MINORS OF A RANDOM BINARY MATROID

COLIN COOPER, ALAN FRIEZE, AND WESLEY PEGDEN

ABSTRACT. Let $\mathbf{A} = \mathbf{A}_{n,m,k}$ be a random $n \times m$ matrix over $\mathbf{GF}_2$ where each column consists of k randomly chosen ones. Let M be an arbirary fixed binary matroid. We show that if m/n and k are sufficiently large then as $n \to \infty$ the binary matroid induced by $\mathbf{A}$ contains M as a minor.

1. INTRODUCTION

There is by now a vast and growing literature on the aymptotic properties of random combinatorial structures. First and foremost in this context are Random Graphs and Hypergraphs, see [3], [7] and [8] for books on this subject. Random groups 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 [12]. See Kahle [9] for a recent survey. Another area of interest in this vein is that of Random Matroids. This paper concerns one aspect of these. Basically we see that two models of a random matroid have been considered so far. In the first we choose a matroid uniformly at random from the set of all matroids with n elements, see for example Oxley, Semple, Warshauer and Welsh [14]. In the second we only consider representable matroids and choose an $m \times n$ matrix with entries chosen independently and uniformly from $\mathbf{GF}_q$, see for example [11]. 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 random graph.

The columns of $\mathbf{A}_{n,m,2}$ define a (random) binary matroid, actually a random graphic matroid. If we want to generalise this to random binary matroids, then one way at least is to replace the two ones in each column by k random ones to obtain the random matrix $\mathbf{A}_{n,m,k}$. 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 [16]. For more on the properties of matroids, see Oxley [13].

Date: January 6, 2017.

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.

A key question therefore in this probabilistic context concerns the number of random columns needed for the matroid $\mathcal{M} = \mathcal{M}_{n,m,k}$ associated with $\mathbf{A}_{n,m,k}$ to contain some fixed matroid as a minor. 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. $\mathcal{M}_{n,m,k}$ 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, 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.

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 is related to our theorem on taking $q = 2$ and $k = n(m)/2$. (We have reversed the roles of m, n from their statement.) 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 because we 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 $\mathbf{A}_m = \mathbf{A}_{n,m,k}$ of the binary matroid 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 let $\mathbf{R}_M = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_\mu]$ be a representation of M by a $\nu \times \mu$ matrix. Assume without loss of generality that $\mathbf{R}_M$ has full row-rank ν.

Let $n_1 = n, m_1 = n/4$. Denote the matrix $n_1 \times m_1$ matrix consisting of the first m_1 columns of $\mathbf{A}_m$ by $\mathbf{X}$. It follows from Theorem 1 of Cooper [5] that w.h.p. the columns of $\mathbf{X}$ are linearly independent.

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

We extend $\mathbf{B}_1$ to an $n_4 \times n_4$ non-singular submatrix $\mathbf{B}$ of $\mathbf{A}_m$. We argue that w.h.p. the rows of $\mathbf{B}^{-1}$ have between $\frac{1}{2}n_4 e^{-k}$ and $n - \frac{1}{2}n_4 e^{-k}$ ones.

We let $\widehat{\mathbf{A}}$ be the submatrix of $\mathbf{A}_m$ that has ones only in rows corresponding to the rows of $\mathbf{B}$. Now suppose that $\widehat{\mathbf{A}} = [\mathbf{B} : \mathbf{M}]$ and consider the matrix $[\mathbf{I} : \mathbf{M}_1 = \mathbf{B}^{-1}\mathbf{M}]$. 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 ν rows and the first μ columns of $\mathbf{M}_1$. Then we get M as a minor by deleting the first ν columns of $\mathbf{B}$ and the last $m - n_4 - \mu$ columns of $\mathbf{M}$ and contracting the last $n_4 - \nu$ 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).

Contraction is simple if we assume that the columns S are a subset of the columns of an identity matrix $\mathbf{I} = \mathbf{I}_{n_4}$. In view of this, we pre-multiply $\widehat{\mathbf{A}}$ by $\mathbf{B}^{-1}$ to obtain $\widehat{\mathbf{A}}_1$. Pre-multiplying by a non-singular matrix does not change the underlying matroid, seeing as column dependence/indepdendence is preserved. We can assume that the first n_4 columns form the $n_4 \times n_4$ identity matrix $\mathbf{I}$. If we contract a set S of the columns of $\mathbf{I}$, then the contracted matroid can be represented by simply 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 $\widehat{\mathbf{A}}_2$ are independent in $\widehat{\mathbf{A}}_2$ if and only if the corresponding columns in $\widehat{\mathbf{A}}_1$ are independent of the columns in S.

