \left( \frac{K}{N} \right)^{\ell-j}.
\]

Further,

\[
\sum_{\ell=1}^{4K} \Pr \left[ \text{Bin} \left( \ell, \frac{4K}{\gamma(1-p)N} \right) \geq K - 1 \right] \cdot \Pr[\text{HyperGeom}(\gamma dN, \gamma d(1-p)N, \gamma dK) = \ell] \leq \left( p^{\gamma d - \log(\gamma d)} \right)^K \left( \frac{K}{N} \right)^{K-1}.
\]

**Proof.** By expanding the binomial distribution, we have,

\[
\Pr \left[ \text{Bin} \left( \ell, 1 - \frac{\ell}{\gamma(1-p)N} \right) = j \right] \leq \sum_{i=0}^{j} \binom{j}{i} \left( \frac{\ell}{\gamma(1-p)N} \right)^i \left( 1 - \frac{\ell}{\gamma(1-p)N} \right)^{\ell-i}
\]

\[
\leq 2^\ell \left( \frac{\ell}{\gamma(1-p)N} \right)^{\ell-j}.
\]
Using the fact that $K \leq \frac{3}{2}N$, we employ the second bound of Lemma 39 to find:

$$\Pr \left[ \text{Bin} \left( \ell, 1 - \frac{\ell}{\gamma(1 - p)N} \right) = j \right] \cdot \Pr[\text{HyperGeom}(\gamma dN, \gamma d(1 - p)N, \gamma dK) = \ell]$$

$$\leq \left( \frac{\gamma dK}{\ell} \right) (1 - p)^{\ell - j} p^{\gamma dK - j e6\ell} 2^{\ell} \left( \frac{\ell}{c(1 - p)N} \right)^{\ell - j}$$

$$\leq \frac{\gamma dK}{2K} p^{\gamma dK - 2K} e^{12eK} 2^{2K} \left( \frac{2K}{\gamma N} \right)^{\ell - j}$$

$$\leq \frac{9e^{12K + 2c^2d^2}}{4} \left( \frac{\gamma dK}{\gamma N} \right) K^\ell - j$$

$$\leq \left( p^{\gamma d + c \log(\gamma d)} \right)^K \left( \frac{K}{N} \right)^{\ell - j}$$

for some universal constant $c$.

Next we show the second statement in the claim. Since for sufficiently large $\gamma d$, the mean of $\text{HyperGeom}(\gamma dN, \gamma d(1 - p)N, \gamma dK)$ is greater than $4K$, the term

$$\Pr[\text{Bin}(\ell, \frac{K}{N}) \geq K - 1] \cdot \Pr[\text{HyperGeom}(\gamma dN, \gamma d(1 - p)N, \gamma dK) = \ell]$$

is maximized (over $\ell \in [1, 4K]$) at $\ell = 4K$. By Equation 115 we have we have

$$\sum_{\ell=1}^{4K} \Pr \left[ \text{Bin} \left( \ell, \frac{K}{N} \right) \geq K - 1 \right] \cdot \Pr[\text{HyperGeom}(\gamma dN, \gamma d(1 - p)N, \gamma dK) = \ell]$$

$$\leq 4K \Pr[\text{HyperGeom}(\gamma dN, \gamma d(1 - p)N, \gamma dK) = 4K] 2^{4K} \left( \frac{K}{\gamma(1 - p)N} \right)^{K - 1}$$

$$\leq 4K \Pr[\text{HyperGeom}(\gamma dN, \gamma d(1 - p)N, \gamma dK) = 4K] 2^{5K} \left( \frac{K}{N} \right)^{K - 1}$$

Using the fact that $K \leq \frac{3}{2}N$, we employ the second bound of Lemma 39 to find:

$$\sum_{\ell=1}^{4K} \Pr \left[ \text{Bin} \left( \ell, \frac{4K}{\gamma(1 - p)N} \right) \geq K - 1 \right] \cdot \Pr[\text{HyperGeom}(\gamma dN, \gamma d(1 - p)N, \gamma dK) = \ell]$$

$$\leq 4K \left( \frac{\gamma dK}{4K} p^{\gamma dK - 4K} e^{24eK} \right) 2^{5K} \left( \frac{4K}{N} \right)^{K - 1}$$

$$\leq 4K \left( \frac{e^d}{4} \right) p^{\gamma dK - 4K} e^{24eK} 2^{5K} \left( \frac{K}{N} \right)^{K - 1}$$

$$\leq \left( p^{\gamma d - c' \log(\gamma d)} \right)^K \left( \frac{K}{N} \right)^{K - 1}$$

for some constant $c'$. The claim follows from choosing $c'$ larger than $c$ and $c'$.

