MINORS OF A RANDOM BINARY MATROID

COLIN COOPER, ALAN FRIEZE, AND WESLEY PEGDEN

ABSTRACT. Let $\mathbf{A}$ be an $n \times m$ matrix over $\mathbf{GF}_2$ where each column consists of k ones, and let M be an arbitrary fixed binary matroid. The matroid growth rate theorem implies that there is a constant C_M such that $m \geq C_M n^2$ implies that the binary matroid induced by $\mathbf{A}$ contains M as a minor. We prove that if the columns of $\mathbf{A} = \mathbf{A}_{n,m,k}$ are chosen *randomly*, then there are constants k_M, L_M such that $k \geq k_M$ and $m \geq L_M n$ implies that $\mathbf{A}$ contains M as a minor w.h.p.

1. INTRODUCTION

There is by now a vast and growing literature on the asymptotic properties of random combinatorial structures. First and foremost in this context are Random Graphs and Hypergraphs, see [3], [8] and [10] for books on this subject. Random groups in their own right and in the guise of random permutations are included in this. Going further afield into Algebraic Geometry we see a recent surge of interest in Random Simplicial Complexes, initiated by the paper of Linial and Meshulam [15]. See Kahle [11] for a recent survey. Another area of interest in this vein is that of Random Matroids. This paper concerns one aspect of these. For the basic facts on matroids see Welsh [22] or Oxley [16]. Basically we see that two models of a random matroid have been considered so far.

In the first model a matroid is chosen uniformly at random from the set of all matroids with n elements, see for example Oxley, Semple, Warshauer and Welsh [17]. Recently, there have been some breakthrough results in this subject. Bansal, Pendavingh and van der Pol [4] give a very close estimate for $\log \log m_n$ where m_n is the number of matroids on a fixed ground set with n elements. And Nelson [19] showed that almost all matroids are non-representable. Pendavingh and van der Pol [20] considered random matroids of rank r and showed that almost all r-sets will be bases in this model.

The second model considers representable matroids. Given a matrix $\mathbf{A}$ we let $\mathcal{M}(\mathbf{A})$ denote the representable matroid with ground set equal to the columns of $\mathbf{A}$ and independence given by linear independence. An example of this model is the space of $n \times m$ matrix with entries chosen independently and uniformly from $\mathbf{GF}_q$, see for example Kelley and Oxley [12].

Date: June 14, 2018.

Research supported in part by EPSRC grant EP/M005038/1.

Research supported in part by NSF Grants DMS1362785, CCF1522984 and a grant(333329) from the Simons Foundation.

Research supported in part by NSF grant DMS1363136.

The random graph $G_{n,m}$ can be identified with a random $n \times m$ 0/1 matrix $\mathbf{A}_{n,m,2}$ where each row represents a vertex and each column has exactly two ones and defines an edge. If the entries are considered to be in $\mathbf{GF}_2$ and the ones in each column are chosen at random, then we have a matrix representation of a random graph and a random graphic matroid.

The columns of $\mathbf{A}_{n,m,2}$ define a (random) graphic matroid. If we want to generalize this to random sample from a larger class of binary matroids, then one natural way is to take k random ones instead of 2 ones in each column, to obtain the random matrix $\mathbf{A}_{n,m,k}$, the vertex-edge incidence matrix of a random k-uniform hypergraph. It is this model of a random binary matroid that is the subject of this paper.

Many properties of a matroid are determined by whether or not it contains some particular fixed matroid as a minor. For example a binary matroid is regular if and only if it does not contain the Fano plane or its dual as a minor, see Tutte [21]. We are interested in the event that $\mathbf{A}_{n,m,k}$ contains a fixed binary matroid M as a minor. The matroid growth rate theorem of Geelen, Kung and Whittle [9] implies that there is a constant C_M depending only on M such that *any* binary matroid of rank n on $m > C_M n^2$ elements must contain M has a minor (see also Kung [13]. We prove that (when k is large), for the random matroid induced by $\mathbf{A}_{n,m,k}$ this quadratic condition can be replaced by a linear one. We prove the following:

Theorem 1.1. *Let M be a fixed binary matroid. Then there exist constants k_M, L_M such that if $k \geq k_M$ and $m \geq L_M n$ then w.h.p.*[1] *$\mathcal{M}_{n,m,k}$ contains M as a minor.*

We briefly recall the definition of the minor relation for matroids. Given a matroid $\mathcal{M}$ on the ground set E and with the family $\mathcal{I}$ of independent sets, for $X \subseteq E$, the deletion $\mathcal{M} \setminus X$ is the matroid on $E \setminus X$ whose independent sets consist of $\{I \in \mathcal{I} : \ I \subset E \setminus X\}$. The contraction $\mathcal{M}/X$ ($X \in \mathcal{I}$) is the matroid on $E \setminus X$ whose independent sets are $\{I \subset E \setminus X : \ I \cup X \in \mathcal{I}\}$. M is a minor of $\mathcal{M}$ if it can be obtained from $\mathcal{M}$ by deletion and contraction operations. (For $X \notin \mathcal{I}$, the contraction can be defined by $\mathcal{M}/X := (\mathcal{M}/Y) \setminus X$, where Y is any basis of X.)

Theorem 1.1 is related to the result of Altschuler and Yang [1]. They prove that if matrix $\mathbf{M}$ is an $n(m) \times m$ matrix with random entries in $\mathbf{GF}_q$ and $m - n(m) \to \infty$ then w.h.p. the matroid associated with $\mathbf{M}$ contains any fixed minor. This can be related to our theorem on taking $q = 2$ and $k = n(m)/2$. (We have reversed the roles of m, n from their statement.) However, the results of [1] rely heavily on the fact that pre-multiplying a uniform random matrix in this model by a non-singular matrix yields another uniform random matrix. Our model lacks this property. Furthermore, multiplying $\mathbf{A}_{n,m,k}$ by a non-singular matrix will not fix this property. This is because whatever matrix we use as a pre-multiplier, we will only have a sample space of size at most $\binom{n}{k}$ for the resulting column set, as opposed to 2^n.

2. Proof of Theorem 1.1

2.1. **Outline of our proof.** Fix k and let the matrix $\mathbf{A}_m = \mathbf{A}_{n,m,k}$ have columns $[\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m]$ where $m = Kn$ for K sufficiently large. Let M be a fixed binary matroid and

[1] A sequence of properties $\mathcal{E}_n, n \geq 1$ is said to hold *with high probability* (w.h.p.) if $\lim_{n\to\infty} \mathbf{Pr}(\mathcal{E}_n) = 1$.

let $\mathbf{R}_M = [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_q]$ be a representation of M by a $p \times q$ matrix. Assume without loss of generality that $\mathbf{R}_M$ has full row-rank p. In this outline we will assume that k is odd. There are some minor adjustments needed for k even.

Let

$$n_1 = n \text{ and } m_1 = \frac{n}{4}. \tag{1}$$

Denote the $n_1 \times m_1$ matrix consisting of the first m_1 columns of $\mathbf{A}_m$ by $\mathbf{X}$. It follows from Theorem 1 of Cooper [6] that w.h.p. the columns of $\mathbf{X}$ are linearly independent.

We will use results on hypergraph cores to find a sub-matrix $\mathbf{B}_1$ of $\mathbf{X}$ that has n_2 rows and m_2 linearly independent columns where n_2 is close to n_1 and $m_2 \geq n/5$, and with the property that $\mathbf{B}_1$ has k random ones in each column and at least $k/10$ ones in each row.

We extend $\mathbf{B}_1$ to an $n_2 \times n_2$ non-singular submatrix $\mathbf{B}$ of $\mathbf{A}_m$ which again has exactly k ones in each column, as follows. Let I_1 denote the index set of the rows of $\mathbf{B}_1$. We extend $\mathbf{B}_1$ by choosing Ln columns of $\mathbf{A}_m$, disjoint from $\mathbf{X}$ to create a submatrix $\mathbf{L}$. Here L is a sufficiently large constant. These columns will only have ones in rows indexed by I_1. Again using properties of hypergraph cores, we show that w.h.p. $\mathbf{L}$ contains a submatrix $\mathbf{L}_1$ which has n_3 rows and m_3 columns which (i) has full row rank and (ii) each row has at least ζkL ones. Here n_3 is close to n_2 and $0 < \zeta < 1$. We then argue that the matrix $\mathbf{L}_2 = [\mathbf{B}_1 : \mathbf{L}_3]$ has n_2 rows and has full row rank. Here $\mathbf{L}_3$ is obtained from $\mathbf{L}_1$ by adding $n_2 - n_3$ rows of zeros. The matrix $\mathbf{B}$ is an arbitrary extension of $\mathbf{B}_1$ to a square non-singular $n_2 \times n_2$ sub-matrix of $\mathbf{L}_2$.