To prove that $\mathbf{R}_M$ appears, we will consider $\mathbf{B}^{-1}\mathbf{c}$ where $\mathbf{c}$ is a random column of $\widehat{\mathbf{A}}$. We argue that $\mathbf{Pr}(\mathbf{B}^{-1}\mathbf{c} = \mathbf{c}_1) = \Omega(1)$. 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.

We now give a detailed proof of Theorem 1.1.

2.2. **Building $\mathbf{B}_1$.** Consider the k-uniform hypergraph H_1 induced by the first $n/4$ columns of $\mathbf{X}$. I.e. the hypergraph with a vertex for each row and where each edge $e_i, i \leq n/4$ 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.

We use some results on cores of random k-uniform hypergraphs (see e.g. Cooper [6] or Molloy [10]). Let $c = km_1/n_1 = k/4$, and let x be the greatest solution to

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

Then w.h.p.,

$$n_2 = |C_1| \approx n_1 \left(1 - e^{-x} \sum_{i=0}^{k/10-1} \frac{x^i}{i!}\right), \tag{2}$$

and

$$m_2 = |E(C_1)| \approx m_1 \left(\frac{x}{c}\right)^{k/(k-1)}. \tag{3}$$

Here, $A(x) \approx B(x)$ is short for $A(x) = (1 + o(1))B(x)$ as $x \to \infty$.

We will first argue that for k large we have

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

The upper bound follows directly from the definition (1). 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 $\theta=1,2$,

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

and

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

Thus $S(k/5)>0$ for k large. As $S(k/4)<0$, the lower bound in (4) follows from the continuity of $S(x)$. It then follows from (2) that w.h.p.

$$n_2=|V(C_1)|\geq n_1\left(1-\frac{1}{k}\right).\tag{5}$$

Similarly, using (3) along with $c=k/4$ and $x>k/5$ from (4) gives

$$m_2\gtrsim\frac{n_1}{4}\left(\frac{4}{5}\right)^{k/(k-1)}\geq\frac{n_2}{5},\tag{6}$$

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 has 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.

2.3. **Extend $\mathbf{B}_1$ 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 $\mathbf{L}$. Now let $0<\gamma<1$ 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 γLk core. Using [6], [10] once again we see that we have to let x be the greatest solution to

$$c=Lk=\frac{x}{\left(1-e^{-x}\sum_{i=0}^{\gamma Lk-2}\frac{x^i}{i!}\right)^{k-1}}.\tag{7}$$

Then w.h.p.,

$$n_3=|C_2|\approx n\left(1-e^{-x}\sum_{i=0}^{\gamma Lk-1}\frac{x^i}{i!}\right).\tag{8}$$

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

$$\frac{(1+\gamma)Lk}{2}\leq x\leq Lk.\tag{9}$$

The upper bound follows directly from the definition (7). To prove the lower bound let now

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

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

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

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

$$e^{-x}\sum_{i=0}^{\gamma Lk-\theta}\frac{x^i}{i!} \leq \frac{2\left(\eta e^{1-\eta}\right)^{\gamma Lk}}{(1-\xi)\sqrt{\gamma Lk\pi}}.$$

Thus $S(\frac{(1+\gamma)Lk}{2}) > 0$ and the lower bound in (9) follows by continuity.

It then follows from (2) that w.h.p.

$$n_3 = |C_2| \geq n_2\left(1 - \frac{1}{k}\right). \tag{10}$$

Similarly, using (3) along with $c = Lk$ and $x \geq (1+\gamma)Lk/2$ in (4) gives us that the number m_3 of edges in C_2 satisfies

$$m_3 \gtrsim Ln_2\left(\frac{1+\gamma}{2}\right)^{k/(k-1)} \geq \frac{Ln_2}{2}, \tag{11}$$

for k large.