We now prove Claim 6.

**Claim 6.** Let $p < 1/2$. There exists $\gamma_0$ and $d_0$ such that for all $\gamma > \gamma_0$ and $d > d_0$, if $k \leq \frac{n}{38e^2d^2}$, we have:

$$\sum_{\ell=2k+1}^{\gamma dk} \Pr \left[ \text{Bin}(\ell, 1 - \frac{\ell}{\gamma n}) \leq \left\lfloor \frac{\ell}{2} \right\rfloor \right] \cdot \Pr[\text{HyperGeom}(\gamma dn, \gamma d(1 - p)n, \gamma dk) = \ell] \leq e^{-k} \left( \frac{k}{n} \right)^{k+1/2}$$

53
Proof. We write:

\[
\sum_{\ell = 2k+1}^{\gamma dk} \Pr[\text{Bin}(\ell, 1 - \frac{\ell}{\gamma n}) \leq \lceil \frac{\ell}{2} \rceil] \cdot \Pr[\text{HyperGeom}(\gamma dn, \gamma d(1-p)n, \gamma dk) = \ell]
\]

\[
\leq \sum_{\ell = 2k+1}^{\gamma dk} \Pr[\text{HyperGeom}(\gamma dn, \gamma d(1-p)n, \gamma dk) = \ell] \sum_{i=0}^{\left\lfloor \frac{\ell}{2} \right\rfloor} \binom{\ell}{i} \left( \frac{\ell}{\gamma n} \right)^{\ell-i} \left( 1 - \frac{\ell}{\gamma (1-p)n} \right)^{\ell}
\]

\[
\leq \sum_{\ell = 2k+1}^{\gamma dk} \Pr[\text{HyperGeom}(\gamma dn, \gamma d(1-p)n, \gamma dk) = \ell] \gamma dk \left( \frac{\ell}{\gamma (1-p)n} \right)^{\ell/2}
\]

Applying the first bound of Lemma 39, we find:

\[
\sum_{\ell = 2k+1}^{\gamma dk} \Pr[\text{Bin}(\ell, 1 - \frac{\ell}{\gamma n}) \leq \lceil \frac{\ell}{2} \rceil] \cdot \Pr[\text{HyperGeom}(\gamma dn, \gamma d(1-p)n, \gamma dk) = \ell]
\]

\[
\leq \gamma dk \sum_{\ell = 2k+1}^{\gamma dk} \left( \frac{6\gamma dk}{\sqrt{n}} \right)^{\ell/2}
\]

\[
\leq (0.625)^{-\gamma dk} \sum_{\ell = 2k+1}^{\gamma dk} \left( \frac{6\gamma dk}{\sqrt{n}} \right)^{\ell/2}
\]

To show the function in the expression above is maximized at \(2k + 1\) when \(k \leq \frac{n}{18\gamma d^2}\), we consider the derivative. Taking \(C = \frac{6\gamma dk}{\sqrt{n}}\), we write:

\[
\frac{d}{d\ell} \left[ \left( \frac{C}{\sqrt{\ell}} \right)^{\ell} \right] = \frac{d}{d\ell} \left[ e^{\ell \ln \left( \frac{C}{\sqrt{\ell}} \right)} \right]
\]

\[
= e^{\ell \ln \left( \frac{C}{\sqrt{\ell}} \right)} \frac{d}{d\ell} \left[ \ell \ln \left( \frac{C}{\sqrt{\ell}} \right) \right]
\]

\[
= e^{\ell \ln \left( \frac{C}{\sqrt{\ell}} \right)} \left[ \ln \left( \frac{C}{\sqrt{\ell}} \right) + \ell \frac{d}{d\ell} \left[ \ln \left( \frac{C}{\sqrt{\ell}} \right) \right] \right]
\]

\[
= e^{\ell \ln \left( \frac{C}{\sqrt{\ell}} \right)} \left[ \ln \left( \frac{C}{\sqrt{\ell}} \right) + \ell \frac{1}{\sqrt{\ell}} \frac{d}{d\ell} \left[ \left( \frac{C}{\sqrt{\ell}} \right) \right] \right]
\]