We then argue that w.h.p. the rows of $\mathbf{B}^{-1}$ have between $\varepsilon_0 n_2 = \frac{1}{2}e^{-k}n_2$ and $n_2 - \varepsilon_0 n_2$ ones.

We let $\widehat{\mathbf{A}}$ be the $n_2 \times m_3$ submatrix of $\mathbf{A}_m$ whose rows are the rows of $\mathbf{B}$, and whose columns are those columns of $\mathbf{A}_m$ which have ones only in rows of $\mathbf{B}$. Note that $\mathcal{M}(\widehat{\mathbf{A}})$ is a minor of $\mathcal{M}(\mathbf{A}_m)$. Now write $\widehat{\mathbf{A}} = [\mathbf{B} : \mathbf{M}]$ and consider the matrix $\widehat{\mathbf{A}}_1 = [\mathbf{I} : \mathbf{M}_1]$ for $\mathbf{M}_1 = \mathbf{B}^{-1}\mathbf{M}$, where we assume that the first n_2 columns form the $n_2 \times n_2$ identirt matrix. Suppose that $\mathbf{M}_1$ contains a submatrix equal to our target matrix $\mathbf{R}_M$. Then we are done. Indeed, suppose w.l.o.g. that $\mathbf{R}_M$ lies in the first p rows and the first q columns of $\mathbf{M}_1$. Then we get M as a minor of $\mathcal{M}(\widehat{\mathbf{A}})$ (and hence of $\mathcal{M}(\mathbf{A}_m)$) by deleting the first p columns of $\mathbf{B}$ and the last $m_3 - n_2 - q$ columns of $\mathbf{M}$ and contracting the last $n_2 - p$ columns of $\mathbf{B}$, as we explain next.

Recall that a minor of $\mathbf{A}_m$ is obtained by deleting and contracting columns. Recall from the definition of contraction that if S denotes an independent set (of column indices), then a set T (of column indices) disjoint from S is independent in the contraction $\mathcal{M}/S$ iff $S \cup T$ is an independent set (of columns) in $\mathcal{M}$.

Contraction is simple if the columns S are a subset of the columns of an identity matrix $\mathbf{I} = \mathbf{I}_{n_2}$. In view of this, we pre-multiply $\widehat{\mathbf{A}} = [\mathbf{A} : \mathbf{M}]$ by $\mathbf{B}^{-1}$ to obtain $\widehat{\mathbf{A}}_1 = [I : \mathbf{M}_1]$. Pre-multiplying by a non-singular matrix does not change the underlying matroid, seeing as column dependence/independence is preserved. We can assume that the first n_2 columns form the $n_2 \times n_2$ identity matrix $\mathbf{I}$. If we contract a set S of the columns of $\mathbf{I}$, then a representation of the contracted matroid is given by deleting the $|S|$ rows of $\widehat{\mathbf{A}}_1$ that have a

one in a column of S to obtain a matrix $\widehat{\mathbf{A}}_2$. In which case we see that a set T of columns of $\widehat{\mathbf{A}}_2$ is independent in $\widehat{\mathbf{A}}_2$ if and only if the set of columns corresponding to $S \cup T$ is independent in the matroid represented by $\mathbf{A}_m$.

To prove that $\mathbf{R}_M = [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_q]$ appears as a submatrix of $\mathbf{M}$, we will consider $\mathbf{B}^{-1}\mathbf{c}$ where $\mathbf{c}$ is a random column of $\widehat{\mathbf{A}}$ outside of the $n/4 + Ln$ columns considered so far in the construction of $\mathbf{B}$. For a set R of rows and a column $\mathbf{x}$ of $\mathbf{B}^{-1}\widehat{\mathbf{A}}$, let $\phi_R(\mathbf{x})$ be the column $\mathbf{x}$ restricted to the rows R. We argue next that we can find R of size ρ such that

$$\mathbf{Pr}(\phi_R(\mathbf{B}^{-1}\mathbf{c}) = \mathbf{m}_j) = \Omega(1) \text{ for } 1 \le j \le q. \tag{2}$$

This means that w.h.p. we can find a copy of each column of $\mathbf{R}_M$ by searching through ω random columns, where $\omega = o(n)$ is any function tending to infinity with n.

To justify (2), let S_i denote the support of the ith row of $\mathbf{B}^{-1}$. Our strategy for analyzing $\mathbf{B}^{-1}\mathbf{c}$ is to show that there is a set R of p rows of $\mathbf{B}^{-1}$ and a partition $A_0, A_1, \dots, A_\ell$ of $[n_2]$ and a corresponding $p \times \ell$, $\{0,1\}$ matrix $\mathbf{D} = (\mathbf{D}[i,j])$ with certain useful properties. We will have that for all i, j, S_i contains A_j or is disjoint from it. And furthermore for some constants $0 < \varepsilon_1 \ll \varepsilon_0 \ll 1$, $\mathbf{D}[i,j] = 1$ implies (i) $r_{i,k} = 1$ for $k \in A_j, i \in R$, ($\mathbf{r}_i = (r_{i,\cdot})$ being the ith row of $\mathbf{B}^{-1}$), (ii) $|A_j| \ge \varepsilon_1 n_2$ for $j \ge 0$ and (iii) $\mathbf{D}$ has full row rank.

Given R and $\mathbf{D}$ we proceed as follows: Let $\mathbf{c} = (c_1, c_2, \dots, c_{n_2})$ be a random column with k 1's. Let $\mathbf{v}$ satisfy $\mathbf{D}\mathbf{v} = \mathbf{m}_1$ (the first column of $\mathbf{R}_M$) and $v_j = 0$ if $|A_j| < \varepsilon_1 n$. We can assume that $\mathbf{v}$ has at most p ones and that $k \ge p$. Equation (2) follows from

$$\mathbf{Pr}(\phi_R(\mathbf{B}^{-1}\mathbf{c}) = \mathbf{m}_1) \ge \mathbf{Pr}(\mathbf{c}_R = \mathbf{v}) = \Omega(1), \tag{3}$$

where $\mathbf{c}_R = (d_0, d_1, \dots, d_\ell)$ and where $d_j = \sum_{l \in A_j} c_l$.

The condition in (3) will be satisfied if exactly one element is chosen from each A_j such that $v_j = 1$ and the rest are chosen from A_0. This has probability $\Omega(1)$.

We will show in Section 2.7 how to choose the set of rows R so that they contain at least $\varepsilon_1 n_2$ common zeros. Then in Section 2.8 we will show that if $S_i^1 = S_1, S_i^0 = \bar{S}_i$ that the partition $A_{\boldsymbol{\sigma}} = \bigcap_{j=1}^K S_j^{\xi_j}, \boldsymbol{\sigma} = (\xi_1, \xi_2, \dots, \xi_p)$ (as $\xi_1, \xi_2, \dots, \xi_p$ runs over $\{0,1\}^p$) suffices as a partition. We will take $d_{i,\boldsymbol{\sigma}} = 1$ only if $\xi_i = 1$ and $|A_{\boldsymbol{\sigma}}| \ge \varepsilon_1 n_2$.

2.2. **Some Notation.** We summarize here the meaning of some parameters. The reader might find this useful to refer back to.

(i) $\mathbf{B}_1$ is the $n_2 \times m_2$ submatrix derived from the first m_1 columns of $\mathbf{A}_{n,m,k}$. Every column has k ones and every row has at least $k/10$ ones. The columns of $\mathbf{B}_1$ are linearly independent and the values n_2, m_2 satisfy (8), (9) below. The set I_1 is the index set of rows of $\mathbf{B}_1$.

(ii) $\mathbf{L}$ is an $n_2 \times Ln$ submatrix of $\mathbf{A}_{n,m,k}$, whose columns are disjoint from those of $\mathbf{B}_1$ and whose rows have index $I_2 \subseteq I_1$.

(iii) $\mathbf{L}_1$ is an $n_3 \times m_3$ submatrix of $\mathbf{L}$ which has rank n_3, where n_3, m_3 satisfy (13), (14).

(iv) $\mathbf{L}_2$ is an n_2 row matrix that contains $\mathbf{B}_1$ as a sub-matrix and has rank n_2 and many more than n_2 columns. It is therefore possible to find an $n_2 \times n_2$ non-singular submatrix $\mathbf{B}$ that contains $\mathbf{B}_1$ and is contained in $\mathbf{L}_2$.
(v) In general bold named variables are either matrices or vectors.