We argue next that w.h.p. the matrix $\mathbf{L}_1$ induced by C_2 has rank $N^* = N - 1_{k \text{ is even}}$. For this we rely on the following lemma, which we will need for several purposes:

Lemma 2.1. *Let $\mathbf{A}$ be an $N \times M$ matrix over $\mathbf{GF}_2$ chosen uniformly randomly over 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 $\mathbf{GF}_2^M$. Let $\mathcal{E}_{s,\boldsymbol{\alpha}}$ be the event that there exists a set S of rows with $|S| = s$ whose sum is $\boldsymbol{\alpha}$. Then,*

(a) If $2 \leq s \leq Ne^{-k}$ then

$$\mathbf{Pr}(\mathcal{E}_{s,\boldsymbol{\alpha}}) \leq_b M^{1/2}\left(\frac{se^{2k/3}}{N}\right)^{\gamma k\sigma s/3}. \tag{12}$$

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

$$\mathbf{Pr}(\exists Ne^{-k} < s \leq N^* : \mathcal{E}_{s,\boldsymbol{\alpha}}) = O(e^{-\xi N}) \qquad \text{for some fixed } \xi > 0. \tag{13}$$

Here the notation $A \leq_b B$ is short for $A = O(B)$.

We apply the lemma to $\mathbf{L}_1$ by taking $N = n_3, M = m_3$ and γ equal to $1/2$. Assume that $L \geq 4e^{5k}$. Now if $\mathbf{L}_1$ does not have full row rank, then $\mathcal{E}_{s,\boldsymbol{\alpha}}$ occurs with $\boldsymbol{\alpha} = \mathbf{0}$ for some

$2 \leq s \leq n_3$. But Lemma 2.1 implies that

$$\mathbf{Pr}(\exists 2 \leq s \leq n_3^* : \mathcal{E}_{s,\boldsymbol{\alpha}} \text{ occurs}) \leq_b e^{-\xi n} + n^{1/2} \sum_{s=2}^{e^{-k} n_3} \left(\frac{s e^{2k/3}}{n_3} \right)^{kLs/12} = O(n^{-kL/10}). \tag{14}$$

The $e^{-\xi n}$ in (14) comes from (13).

So, w.h.p. we have found an $n_3 \times m_3$ matrix $\mathbf{L}_1$ of full row rank. In the case k is even we remove an arbitrary row and replace n_3 by $n_3 - 1$. Now $\mathbf{B}_1$ and $\mathbf{L}_1$ are of full row rank but they may not have exactly the same set of row indices. So we let $\mathbf{L}_2$ be the matrix $[\mathbf{B}_1 : \mathbf{L}_1]$ where the row indices of $\mathbf{L}_2$ are the union of the row indices of $\mathbf{B}_1$ and $\mathbf{L}_1$. If a row index occurs only in one of the two matrices, then we pad out the other matrix with a row of zeros. $\mathbf{L}_2$ has full row rank, n_4 say and it is a submatrix of $\mathbf{A}_m$. We obtain $\mathbf{B}$ as an arbitrary extension of $\mathbf{B}_1$ (plus additional padded zero rows) to an $n_4 \times n_4$ non-singular sub-matrix of $\mathbf{L}_2$. After this we order the rows of $\mathbf{B}$ so that the rows of $\mathbf{B}_1$ come first.

2.4. **Proof of Lemma 2.1.** We first deal with small s. Suppose that $2 \leq s \leq N e^{-k}$. 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 $S \setminus T$.

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

Explanation: We sum over sets S and then fix the number of ones d_j in column $j \in [M]$ 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 allow the ones in a column to be chosen with repetition. More precisely, for the purposes of the lemma, the ones in a column of $\mathbf{A}_m$ will be represented by a random member of $[N]^k$. Now the probability that any one column contains a repetition in this model is bounded above by k^2/N and so the probablity that there are no repetitions is bounded below by $\left(1 - \frac{k^2}{N}\right)^M \geq e^{-2k^2\sigma} = \Omega(1)$, assuming that $\sigma = O(1)$. It follows that events that occur w.h.p. in the relaxed model, occur w.h.p. in the model without repetition. The parameter S_j in $\mathcal{E}_{j,S,S_j}$ is now to be interpreted as a multigraph or equivalently as a member of S^{d_j}.