\[
= e^{\ell \ln \left( \frac{C}{\sqrt{\ell}} \right)} \left[ \ln \left( \frac{C}{\sqrt{\ell}} \right) + \ell^{1.5} \frac{d}{d\ell} \ell^{-0.5} \right]
\]

\[
= e^{\ell \ln \left( \frac{C}{\sqrt{\ell}} \right)} \left[ \ln \left( \frac{C}{\sqrt{\ell}} \right) - \frac{1}{2} \right]
\]

\[
= \frac{1}{2} e^{\ell \ln \left( \frac{C}{\sqrt{\ell}} \right)} \left[ \ln \left( \frac{C^2}{\ell} \right) - 1 \right]
\]

\[
= \frac{1}{2} e^{\ell \ln \left( \frac{C}{\sqrt{\ell}} \right)} \left[ \ln \left( \frac{C^2}{\ell} \right) \right].
\]

Thus, we see that the derivative is zero when \(\ell = \frac{C^2}{\ell} = \frac{36\gamma dk^2}{n}\) and strictly negative afterwards. For \(k \leq \frac{n}{18\gamma d^2}\), this maximum occurs before \(\ell = 2k + 1\). As the function is monotonically decreasing after the
Observation 3. For any matrix $A$, $\text{Cov}(\gamma, \gamma, \gamma)$, where all big-O notation is in terms of $\gamma$ for all $\gamma$, is at least $\gamma$.

Let $S = \{1, \ldots, N\}$ be a family of sets of $\{0,1\}$-random variables such $E[I_i^{(N)}] = \Theta(1)$ for all $i$, and $\text{Cov}(I_i^{(N)}, I_j^{(N)}) = O(1/N)$ for all $i \neq j$. Let $X^{(N)} = \sum_{i=1}^{N} I_i^{(N)}$. Then with probability $1 - o(1)$,

$$X^{(N)} \geq (1 - o(1))E[X^{(N)}],$$

where all big-O notation is in terms of $N \to \infty$.

Proof. $\text{Cov}(I_i^{(N)}, I_j^{(N)}) = O(1/N)$ for $i \neq j$, and $\text{Var}(I_i^{(N)}) = O(1)$ because $I_i^{(N)}$ is bounded, so $\text{Var}(X^{(N)}) = \sum_{i=1}^{N} \text{Var}(I_i^{(N)}) + \sum_{1 \leq i \neq j \leq N} \text{Cov}(I_i^{(N)}, I_j^{(N)}) = O(N)$. By Chebyshev’s inequality,

$$\Pr\left[|X^{(N)} - E[X^{(N)}]| > N^{1/3} \sqrt{\text{Var}(X^{N})}\right] < N^{-1/3}$$

Because $\text{Var}(X^{(N)}) = O(N)$, $N^{1/3} \sqrt{\text{Var}(X^{N})} = o(N)$. Thus, we have with probability $1 - o(1)$ that $X^{(N)} \geq E[X^{(N)}] - o(N)$. Because $E[X^{(N)}] = \sum_{i=1}^{N} E[I_i^{(N)}] = \Theta(N)$, we equivalently assert that with probability $1 - o(1)$, $X^{(N)} \geq (1 - o(1))E[X^{(N)}]$. □
Lemma 42. Let $A \sim SB(n, d)$. With probability $1 - o(1)$, the number of rows that are all zeros in $A$ is at least $(1 - o(1))e^{-d}n$.

Proof. Define $I_i$ as the indicator variable of the event that row $i$ is all zero. We have $E[I_i] = (1 - d/n)^n$ for all $i$, but the $I_i$ are not independent. Let $X = \sum_{i=1}^N I_i$. We compute $\text{Var}(X)$.

$$\text{Var}(X) = \sum_{i,j=1}^n \text{Cov}(I_i, I_j) = \sum_{i,j=1}^n E[I_iI_j] - E[I_i]E[I_j]$$