We now give a detailed proof of Theorem 1.1.

2.3. **Building $\mathbf{B}_1$.** Consider the k-uniform hypergraph H_1 induced by the first $m_1 = n/4$ columns $\mathbf{X}$ of $\mathbf{A}_{n,m,k}$. I.e. the hypergraph with a vertex for each row and where each edge $e_j, j \le n/4$ corresponds to the column $\mathbf{c}_j$ of $\mathbf{X}$ via e_j contains an element $i \in [n]$ if and only if $\mathbf{X}[i,j] = 1$. H_1 is distributed as a random k-uniform hypergraph with n_1 vertices and m_1 edges. We show next that w.h.p. the $k/10$-core C_1 of H_1 is large. This will provide us with a matrix $\mathbf{B}_1$ with at least $k/10$ ones in each row. The r-core of a hypergraph $H = (V, E)$ is the largest set $S \subseteq E$ such that each $x \in S$ has degree at least r in the sub-hypergraph of H induced by S i.e. x lies in at least r edges e, $e \subseteq S$.

We use some results on the cores of random k-uniform hypergraphs (see e.g. Cooper [7] or Molloy [14]). Let $c = km_1/n_1 = k/4$, and let x be the greatest solution to

$$c = \frac{k}{4} = \frac{x}{\left(1 - e^{-x}\sum_{i=0}^{k/10-2}\frac{x^i}{i!}\right)^{k-1}}. \tag{4}$$

We will use a simple continuity argument to prove the existence of x and bound it as in (7) below.

It is known that w.h.p.,

$$n_2 = |V(C_1)| \approx n_1\left(1 - e^{-x}\sum_{i=0}^{k/10-1}\frac{x^i}{i!}\right), \tag{5}$$

and

$$m_2 = |E(C_1)| \approx m_1\left(\frac{x}{c}\right)^{k/(k-1)}. \tag{6}$$

Here, $A(x) \approx B(x)$ stands for $A(x) = (1 + o(1))B(x)$ as $x \to \infty$ $A(x) \gtrsim B(x)$ stands for $A(x) \ge (1 + o(1))B(x)$ as $x \to \infty$.

We will first argue that for k large we have

$$\frac{k}{5} < x \le \frac{k}{4}. \tag{7}$$

The upper bound follows directly from the definition (4). To prove the lower bound let

$$S(x) = \frac{k}{4} - \frac{x}{\left(1 - e^{-x}\sum_{i=0}^{k/10-2}\frac{x^i}{i!}\right)^{k-1}}.$$

If $x \geq 2(i+1)$, then $\frac{x^i}{i!} \leq \frac{x^{i+1}}{2(i+1)!}$. Thus for $x \geq k/5$ and $\theta = 1, 2$,

$$\sum_{i=0}^{k/10-\theta} \frac{x^i}{i!} \leq \frac{x^{k/10}}{(k/10)!} \leq \left(\frac{10xe}{k}\right)^{k/10}$$

and so

$$e^{-x} \sum_{i=0}^{k/10-\theta} \frac{x^i}{i!} \leq \left(\frac{10xe}{k}\, e^{-10x/k}\right)^{k/10} \leq \left(\frac{2}{e}\right)^{k/10} \qquad \text{since } x \geq k/5.$$

Thus $S(k/5) > 0$ for k large. As $S(k/4) < 0$, the lower bound in (7) follows from the continuity of $S(x)$. It then follows from (5) that w.h.p.

$$n_1 \geq n_2 = |V(C_1)| \geq n_1 \left(1 - \frac{1}{k}\right). \tag{8}$$

Similarly, using (6) along with $c = k/4$ and $x > k/5$ from (7) gives

$$\frac{n}{4} \geq m_2 \gtrsim \frac{n_1}{4} \left(\frac{4}{5}\right)^{k/(k-1)} \geq \frac{n_2}{5}, \tag{9}$$

for k large.

Now consider the submatrix $\mathbf{B}_1$ of $\mathbf{X}$ comprised of the columns corresponding to the edges of H_1 that are contained in C_1. The distribution of ones in $\mathbf{B}_1$ is that each of the m_2 columns chooses k random ones from n_2 rows, subject only to each row having at least $k/10$ ones. This is an interpretation of a standard result on cores of graphs being random subject to a lower bound on minimum degree. Let I_1 denote the index set of the rows of $\mathbf{B}_1$. Thus $|I_1| = n_2$.

2.4. **Extending $\mathbf{B}_1$ to a basis.** We fix some constant $L > 1$ and begin by choosing Ln columns of $\mathbf{A}_m$ disjoint from $\mathbf{X}$ to make a sub-matrix $\mathbf{L}$. We choose the first Ln columns following $\mathbf{X}$ that only have ones in rows indexed by I_1. The probability that a random column only has ones in rows I_1 is $\frac{\binom{n_2}{k}}{\binom{n}{k}} = \Omega(1)$ and so w.h.p. we only need to examine $O(n)$ columns of $\mathbf{A}_m$ in order to find these Ln columns. Now let $0 < \zeta < 1$ be a small constant. Let now H_2 denote the k-uniform hypergraph induced by the columns of $\mathbf{L}$ and let $C_2 = C_2(H_2)$ denote its ζLk-core. Using [7], [14] once again we see that we have to let x be the greatest solution to

$$c = Lk = \frac{x}{\left(1 - e^{-x} \sum_{i=0}^{\zeta Lk-2} \frac{x^i}{i!}\right)^{k-1}}. \tag{10}$$

Then w.h.p.,

$$n_3 = |V(C_2)| \approx n_2 \left(1 - e^{-x} \sum_{i=0}^{\zeta Lk-1} \frac{x^i}{i!}\right). \tag{11}$$

We will next argue that for k, L large we have

$$\frac{(1+\zeta)Lk}{2} \leq x \leq Lk. \tag{12}$$

The upper bound follows directly from the definition (10). To prove the lower bound let now

$$S(x) = Lk - \frac{x}{\left(1 - e^{-x}\sum_{i=0}^{\zeta Lk-2}\frac{x^i}{i!}\right)^{k-1}}.$$

If $x \geq \frac{(1+\zeta)(i+1)}{2\zeta}$ then $\frac{x^i}{i!} \leq \frac{\xi x^{i+1}}{(i+1)!}$ where $\xi = \frac{2\zeta}{1+\zeta} < 1$. Thus for $x \geq \frac{(1+\zeta)Lk}{2}$ and $\theta = 1, 2$,

$$\sum_{i=0}^{\zeta Lk-\theta}\frac{x^i}{i!} \leq \frac{1}{1-\xi}\cdot\frac{x^{\zeta Lk}}{(\zeta Lk)!} \leq \frac{1}{1-\xi}\left(\frac{ex}{\zeta Lk}\right)^{\zeta Lk},$$

and if $\eta = \frac{1+\zeta}{2\zeta} < 1$ then

$$e^{-x}\sum_{i=0}^{\zeta Lk-\theta}\frac{x^i}{i!} \leq \frac{1}{1-\xi}\left(\frac{ex}{\zeta Lk}e^{-x/(\zeta Lk)}\right)^{\zeta Lk} \leq \frac{\left(\eta e^{1-\eta}\right)^{\zeta Lk}}{1-\xi}.$$

Thus $S(\frac{(1+\zeta)Lk}{2}) > 0$ for large k and the lower bound in (12) follows by continuity.

It then follows from (11) that for large enough L, we have that w.h.p.

$$n_2 \geq n_3 = |V(C_2)| \geq n_2\left(1 - e^{-2k}\right). \tag{13}$$

Similarly, using a similar expression to (6) along with $c = Lk$ and $x \geq (1+\zeta)Lk/2$ in (7) gives us that the number m_3 of edges in C_2 satisfies

$$Ln \geq m_3 \gtrsim Ln_2\left(\frac{1+\zeta}{2}\right)^{k/(k-1)} \text{ and so } m_3 \geq \frac{4(1+\zeta)Ln_2}{9}, \tag{14}$$

for k large.

We argue next that w.h.p. the matrix $\mathbf{L}_1$ induced by C_2 has rank n_3. For this we rely on the following lemma, which we will need for several purposes:

Lemma 2.1. *Let $\mathbf{A} = (\mathbf{A}[i,j])$ be an $N \times M$ matrix over $\mathbf{GF}_2$ chosen uniformly at random from matrices where each column has k ones, and condition on the event that each row has at least $\gamma k\sigma$ ones, where $\gamma < 1$ and $\gamma k > 1$ and $\sigma = M/N = O(1)$. Let $\boldsymbol{\alpha}$ be a fixed member of $\mathbf{GF}_2^M$. If $\mathcal{E}_{s,\boldsymbol{\alpha}}$ is the event that there exists a set S of rows with $|S| = s$ whose sum is $\boldsymbol{\alpha}$, then*

(a)

$$\mathbf{Pr}(\exists 1 \leq s \leq Ne^{-k} : \mathcal{E}_{s,\boldsymbol{\alpha}}) = O(N^{-K}).$$

(b) If $\sigma \geq e^{5k}/(1-\gamma)^2$ then

$$\mathbf{Pr}(\exists Ne^{-k} < s < N : \mathcal{E}_{s,\boldsymbol{\alpha}}) = O(N^{-K}).$$

Here K can be made arbitrarily large, by taking k sufficiently large. Note also that we exclude $|S| = N$ from the statement of the lemma, since this would be false with $\boldsymbol{\alpha}$ equal to all ones (k odd) or all zeros (k even).

We apply the lemma to $\mathbf{L}_1$ by taking $N = n_3, M = m_3$ where $\zeta = 1/2$ and then let γ be equal to $\frac{Ln_3}{2m_3} \in \left[\frac{k-1}{2k}, \frac{3}{4}\right]$. The bounds on γ being justified by (1), (8), (13) and (14). Assume that $L \geq 10e^{5k}$, so that the lower bound on σ in (b) is satisfied. We will now make the following:

Assumption A: k is odd.

We will deal with the case of k even in Section 3. Now if k is odd and $\mathbf{L}_1$ does not have full row rank, then $\mathcal{E}_{s,\boldsymbol{\alpha}}$ occurs with $\boldsymbol{\alpha} = \mathbf{0}$ for some $1 \leq s \leq n_3$. But Lemma 2.1 implies that

$$\mathbf{Pr}(\exists 1 \leq s \leq n_3 : \mathcal{E}_{s,\boldsymbol{\alpha}} \text{ occurs}) = O(n^{-K}).$$

So, w.h.p. we have found an $n_3 \times m_3$ matrix $\mathbf{L}_1$ of rank n_3. Now consider the matrix $\mathbf{L}_2 = [\mathbf{B}_1 : \mathbf{L}_3]$. Here $\mathbf{L}_3$ is obtained from $\mathbf{L}_1$ by adding $n_2 - n_3$ rows of zeros. We claim that w.h.p. $\mathbf{L}_2$ has rank n_2. Let $I_2 \subseteq I_1$ be the row indices of $\mathbf{L}_1$. Let the rows of $\mathbf{L}_2$ be $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{n_2}$ and suppose that there exists $J \subseteq I_1$ such that $\sum_{i \in J} \mathbf{a}_i = 0$. Then we have $J \cap I_2 = \emptyset$ else $\mathbf{L}_1$ does not have rank n_3. We have $J \subseteq I_1 \setminus I_2$ and then (13) implies that $|J| \leq ne^{-2k}$. But then we obtain a contradiction from Lemma 2.1(a) applied to the rows of $\mathbf{B}_1$.

Because $\mathbf{L}_2$ has full row rank, we can obtain $\mathbf{B}$ as an extension of $\mathbf{B}_1$ to an $n_2 \times n_2$ non-singular sub-matrix of $\mathbf{L}_2$. After this we order the columns of $\mathbf{B}$ so that the columns of $\mathbf{B}_1$ come first.

2.5. **Proof of Lemma 2.1.** We first deal with small s. Suppose that $1 \leq s \leq Ne^{-k}$. If $T \subseteq [N], |S| = s$, let $\mathcal{E}_{j,T,S}$ denote the event that column j of $\mathbf{A}$ has ones in all of the rows T and zero's in the rows $S \setminus T$. Then where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_M)$,

$$\mathbf{Pr}(\mathcal{E}_{s,\boldsymbol{\alpha}}) \leq \sum_{S \subseteq [N], |S| = s} \sum_{\substack{d_j = \alpha_j \bmod 2, j \in [M] \\ d_1 + d_2 + \cdots + d_M \geq \gamma k \sigma s}} \sum_{S_j \subseteq S, |S_j| = d_j} \mathbf{Pr}\left(\bigcap_{j=1}^{M} \mathcal{E}_{j,S_j,S}\right). \tag{15}$$

Explanation: We sum over sets S and then for each $j \in [M]$ we fix the number of ones $d_j = |\{i \in S : \mathbf{A}[i,j] = 1\}$ of column j that appear in the rows S. We then choose the rows S_j where these ones appear and multiply by the probability that things are just so.

To estimate the probabilities in the RHS of (15) we will use the following model: we choose $\mathbf{X}$ uniformly from $[N]^{kM}$. Then column i of $\mathbf{A}$contains a one in positions $X_{k(i-1),j}$ for $1 \leq j \leq k$ and $1 \leq i \leq M$. It is possible that $X_{i,j_1} = X_{i,j_2}$ for some i, j_1, j_2. Let $\mathcal{S}$ be the event that this does not happen. Then

$$\mathbf{Pr}(\mathcal{S}) \geq \left(1 - \frac{\binom{k}{2}}{N}\right)^M \geq e^{-k^2\sigma}. \tag{16}$$

Explanation: $\frac{\binom{k}{2}}{N}$ bounds from above, the probability that a fixed column contains a repeat, by the expected number of repeats. Each column is independently generated and (16) follows.

Thus $\mathbf{Pr}(\mathcal{S}) = \Omega(1)$ and events involving $\mathbf{X}$ that occur w.h.p. will also occur w.h.p. if we condition on $\mathcal{S}$.

We see next that given $\mathcal{S}$, each matrix with exactly k ones in each column is equally likely. Indeed, each such matrix arises from the same number $(k!)^M$ of choices of $\mathbf{X}$. Thus we can use $\mathbf{X}$ to generate our matrix $\mathbf{A}$ in a uniform way. It remains to deal with the lower bounds on row sums.

The row-sums $\rho_i = |\{l : x_l = i\}|, 1 \le i \le N$ will be independent Poisson random variables, subject to $\rho_i \ge \gamma k\sigma, i \in [N]$ and $\rho_1 + \rho_2 + \ldots + \rho_N = kM$. This was proved in [2] where the lower bound of $\gamma k\sigma$ is replaced by 2. We include a proof in an appendix for completeness. Thus

$$\mathbf{Pr}(\rho = l) = \frac{\lambda^l}{l! f_{\gamma k\sigma}(\lambda)} \text{ where } f_l(\lambda) = e^\lambda - \sum_{i=0}^{l-1} \frac{\lambda^i}{i!}. \tag{17}$$

Here we choose λ so that $\mathbf{E}(\rho) = k\sigma$, which implies that

$$\frac{\lambda f_{\gamma k\sigma - 1}(\lambda)}{f_{\gamma k\sigma}(\lambda)} = k\sigma. \tag{18}$$

This choice of λ ensures that $\mathbf{Pr}(\rho_1 + \rho_2 + \ldots + \rho_N = kM) = \Omega(M^{-1/2})$. This follows from a version of the local central limit theorem, proved in [2].

It follows that for large k, we have

$$\frac{k\sigma}{2} \le \lambda \le k\sigma \text{ and } f_{\gamma k\sigma}(\lambda) \ge e^{\gamma k\sigma/2}. \tag{19}$$

The upper bound in (19) follows from the fact that $f_{\gamma k\sigma-1}(\lambda) > f_{\gamma k\sigma}(\lambda)$. The lower bound follows from the fact that if k is large, then the RHS of (18) is large and then λ approaches $k\sigma$ which is large. This then implies that $f_{\gamma k\sigma-1}(\lambda)$ approaches $f_{\gamma k\sigma}(\lambda)$ as k grows.

Suppose now that we condition on the row sums $\rho_1 = \theta_1, \rho_2 = \theta_2, \ldots, \rho_N = \theta_N$. If $d_1 + d_2 + \cdots + d_M = \ell$ then

$$\mathbf{Pr}\left(\bigcap_{j=1}^{M} \mathcal{E}_{j,S_j,S}\right) \le \frac{(kM-\ell)!}{(kM)!} \prod_{j=1}^{M} \prod_{i \in S_j} (\theta_i k) \le \frac{e^{\ell^2/kM}}{M^\ell} \prod_{i \in S} \theta_i^{\theta_i}. \tag{20}$$

Explanation of (20)**:** The conditioned model involves a vector $\mathbf{X} \in [N]^{kM}$ that can be viewed as a random permutation of ρ_i copies of i for $i \in M$. We can assume that these copies are distinguishable. Then, if $i \in S_j$ and $(i_1, j_1), \ldots, (i_l, j_l)$ represent prior assignments,

$$\mathbf{Pr}(\mathbf{A}[i,j] = 1 \mid \mathbf{A}[i_1, j_1] = 1, \ldots, \mathbf{A}[i_l, j_l] = 1) \le \frac{k\theta_i}{kM - l}. \tag{21}$$

To see (21), observe that there are at most k positions in $\mathbf{X}$ that give us $\mathbf{A}[i,j] = 1$ and for each there are at most ρ_i out of $kM - l$ equally likely choices of being i. The middle term in equation (20) follows. The final estimate follows by using Stirling's inequality.