Note that $E[I_iI_j]$ is the probability of the event that both row $i$ and $j$ are 0. When $i \neq j$, rows $i$ and $j$ have identical entry $A_{ij} = A_{ji}$, but all other entries are independent. Thus the probability of both rows $i, j$ having all zero entries is $(1 - d/n)^{2n-1}$. Thus,

$$\text{Cov}(I_i, I_j) = (1 - d/n)^{2n-1} - (1 - d/n)^{2n}$$

$$= \left(\frac{1}{1 - d/n} - 1\right) (1 - d/n)^{2n}$$

$$= \frac{d/n}{1 - d/n} (1 - d/n)^{2n} = O(1/n)$$

Then from Lemma 41, with probability $1 - o(1)$, we have $E[X] \geq (1 - o(1))n(1 - d/n)^n = (1 - o(1))ne^{-d}$. \qed

Lemma 43. Let $A \sim ABC(n, \gamma, d)$. With probability $1 - o(1)$, the number of rows that are all zeros in $A$ is at least $(1 - o(1))p^\gamma d n$.

Proof. Define $I_i$ as the indicator variable of the event that row $i$ is all zero. Because we are interested in upper bounding the variance, we want to upper bound $E[I_iI_j] - E[I_i]E[I_j]$.

A row $i$ of $A \sim ABC_p(n, \gamma, d)$ is zero only if all of the $\gamma d$ 1’s in row $i$ of $A_0$ fell into the $p\gamma n$ last columns. In the setting of the configuration model, this means all of row $i$’s $\gamma d$ half edges are matched to the $p(\gamma dn)$ half edges corresponding to the dropped columns. So

$$\text{Pr}[E_i] = \frac{(p(\gamma dn))^{\gamma d}}{(\gamma d)}$$

Let $j \neq i$. Observe that $E[I_j|I_i]$ can be computed similarly, but the conditioning implies that there are already $cd$ half edges from row $i$ assigned all to the $p(\gamma dn)$ dropped column half-edges.

$$\text{Pr}[E_j|E_i] = \frac{(p(\gamma dn)-cd)^{\gamma d-\ell}}{(\gamma d-\ell)}$$

Observe that if two pairs of functions $a_i(n), b_i(n), i = 1, 2$ satisfy $a_i(n), b_i(n) = O(1)$ and $a_i(n) - b_i(n) = O(1/n)$, then $a_1(n)a_2(n) = (b_1(n) + O(1/n))(b_2(n) + O(1/n)) = b_1(n)b_2(n) + O(1/n)O(1/n) \rightarrow a_1(n)a_2(n) - b_1(n)b_2(n) = O(1/n)$. This can be applied iteratively to any constant number of functions. Observe that

$$\text{Pr}[E_i] - \text{Pr}[E_j|E_i] = \sum_{\ell=0}^{\gamma d-1} \frac{p(\gamma dn) - \ell}{\gamma dn - \ell} - \sum_{\ell=0}^{\gamma d-1} \frac{p(\gamma dn) - \gamma d - \ell}{\gamma dn - \gamma d - \ell}$$

Note that the terms corresponding to a fixed $\ell$ in each product have difference that is $O(1/n)$:

$$\frac{p(\gamma dn) - \ell}{\gamma dn - \ell} - \frac{p(\gamma dn) - cd - \ell}{\gamma dn - cd - \ell} = \frac{O(n)}{(\gamma dn - \ell)(\gamma dn - cd - \ell)} = O(1/n).$$

It follows from the observation (and that $\gamma d$ is a constant) that $\text{Pr}[E_i] - \text{Pr}[E_j|E_i] = O(1/n)$. It follows that

$$\text{Cov}(I_i, I_j) = E[I_iI_j] - E[I_i]E[I_j] = E[I_iI_j] - E[I_i]^2 = (\text{Pr}[E_i])(\text{Pr}[E_i] - \text{Pr}[E_j|E_i]) = O(1/n)$$

Then by Lemma 41 we have with probability $1 - o(1)$ that $X \geq (1 - o(1))E[X] = (1 - o(1))p^\gamma d n$. \qed

56
F Implication of Theorem 2 on exact rank of random matrices.

Define the 2-core of a matrix $A$ to be the matrix remaining after repeating the following peeling process: If there is a row with zero or one non-zero entries, remove that row, and remove the column corresponding to the position of the non-zero entry (if any). Note that this 2-core corresponds to viewing $A$ as the vertex-hyperedge incidence matrix of a hypergraph. Let $n^*$ and $m^*$ be the number of rows (vertices) and columns (hyperedges) respectively in the 2-core.