Next let

$$D_\ell = \left\{ \mathbf{d} = (d_1, d_2, \ldots, d_M) : d_j = \alpha_j \bmod 2, d_j \le k, j \in [M], \sum_{j\in[M]} d_j = \ell \right\}$$

and

$$E_\ell = \left\{ \boldsymbol{\theta} = (\theta_i, i \in S) : \sum_{i\in S} \theta_i = \ell,\ \theta_i \ge \gamma k\sigma, i \in S \right\}.$$

Note that

$$|D_\ell| \le_b \binom{M+\ell/2-1}{\ell/2-1} \le \frac{M^{\ell/2} e^{\ell^2/4M}}{(\ell/2)!} \text{ and } |E_\ell| = \binom{\ell-\gamma k\sigma s+s-1}{s-1} < 2^\ell. \tag{22}$$

Here the notation $A \le_b B$ is used in place of $A = O(B)$.

The first inequality in (22) is obtained as follows: Let $d'_j = (d_j-1)/2$ if $\alpha_j = 1$ and let $d'_j = d_j/2$ if $\alpha_j = 0$. Then $\sum_j d'_j = (\ell-\ell_1)/2$ where ℓ_1 is the number of α_j equal to one. Knowing $\boldsymbol{\alpha}$, which is fixed, we can re-construct the d_j's from the d'_j's. This explains the binomial coefficient. After this we use

$$B!\binom{A+B}{B} = A^B \prod_{i=0}^{B-1} \left(1 + \frac{B-i}{A}\right) \le A^B e^{B(B+1)/2A}.$$

Plugging (20) into (15) we obtain,

$$\begin{aligned}
&\mathbf{Pr}(\mathcal{E}_{s,\boldsymbol{\alpha}}) \\
&\le \sum_{S\subseteq[N],|S|=s} \sum_{\ell=\gamma k\sigma s}^{kM} \sum_{\mathbf{d}\in D_\ell} \sum_{S_j\subseteq S,|S_j|=d_j} \sum_{\boldsymbol{\theta}\in E_\ell} \mathbf{Pr}(\rho_i = \theta_i, i \in S) \times \frac{e^{\ell^2/kM}}{M^\ell} \prod_{i\in S} \theta_i^{\theta_i} \\
&\le_b M^{1/2} \sum_{S\subseteq[N],|S|=s} \sum_{\ell=\gamma k\sigma s}^{kM} \sum_{\mathbf{d}\in D_\ell} \sum_{S_j\subseteq S,|S_j|=d_j} \sum_{\boldsymbol{\theta}\in E_\ell} \prod_{i\in S} \frac{\lambda^{\theta_i}}{\theta_i! f_{\gamma k\sigma}(\lambda)} \frac{e^{\ell^2/kM}}{M^\ell} \prod_{i\in S} \theta_i^{\theta_i}. \\
&\le_b M^{1/2} \binom{N}{s} \sum_{\ell=\gamma k\sigma s}^{kM} \sum_{\mathbf{d}\in D_\ell} \sum_{\boldsymbol{\theta}\in E_\ell} \frac{\lambda^\ell}{f_{\gamma k\sigma}(\lambda)^s} \left(\frac{se^{\ell/kM}}{M}\right)^\ell \prod_{i=1}^{s} \frac{\theta_i^{\theta_i}}{\theta_i!}
\end{aligned}$$

To obtain the last line we used $\sum_{S_j\subseteq S,|S_j|=d_j} 1 \le s^{d_1+\cdots+d_M} = s^\ell$.

Thus,

$$\begin{aligned}
\mathbf{Pr}(\mathcal{E}_{s,\boldsymbol{\alpha}}) &\le_b M^{1/2} \binom{N}{s} \sum_{\ell=\gamma k\sigma s}^{kM} \sum_{\mathbf{d}\in D_\ell} \sum_{\boldsymbol{\theta}\in E_\ell} \frac{\lambda^\ell}{f_{\gamma k\sigma}(\lambda)^s} \left(\frac{se^{1+\ell/kM}}{M}\right)^\ell \\
&\le_b M^{1/2} \left(\frac{Ne}{s}\right)^s \sum_{\ell=\gamma k\sigma s}^{kM} \frac{M^{\ell/2} e^{\ell^2/4M}}{(\ell/2)!} \frac{(2k\sigma)^\ell}{e^{\gamma k\sigma s/2}} \left(\frac{se^{1+\ell/kM}}{M}\right)^\ell
\end{aligned}$$

$$\leq_b M^{1/2}\left(\frac{Ne}{s}\right)^s \sum_{\ell=\gamma k\sigma s}^{kM}\left(\frac{e^{k/3}\sigma s}{\ell^{1/2}M^{1/2}}\right)^\ell e^{-\gamma k\sigma s/2}, \tag{23}$$

since k is large. Now if u_ℓ is the ℓth root of the summand in (23) then

$$\frac{u_\ell}{u_{\ell-2}} \leq \frac{e^{2k/3}\sigma^2 s^2}{\ell M} \leq \frac{e^{2k/3}\sigma^2 s^2}{\gamma k s\sigma^2 N} = \frac{e^{2k/3}s}{\gamma k N} \leq \frac{1}{2},$$

since $\gamma k > 1$.

It now follows, since $\gamma k > 1$ and the largest term in the sum in (23) is at $\ell = \gamma k\sigma s$, that

$$\mathbf{Pr}(\mathcal{E}_{s,\boldsymbol{\alpha}}) \leq_b M^{1/2}\left(\frac{Ne}{s}\right)^s\left(\frac{e^{k/3}s}{N}\right)^{\gamma k\sigma s/2} \leq M^{1/2}\left(\frac{se^{2k/3}}{N}\right)^{\gamma k\sigma s/3}. \tag{24}$$

Summing the RHS of (24) for $1 \leq s \leq Ne^{-k}$ and taking k large completes the proof of part (a) of the lemma.

Assume now that $Ne^{-k} \leq s \leq N/2$. If the sum of the rows in S is $\mathbf{0}$, (resp. $\mathbf{1}$), then no column has has exactly one one (resp. exactly two ones) in the rows of S. Let these events be $\mathcal{A}_{S,i}, i = 0, 1$. If the ones in each column were generated completely at random then

$$\begin{aligned}\mathbf{Pr}(\mathcal{A}_{S,0}) &= \left(\sum_{i\neq 1}^{k}\frac{\binom{s}{i}\binom{N-s}{k-i}}{\binom{N}{k}}\right)^M = \left(1-\frac{s\binom{N-1}{k-1}}{\binom{N}{k}}\right)^M\\ &= \left(1-\frac{ks}{N-s-k+1}\prod_{i=0}^{k-1}\left(1-\frac{s}{N-i}\right)\right)^M. \qquad (25)\\ \mathbf{Pr}(\mathcal{A}_{S,1}) &= \left(\sum_{i\neq 2}^{k}\frac{\binom{s}{i}\binom{N-s}{k-i}}{\binom{N}{k}}\right)^M = \left(1-\frac{\binom{s}{2}\binom{N-2}{k-2}}{\binom{N}{k}}\right)^M =\\ &= \left(1-\frac{k(k-1)s(s-1)}{2(N-s-k+2)(N-s-k+1)}\prod_{i=0}^{k-1}\left(1-\frac{s}{N-i}\right)\right)^M. \qquad (26)\end{aligned}$$

Now we can, for some r, bound the probability of $\mathcal{E}_{s,\boldsymbol{\alpha}}$ by the product of the RHS of (25) with M replaced by r and the RHS of (26) with M replaced by $M-r$. It follows therefore that, after ignoring conditioning on the event $\mathcal{B}$ that every row of $\mathbf{A}$ contains at least $\gamma k\sigma$ ones, we have

$$\mathbf{Pr}(\mathcal{E}_{s,\boldsymbol{\alpha}}) \leq \left(1-\frac{(1+o(1))k(k-1)s^2e^{-ks/N}}{2(N-s)^2}\right)^M \leq \left(1-\frac{k^2e^{-k}}{3}\right)^M \leq (1-e^{-k})^M.$$

So, in fact, taking account of $\mathcal{B}$, we have

$$\mathbf{Pr}(\mathcal{E}_{S,\boldsymbol{\alpha}} \mid \mathcal{B}) \leq \frac{(1-e^{-k})^M}{\mathbf{Pr}(\mathcal{B})}. \tag{27}$$

We need a lower bound for $\mathbf{Pr}(\mathcal{B})$. By (17) above, we have,

$$\mathbf{Pr}(\mathcal{B}) \geq_b \frac{1}{N^{1/2}}\left(1-e^{-\lambda}\sum_{i=0}^{k\gamma\sigma-1}\frac{\lambda^i}{i!}\right)^N \geq \frac{1}{N^{1/2}}\left(1-e^{-(1-\gamma)^2k\sigma/3}\right)^N.$$

Plugging this into (27) we see that for large k, since $(1-\gamma)^2\sigma \geq e^{5k}$,

$$\mathbf{Pr}(\mathcal{E}_{S,\boldsymbol{\alpha}} \mid \mathcal{B}) \leq (1-e^{-2k})^M \leq (1-e^{-2k})^{e^{5k}N}.$$

So,

$$\mathbf{Pr}(\exists S, |S| \geq Ne^{-k} : \mathcal{E}_{S,\boldsymbol{\alpha}}) \leq \sum_{s=Ne^{-k}}^{N} \binom{N}{s} (1-e^{-2k})^{e^{5k}N} = O(N^{-K}).$$

Finally, if $N/2 < |S| \leq N-1$, then the complement $\bar{S}$ of S is non-empty, and the rows S sum to $\boldsymbol{\alpha}$ if and only if the rows $\bar{S}$ sum to $\boldsymbol{\beta}-\boldsymbol{\alpha}$, where $\boldsymbol{\beta}$ is the row-sum of $\mathbf{A}$. But this probability is controlled by the cases above, since $1 \leq |\bar{S}| < N/2$. □

2.6. **The initial rows of $\mathbf{B}^{-1}$ have many, but not too many, ones.** We argue next that the rows of $\mathbf{B}^{-1}$ must contain many ones. Let $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{n_2}$ denote the rows of $\mathbf{B}^{-1}$. We consider its first row $\mathbf{r}_1$. Let $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{m_2}$ be the columns of $\mathbf{B}_1$. Then we must have $\mathbf{r}_1\mathbf{b}_1 = 1$ and $\mathbf{r}_1\mathbf{b}_i = 0$ for $i = 2, 3, \ldots, m_2$. Let this event be $\mathcal{E}_0$ and suppose that $\mathbf{r}_1$ has s ones. Then, for $\mathcal{E}_0$ to occur there must be s rows of $\mathbf{B}_1$ whose sum is $(1, 0, 0, \ldots, 0)$.

We apply Lemma 2.1 to $\mathbf{B}_1$ with $N = n_2, M = m_2, \gamma = \frac{n_2}{10m_2} \leq \frac{1}{2}$ and $\boldsymbol{\alpha} = (1, 0, 0, \ldots, 0)$. We consider case (a) and we assume that $s \leq s_0 = n_2e^{-k}$. In which case we find that

$$\mathbf{Pr}(\mathcal{E}_0) \leq_b n^{1/2} \sum_{s=2}^{s_0} \left(\frac{se^{2k/3}}{n_2}\right)^{ks/30} = O(n^{-k/50}). \tag{28}$$

Now suppose that $\mathbf{r}_i$ has $\beta_i n_2$ ones. We can assume from (28) that

$$\beta_i \geq \varepsilon_0 n_2 \text{ where } \varepsilon_0 = e^{-k}. \tag{29}$$

We also need a bound on $1-\beta_i$. Again consider $\mathbf{r}_1$. Suppose that this has at least $n_2(1-\varepsilon_0)$ ones in positions S. Now since each column of $\mathbf{B}_1$ has exactly k ones, we know that the sum of the rows of $\mathbf{B}_1$ is either $\mathbf{0}$ (if k is even) or $\mathbf{1}$=(1,1,...,1) (if k is odd. (We take care of k even, even though the assumption is still that k is odd.) Thus the $n_2 - s$ rows of $\mathbf{B}_1$ corresponding to $[n_2] \setminus S$ will sum to $(1, 0, 0, \ldots, 0)$ or $(0, 1, 1, \ldots, 1)$ according as k is even or odd. We can apply Lemma 2.1 once more. This deals with all rows because the probability in (28) is bounded by $o(n^{-1})$.

Remark 2.2. *We see that if we fix a positive integer K and if k is sufficiently large, then $\sum_{i\in I} \mathbf{r}_i$ contains at least s_0 ones for all $|I| \leq K$. This is because each such I gives us an $\boldsymbol{\alpha}$ with only $|I|$ ones. There are $O(n^K)$ such $\boldsymbol{\alpha}$ and the probability bound in* (28) *will be small enough to deal with all such I if $K < k/50$.*

2.7. **A few rows of $\mathbf{B}^{-1}$ are not enough to cover $[n_2]$.** Let $S_i, i \in [n_2]$ be the indices of the columns where row i of $\mathbf{B}^{-1}$ has a one. We will apply the following lemma to the complements of the S_i's. In which case we will have $N = n_2$ and $X_i = [n_2] \setminus S_i$.

Lemma 2.3. *Let $X_1, X_2, \ldots, X_N \subseteq [N]$ satisfy $|X_i| \geq \delta N$. Let r be a fixed positive integer independent of N. If N is sufficiently large, then there exists a set $I \subseteq [N], |I| = r$ and $s = \lceil \log_2 r \rceil$ such that $\left|\bigcap_{i\in I} X_i\right| \geq \delta_s N/2r$. Here $\delta_0 = \delta$ and $\delta_{i+1} = \delta_i^2/4$ for $i \geq 0$.*

Proof We will assume that $r = 2^s$ is a power of two. For general r we take the smallest power of two greater than r. This will explain the extra factor of two in the denominator in our lower bound on $\left|\bigcap_{i\in I} X_i\right|$.

We will prove this by induction on s. As a base case, consider $s = 1$. Now suppose that for some $t \geq 2$ we find that $|X_t \cap X_i| \leq \delta N/(2t)$ for all $i < t$. This implies that $\left|X_t \setminus \bigcup_{i=1}^{t-1} X_i\right| \geq \delta N/2$ and so $\left|\bigcup_{i=1}^{t} X_i\right| \geq t\delta N/2$. This process must stop after $2/\delta$ steps and our induction on s has a base case, i.e. there exists $i, t \leq 2/\delta$ such that $|X_i \cap X_t| \geq \delta^2 N/4$.

Suppose that for some s we can we can find $\{i_1, i_2, \ldots, i_{2^s}\} \subseteq [\prod_{i=1}^{s}(2/\delta_i)]$ such that $|Y_1| \geq \delta_s N$ where $Y_1 = \bigcap_{j=1}^{2^s} X_{i_j}$. Assuming N is sufficiently large, we can generate a sequence $Y_1, Y_2, \ldots, Y_{2/\delta_s}$ where (i) $|Y_i| \geq \delta_s N$ for $i = 1, 2, \ldots, 2/\delta_s$ and (ii) each Y_i is the intersection of 2^s distinct X_j and (iii) no X_j appears in more than one of these intersections. Applying the argument that gave us the base case we see that there exists $i, t \leq 2/\delta_s$ such that $|Y_i \cap Y_t| \geq \delta_{s+1} N$. □

Putting $X_i = [n_2] \setminus S_i$ for $i \in [n_2]$ we see that we can find for any r, a set of rows, such that there are $\Omega(n)$ columns without a one in the union of the rows.