We prove the following corollary to Theorem 2.

Corollary 1. Let $A \sim SB(n, d)$ with $d = \omega(1)$. With probability $1 - o(1)$,

$$\text{rank}(A) = n^* + n - m^*$$

(116)

Proof. Let $D$ be the set of rows involved in linear dependencies:

$$D = \bigcup_{x : x^T A = 0} \text{supp}(x).$$

Let $I := [n] \setminus D$ be the remaining rows.

Let $P_R$ be the set of rows removed during the peeling process, and let $P_C$ be the set of columns removed. Let $A' \in \mathbb{R}^{n^* \times m^*}$ be the 2-core matrix which remains after the peeling process.

The following claim follows immediately from the definition of the peeling process:

Claim 22. The row-span of $A_{P_R}$ is $\mathbb{R}^{P_C}$, which has rank $n - m^*$.

Then it suffices to show that following claim holds conditioned on the event in Theorem 2 holding.

Claim 23. $A'$ has full row rank, which is rank $n^*$.

First we show that conditioned on the event in Theorem 2 holding, we have $D \subseteq P_R$.

Indeed, conditioned on this event holding, we have

$$D = \{i \in [n] : \exists S \ni i : A_S \in T_{|S|}\}.$$  

Consider any set $S$ for which $A_S \in T_{|S|}$. By the definition of $T_{|S|}$, $A_S$ is the vertex-edge incidence matrix of a tree. Hence the peeling process defined above to create the 2-core will necessarily remove all the rows in $S$.

Now to prove the claim, suppose for contradiction that there existed a linear dependency among the rows of $A'$. Then since the row span of $A_{P_R}$ is $\mathbb{R}^{P_C}$, it must be the case that there is a linear dependency in $A$ which contains these rows of $A'$. That is, if $x^T A' = 0$, let $y := x^T A_{[n] \setminus P_R}$ such that $\text{supp}(y) \subseteq P_C$. Then we can find $x'$ such that $x'^T A_{P_R} = -y$, and hence combining $x$ and $x'$ yields a vector $x$ such that $x^T A = 0$.

However, this is a contradiction, because it means some row in $[n] \setminus P_R$ must be involved in a linear dependency — and hence a minimal linear dependency — contradicting the fact that $D \subseteq P_R$.

G Proof of Lemma 3

In this section we prove the following anticoncentration lemma.

Lemma 3 (c.f. [10] Lemma 8.2). Let $v \in \mathbb{R}^n$ be a deterministic vector with support at least $m$. Let $z \in \mathbb{R}^n$ be the random vector with i.i.d. Bernoulli entries with parameter $p \leq 1/2$. Then for any fixed $c$,

$$\Pr[v^T z = c] \leq \left( \frac{1}{\sqrt{\pi/2}} \right) \frac{1}{\sqrt{mp}} + \left( e^{(\ln(2)) - 1)mp} \right).$$

In particular, for $mp \geq 9$, we have:

$$\Pr[v^T z = c] \leq \frac{1}{\sqrt{mp}}.$$  

57
Proof. Let \( r_i \) be an independent Bernoulli random variable with parameter \( 2p \) and let \( s_i \) be an independent Bernoulli random variable with parameter \( \frac{1}{2} \). It follows that \( \Pr[r_i s_i = 1] = \Pr[z_i = 1] \). Thus, we have:

\[
\Pr(u^T z = c) = \Pr \left( \sum_{i=1}^{n} v_i r_i s_i = c \right) \\
\leq \Pr \left( \sum_{i=1}^{n} v_i r_i s_i = c \mid \sum_{i=1}^{n} r_i \geq mp \right) + \Pr \left( \sum_{i=1}^{n} r_i \leq mp \right)
\]

We employ a result of Erdős [11] to note:

\[
\Pr \left( \sum_{i=1}^{n} v_i r_i s_i = c \mid \sum_{i=1}^{n} r_i \geq mp \right) \leq \left( \frac{1}{\sqrt{\pi/2}} \right) \frac{1}{\sqrt{mp}}
\]

Finally, a Chernoff bound gives us:

\[
\Pr(\sum_{i=1}^{n} r_i \leq mp) \leq \left( 2^p \left( \frac{1 - 2p}{1-p} \right)^{1-p} \right)^m \leq e^{(\ln(2) - 1)mp}
\]

Combining these two expressions gives the desired result. \( \square \)

References

[1] Rawad Bitar, Mary Wootters, and Salim El Rouayheb. Stochastic gradient coding for straggler mitigation in distributed learning. *IEEE Journal on Selected Areas in Information Theory*, 2020.