2.8. **Constructing a representative matrix.** We now consider the construction of the partition $A_0, A_1, \ldots, A_\ell$ in Section 2.1. We choose a constant K and consider an arbitrary set R of K rows of $\mathbf{B}^{-1}$. Let $\varepsilon_1 = 2^{-2K}\varepsilon_0$, where $\varepsilon_0 = e^{-k}n_2$, as in (29), and consider the $K \times 2^K$ matrix $\mathbf{D}$ with entries in $\{0, 1\}$. The ith row $\mathbf{u}_i$ of $\mathbf{D}$ is associated with set S_i and the columns of $\mathbf{D}$ are indexed by $\boldsymbol{\sigma} = (\xi_1, \xi_2, \ldots, \xi_K) \in 2^{[K]}$ and they are associated with an atom $\mathbf{a}_{\boldsymbol{\sigma}} = \bigcap_{j=1}^{K} S_j^{\xi_j}$ in the Boolean algebra $\mathcal{B}_R$ generated by the sets S_i. Here $\xi_j = \xi_j(\boldsymbol{\sigma}) = 0, 1$ and $S_j^1 = S_j, S_j^0 = \bar{S}_j = [n_2] \setminus S_j$. The columns run over the 2^K sequences $\{0, 1\}^K$. For each $j \in [n_2]$ there is a unique $\boldsymbol{\sigma} = \boldsymbol{\sigma}(j)$ such that $j \in \mathbf{a}_{\boldsymbol{\sigma}}$ i.e. the $\mathbf{a}_{\boldsymbol{\sigma}}$ partition $[n_2]$. Further, if $S_{\boldsymbol{\sigma}} = \bigcap_{i=1}^{K} S_i^{\xi_i}$ then S_i is partitioned into the parts $S_{\boldsymbol{\sigma}}$ such that $\xi_i(\boldsymbol{\sigma}) = 1$.

Row i of $\mathbf{D}$ contains a one in position $\boldsymbol{\sigma}$ if $\xi_i(\boldsymbol{\sigma}) = 1$ and $|\mathbf{a}_{\boldsymbol{\sigma}}| \geq \varepsilon_1 n$. Otherwise, row i of $\mathbf{D}$ contains a zero in position $\boldsymbol{\sigma}$. We now claim that $\mathbf{D}$ has row rank K.

Fix some $\emptyset \neq I \subseteq [K]$ and let $\mathbf{r}^I = \sum_{i\in I} \mathbf{r}_i$ and $S_\oplus = \left\{j : \mathbf{r}_j^I = 1\right\}$ and suppose that $\mathbf{r}^I = \mathbf{0}$. Note that Lemma 2.1 and Remark 2.2 means that we can assume that $|S_\oplus| \geq \varepsilon_0 n$. Now let $\boldsymbol{\eta} = \sum_{i\in I} \mathbf{u}_i$ and $S_{\boldsymbol{\eta}} = \bigcup_{\eta_{\boldsymbol{\sigma}}=1} S_{\boldsymbol{\sigma}}$. We have

$$|S_{\boldsymbol{\eta}}| \geq |S_\oplus| - 2^K\varepsilon_1 n \geq \varepsilon_0 n - 2^K\varepsilon_1 n > 0.$$

Explanation: If an entry $u_{i,\boldsymbol{\sigma}} = 1$ this means (among other things) that $\ell \in S_i$ for all $\ell \in \mathbf{a}_{\boldsymbol{\sigma}}$ and thus $r_{i,\ell} = 1$ for all $\ell \in \mathbf{a}_{\boldsymbol{\sigma}}$. Thus, $S_{\boldsymbol{\eta}}$ is equal to $S_\oplus$ minus sets of the form $S_{\boldsymbol{\sigma}}$ where (i) $\xi_i(\boldsymbol{\sigma}) = 1$ for an odd number of $i \in I$ and (ii) $|S_{\boldsymbol{\sigma}}| \leq \varepsilon_1 n$.

It follows that there exists $\boldsymbol{\sigma}$ such that $\boldsymbol{\eta}_{\boldsymbol{\sigma}} = 1$ i.e. $\boldsymbol{\eta} \neq \mathbf{0}$. Because I is arbitrary, we see that $\mathbf{D}$ has full row rank.

2.9. **Finishing the proof of Theorem 1.1.** Recall that the minor M can be represented by a $p\times q$ matrix $\mathbf{R}_M$. Let R be a set of row indices where (i) $|R|=p$ and (ii) $\left|[n]\setminus\bigcup_{i\in R}S_i\right|\geq \delta_s n, s=\lceil\log_2 p\rceil$ (see Lemma 2.3). Suppose that $\mathbf{c}$ is a column of $\mathbf{A}_m$ not involved in the construction of $\mathbf{B}$. We say that $\mathbf{c}$ is a *candidate* column if $c_j=0$ whenever $j\in\mathbf{a}_{\boldsymbol\sigma}$ for which $|\mathbf{a}_{\boldsymbol\sigma}|<\varepsilon_1 n$. Next let $c_{\boldsymbol\sigma}=\sum_{j\in\mathbf{a}_{\boldsymbol\sigma}}c_j$. If $\mathbf{c}$ is a candidate column then $\mathbf{r}_i\cdot\mathbf{c}=\mathbf{u}_i\cdot\mathbf{c}_R$ where $\mathbf{c}_R$ is the column vector with components $c_{\boldsymbol\sigma},{\boldsymbol\sigma}\in 2^{[p]}$. (Remember that $\mathbf{r}_i$ is row i of $\mathbf{B}^{-1}$ and that $\mathbf{u}_i$ is row i of $\mathbf{D}$.) For a column $\mathbf{x}$ of $\mathbf{A}_m$, let $\phi_R(\mathbf{x})$ be the column $\mathbf{x}$ restricted to the p rows of R. Let $\mathbf{c}_1$ be the first column of the target matrix M and let $\mathbf{c}$ be a random candidate column. Let $\mathbf{v}$ satisfy $\mathbf{D}\mathbf{v}=\mathbf{m}_1$ and $v_{\boldsymbol\sigma}=0$ if $|\mathbf{a}_{\boldsymbol\sigma}|<\varepsilon_1 n$. Assume also that $\mathbf{v}$ has at most p ones and that $k\geq p$. There are always such solutions. Then we have

$$\mathbf{Pr}(\phi_R(\mathbf{B}^{-1}\mathbf{c})=\mathbf{m}_1)=\mathbf{Pr}(\mathbf{r}_i\cdot\mathbf{c}=\mathbf{u}_i\cdot\mathbf{c}_R=m_{1,i},i\in R)\geq\mathbf{Pr}(\mathbf{c}_R=\mathbf{v})\geq\varepsilon_1^k. \tag{30}$$

Explanation of second inequality: Let $J=\{{\boldsymbol\sigma}:v_{\boldsymbol\sigma}=1\}$. Each index ${\boldsymbol\sigma}$ corresponds to a set $\mathbf{a}_{\boldsymbol\sigma}$ of size at least $\varepsilon_1 n$. Now we will have $\mathbf{c}_R=\mathbf{v}$ if column $\mathbf{c}$ has a single one in each $\mathbf{a}_{\boldsymbol\sigma},{\boldsymbol\sigma}\in J$ and its remaining ones $[n_2]\setminus\bigcup_{i\in R}S_i$. All of the sets where we need to place ones are of size at least $\varepsilon_1 n_2$ and (30) follows.

It follows from this that we can find a copy of M w.h.p. by examining a further ω random columns, where $\omega=\omega(n)\to\infty$, is arbitrary. This completes the proof of Theorem 1.1.

3. k EVEN

We now examine the adjustments needed for the case of k even. The problem here is that the rows of $\mathbf{A}_m$ now sum to zero and so we cannot construct $\mathbf{B}$ in quite the same way as for k odd.

Going back to Section 2.4 we define $\mathbf{L}_1,\mathbf{L}_2,\mathbf{L}_3$ in the same way, but now we can only say that w.h.p. the rank of $\mathbf{L}_1$ is $n_3^*=n_3-1$. So now we choose $i\in I_2$ such that if the matrix $\mathbf{L}_2^*=[\mathbf{B}_1^*:\mathbf{L}_3^*]$ is obtained from $\mathbf{L}_2$ by deleting row i then $\mathbf{L}_3^*$ has full row rank. We claim now that w.h.p. $\mathbf{L}_2^*$ also has full row rank. Suppose now that there exists $J\subseteq I_1\setminus\{i\}$ such that $\sum_{j\in J}\mathbf{a}_j=0$. Then we must have $J\cap(I_2\setminus\{i\})=\emptyset$. For otherwise, $\mathbf{L}_3^*$ does not have full row rank. But then $J\subseteq I_1\setminus I_2$ and by (13) we can assume that $|J|\leq n_2e^{-2k}$ and then we can apply Lemma 2.1(a) to J and $\mathbf{B}_1$ to get a contradiction w.h.p.