[2] Béla Bollobás and Bollobás Béla. *Random graphs*. Cambridge university press, 2001.

[3] Gerandy Brito, Ioana Dumitriu, and Kameron Decker Harris. Spectral gap in random bipartite biregular graphs and applications. *arXiv preprint arXiv:1804.07808*, 2018.

[4] Zachary Charles and Dimitris Papailiopoulos. Gradient coding using the stochastic block model. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 1998–2002. IEEE, 2018.

[5] Zachary Charles, Dimitris Papailiopoulos, and Jordan Ellenberg. Approximate gradient coding via sparse random graphs. *arXiv preprint arXiv:1711.06771*, 2017.

[6] Vasek Chvátal. The tail of the hypergeometric distribution. *Discrete Mathematics*, 25(3):285–287, 1979.

[7] Amin Coja-Oghlan, Alperen A Ergür, Pu Gao, Samuel Hettermich, and Maurice Rolvien. The rank of sparse random matrices. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 579–591. SIAM, 2020.

[8] Kevin P Costello, Terence Tao, Van Vu, et al. Random symmetric matrices are almost surely nonsingular. *Duke Mathematical Journal*, 135(2):395–413, 2006.

[9] Kevin P Costello and Van Vu. On the rank of random sparse matrices. *arXiv preprint arXiv:0711.2696*, 2007.

[10] Kevin P Costello and Van H Vu. The rank of random graphs. *Random Structures & Algorithms*, 33(3):269–285, 2008.

[11] Paul Erdős. On a lemma of littlewood and offord. *Bulletin of the American Mathematical Society*, 51(12):898–902, 1945.

[12] Asaf Ferber, Matthew Kwan, and Lisa Sauermann. Singularity of sparse random matrices: simple proofs, 2020.
[13] Margalit Glasgow and Mary Wootters. Approximate gradient coding with optimal decoding. *arXiv preprint arXiv:2006.09638*, 2020.

[14] Han Huang. Rank of sparse bernoulli matrices. *https://arxiv.org/pdf/2009.13726.pdf*, 2021.

[15] Jiaoyang Huang. Invertibility of adjacency matrices for random $d$-regular graphs. *arXiv preprint arXiv:1807.06465*, 2018.

[16] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney. Singularity of discrete random matrices ii. *arXiv preprint arXiv:2010.06554*, 2020.

[17] Swanand Kadhe, O Ozan Koyluoglu, and Kannan Ramchandran. Gradient coding based on block designs for mitigating adversarial stragglers. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 2813–2817. IEEE, 2019.

[18] Antti Knowles and Jun Yin. The isotropic semicircle law and deformation of wigner matrices. *Communications on Pure and Applied Mathematics*, 66(11):1663–1749, 2013.

[19] Netanel Raviv, Rashish Tandon, Alex Dimakis, and Itzhak Tamo. Gradient coding from cyclic mds codes and expander graphs. In *International Conference on Machine Learning*, pages 4305–4313, 2018.

[20] Mark Rudelson and Roman Vershynin. The littlewood–offord problem and invertibility of random matrices. *Advances in Mathematics*, 218(2):600–633, 2008.

[21] Rashish Tandon, Qi Lei, Alexandros G Dimakis, and Nikos Karampatziakiskis. Gradient coding: Avoiding stragglers in distributed learning. In *International Conference on Machine Learning*, pages 3368–3376, 2017.

[22] Konstantin Tikhomirov. Singularity of random bernoulli matrices. *Annals of Mathematics*, 191(2):593–634, 2020.

[23] Van Vu. Recent progress in combinatorial random matrix theory. *arXiv preprint arXiv:2005.02797*, 2020.

[24] Hongyi Wang, Zachary Charles, and Dimitris Papailiopoulos. Erasurehead: Distributed gradient descent without delays using approximate gradient coding. *arXiv preprint arXiv:1901.09671*, 2019.

[25] Sinong Wang, Jiashang Liu, and Ness Shroff. Fundamental limits of approximate gradient coding. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3(3):1–22, 2019.