Then we let $\mathbf{B}^*$ be obtained from $\mathbf{B}_1$ by removing row i and then we can extend it to an $n_2^*\times n_2^*$ non-singular submatrix of $\mathbf{L}_2^*$. We need to argue that $\mathbf{B}^*$ has many ones and zeros in each row. For this we need to argue about sums of rows of the matrix $\mathbf{B}_1^*$ which has k or $k-1$ ones in each column and at least $k/10$ ones in each row. The row we deleted from $\mathbf{B}_1$ came by considering $\mathbf{L}_1$ which is independent of $\mathbf{B}_1$ and so the ones in each column of $\mathbf{B}_1^*$ are still randomly chosen subject to the row constraints. We can then argue via Lemma 2.1 that $(\mathbf{B}^*)^{-1}$ has many ones and zeros in each row.

We write

$$\mathbf{A}_m=\begin{bmatrix}\mathbf{B}^* & \mathbf{L}_1^* & \mathbf{C}^* & \mathbf{R}^*\\ \mathbf{u}_1 & \mathbf{u}_2 & 0 & \mathbf{u}_3\end{bmatrix}$$

where $[\mathbf{u}_1, \mathbf{u}_2, 0, \mathbf{u}_3]$ is row i and $\mathbf{C}^*$ comprises the unviewed random columns that appear where there is a zero in row i. $\mathbf{R}^*, \mathbf{u}_3$ comprise the rest of the matrix. Now let $\widehat{\mathbf{B}}$ be the $n_2 \times n_2$ matrix obtained from $\mathbf{B}^*$ by adding a column $\mathbf{e}_{n_2}$ and a row $\mathbf{e}_{n_3}^T$ where $\mathbf{e}_{n_2}$ has a unique one in position n_3.

The number of ones in a row of $\mathbf{X}$ is dominated by the binomial $Bin(n/4, k/n)$ and so w.h.p. the maximum number of ones in any row is $O(\log n)$. Then we write

$$\widehat{\mathbf{B}}^{-1}\mathbf{A}_m = \begin{bmatrix} \mathbf{I}_1 & 0 & \widehat{\mathbf{L}}_{1,1} & \widehat{\mathbf{C}}_{1,1} & \widehat{\mathbf{C}}_{1,2} & \widehat{\mathbf{R}}_1 \\ 0 & \mathbf{I}_2 & \widehat{\mathbf{L}}_{2,1} & 0 & \widehat{\mathbf{C}}_{1,2} & \widehat{\mathbf{R}}_2 \\ 0 & \mathbf{u}_{1,2} & \mathbf{u}_2 & 0 & 0 & \mathbf{u}_3 \end{bmatrix}.$$

Here we have split $\mathbf{u}_1$ into $[\mathbf{u}_{1,1}, \mathbf{u}_{1,2}]$ where $\mathbf{u}_{1,1} = 0$ and $\mathbf{u}_{1,2}$ is an all ones vector of dimension $O(\log n)$. And then the matroid $\mathcal{M}$ associated with $\mathbf{A}_m$ has a minor isomorphic to M if $\mathbf{R}_M$ appears in $\widehat{\mathbf{C}}_{1,1}$. The argument for this is covered by the case k odd, concentrating on the sub-matrix $[\mathbf{I}_1 : \widehat{\mathbf{C}}_{1,1}]$.

4. Further Questions

We have shown that $\mathbf{A}_m$ contains a copy of an arbitrary fixed binary matroid as a minor under the assumption that $k, m/n$ are sufficiently large. It would be of interest to reduce k, perhaps to three, and to get precise estimates for the number of columns needed for some fixed matroid, the Fano plane for example. In this way we could perhaps get the precise number of columns needed to make the random matroid associated with $\mathbf{A}_m$, non-graphic or non-regular, w.h.p. Behavior of random matroids over fields other than $\mathbf{GF}_2$ are also an interesting target.

Acknowledgement: We thank Peter Nelson for for helpful discussions.

References

[1] J. Altschuler and E. Yang, Inclusion of Forbidden Minors in Random Representable Matroids, arXiv:1507.05332 [math.CO].
[2] J. Aronson, A.M. Frieze and B. Pittel, Maximum matchings in sparse random graphs: Karp-Sipser revisited, *Random Structures and Algorithms* 12 (1998) 111-178.
[3] B. Bollobás, Random Graphs, First Edition, Academic Press, London 1985, Second Edition, Cambridge University Press, 2001.
[4] N. Bansal, R.A. Pendavingh and J.G. van der Pol, On the number of matroids, *Combinatorica* 35 (215) 253-277.
[5] O. Dubois and J. Mandler, The 3-XORSAT Threshold, *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science* (2002) 769-778.
[6] C. Cooper, On the rank of random matrices, *Random Structures and Algorithms* 16 (2000) 209-232.
[7] C. Cooper, The cores of random hypergraphs with a given degree sequence, *Random Structures and Algorithms* 25 (2004) 353-375.
[8] A.M. Frieze and M. Karoński, Introduction to Random Graphs, Cambridge University Press 2015.
[9] J. Geelen, J.P.S. Kung, and G. Whittle, Growth rates of minor-closed classes of matroids, *Journal of Combinatorial Theory, Series B* **99** (2009) 420-427.
[10] S. Janson, T. Łuczak and A. Ruciński, Random Graphs, John Wiley and Sons, New York, 2000.

[11] M. Kahle, Topology of random simplicial complexes: a survey, *AMS Contemporary Volumes in Mathematics*,
[12] D. Kelly and J. Oxley, On random representable matroids, *Studies in Applied Mathematics* 71 (1984) 181-205.
[13] J.P.S. Kung, The long-line graph of a combinatorial geometry. II. Geometries representable over two fields of different characteristics, *Journal of Combinatorial Theory, Series B* **50** (1990) 41–53.
[14] M. Molloy, Cores in random hypergraphs and Boolean formulas, *Random Structures and Algorithms* 27 124-135 (2005).
[15] N. Linial and R. Meshulam, Homological connectivity of random 2-complexes, *Combinatorica* 26 (2006) 475-487.
[16] J. Oxley, Matroid Theory, Second Edition, Oxford University Press, New York, 2011.
[17] J. Oxley, L. Lowrance, C. Semple and D. Welsh, On properties of almost all matroids, *Advances in Applied Mathematics* 50 (2013) 115-124.
[18] M. Molloy, Cores in random hypergraphs and random formulas, *Random Structures and Algorithms* 27 (2005) 124-135.
[19] P. Nelson, Almost all matroids are nonrepresentable, *Bulletin of the London Mathematical Society* 50 (2018) 245-248.
[20] R. Pendavingh and J. van der Pol, On the number of bases of almost all matroids, arXiv preprint arXiv:1602.04763.
[21] W.T. Tutte, A homotopy theorem for matroids. I, II, *Transactions of the American Mathematical Society* 88 (1958) 144174.
[22] D. Welsh, Matroid Theory, Academic Press, 1976.

Appendix A. Proof of (17)

Let $\boldsymbol{\rho}$ be the vector of row counts in $\mathbf{X}$ and let A, B be arbitrary positive integers,

$$S = \left\{ \boldsymbol{\rho} \in [M]^N \,\middle|\, \sum_{1\le j\le N} \rho_j = A \text{ and } \forall j,\ \rho_j \ge B \right\}.$$

Fix $\vec{\xi} \in S$. Then, if $\mathbf{Pr}_1$ refers to a random choice from S,

$$\mathbf{Pr}(\boldsymbol{\rho} = \vec{\xi}) = \left(\frac{M!}{\xi_1!\xi_2!\dots\xi_N!} \right) \Big/ \left(\sum_{\boldsymbol{\rho}\in S} \frac{M!}{\rho_1!\rho_2!\dots\rho_N!} \right).$$

On the other hand, if $\mathbf{Pr}_2$ refers to a random choice via independent Poisson,

$$\begin{aligned}
\mathbf{Pr}_2\left(\boldsymbol{\rho} = \vec{\xi} \,\middle|\, \sum_{1\le j\le N} \rho_j = A \right) &= \left(\frac{\prod_{1\le j\le N} \lambda^{\xi_j}}{f_B(\lambda)\xi_j!} \right) \Big/ \left(\sum_{\boldsymbol{\rho}\in S} \prod_{1\le j\le N} \frac{\lambda^{\rho_j}}{f_B(\lambda)\rho_j!} \right) \\
&= \left(\frac{f_B(\lambda)^{-N}\lambda^s}{\xi_1!\xi_2!\dots\xi_N!} \right) \Big/ \left(\sum_{\boldsymbol{\rho}\in S} \frac{f_B(\lambda)^{-N}\lambda^s}{\rho_1!\rho_2!\dots\rho_N!} \right) \\
&= \mathbf{Pr}_1(\boldsymbol{\rho} = \vec{\xi}).
\end{aligned}$$

□