The degrees (row-sums) $\rho_1, \rho_2, \ldots, \rho_N$ will be independent Poisson, subject to $\rho_j \geq \gamma k \sigma, j \in [N]$ 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

$$\mathbf{Pr}(\rho = l) = \frac{\lambda^l}{l! f_{\gamma k \sigma}(\lambda)} \text{ where } f_l(\lambda) = e^{\lambda} - \sum_{i=0}^{l-1} \frac{\lambda^i}{i!}. \tag{16}$$

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

$$\frac{\lambda f_{\gamma k \sigma - 1}(\lambda)}{f_{\gamma k \sigma}(\lambda)} = k\sigma. \tag{17}$$

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

$$\frac{k\sigma}{2}\leq\lambda\leq k\sigma \text{ and } f_{\gamma k\sigma}(\lambda)\geq e^{\gamma k\sigma/2}. \tag{18}$$

The upper bound in (18) 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 (17) 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 degrees $\rho_1=\theta_1,\rho_2=\theta_2,\ldots,\rho_N=\theta_N$. If $d_1+d_2+\cdots+d_M=\ell$ then

$$\mathbf{Pr}\left(\bigcap_{j=1}^{N}\mathcal{E}_{j,S_j,S}\right)\leq\frac{(kM-\ell)!}{(kM)!}\prod_{j=1}^{M}\prod_{i\in S_j}(\theta_i k)\leq\frac{e^{\ell^2/kM}}{M^\ell}\prod_{i\in S}\theta_i^{\theta_i}. \tag{19}$$

Explanation of (19)**:** In this model, the probability that Γ contains the edge (i,j), given that it contains l prior edges is bounded by $\frac{\theta_i k}{kM-l}$. This is a bound on the expected number of edges between i and j, given the previous edges exist. The final inequality comes from using Stirling's inequality.

Next let

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

and

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

Note that

$$|D_\ell|\leq_b\binom{M+\ell/2-1}{\ell/2-1}\leq\frac{M^{\ell/2}e^{\ell^2/4M}}{(\ell/2)!} \text{ and } |E_\ell|=\binom{\ell-\gamma k\sigma s+s-1}{s-1}<2^\ell. \tag{20}$$

The first inequality in (20) 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

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

Plugging (19) into (15) we obtain,

$$\mathbf{Pr}(\mathcal{E}_{s,\boldsymbol{\alpha}}) \tag{21}$$

$$\leq\sum_{S\subseteq[N],|S|=s}\sum_{\ell=\gamma k\sigma s}^{kM}\sum_{\mathbf{d}\in D_\ell}\sum_{S_j\subseteq S,|S_j|=d_j}\sum_{\boldsymbol{\theta}\in E_\ell}\mathbf{Pr}(\rho_i=\theta_i,i\in S) \tag{22}$$

$$\times \frac{e^{\ell^2/kM}}{M^\ell}\prod_{i\in S}\theta_i^{\theta_i} \tag{23}$$

$$\leq_b M^{1/2}\sum_{S\subseteq[N],|S|=s}\ \sum_{\ell=\gamma k\sigma s}^{kM}\ \sum_{\mathbf{d}\in D_\ell}\ \sum_{S_j\subseteq S,|S_j|=d_j}\ \sum_{\boldsymbol{\theta}\in E_\ell}\prod_{i\in S}\frac{\lambda^{\theta_i}}{\theta_i! f_{\gamma k\sigma}(\lambda)}\frac{e^{\ell^2/kM}}{M^\ell}\prod_{i\in S}\theta_i^{\theta_i} \tag{24}$$

$$\leq_b M^{1/2}\binom{N}{s}\sum_{\ell=\gamma k\sigma s}^{kM}\ \sum_{\mathbf{d}\in D_\ell}\ \sum_{\boldsymbol{\theta}\in E_\ell}\frac{\lambda^\ell}{f_{\gamma k\sigma}(\lambda)^s}\left(\frac{se^{\ell/kM}}{M}\right)^\ell\prod_{i=1}^s\frac{\theta_i^{\theta_i}}{\theta_i!} \tag{25}$$

$$\leq_b M^{1/2}\binom{N}{s}\sum_{\ell=\gamma k\sigma s}^{kM}\ \sum_{\mathbf{d}\in D_\ell}\ \sum_{\boldsymbol{\theta}\in E_\ell}\frac{\lambda^\ell}{f_{\gamma k\sigma}(\lambda)^s}\left(\frac{se^{1+\ell/kM}}{M}\right)^\ell \tag{26}$$

$$\leq_b M^{1/2}\left(\frac{Ne}{s}\right)^s\sum_{\ell=\gamma k\sigma s}^{kM}\frac{M^{\ell/2}e^{\ell^2/4M}}{(\ell/2)!}\frac{(2k\sigma)^\ell}{e^{\gamma k\sigma s/2}}\left(\frac{se^{1+\ell/kM}}{M}\right)^\ell \tag{27}$$

$$\leq_b M^{1/2}\left(\frac{Ne}{s}\right)^s\sum_{\ell=\gamma k\sigma s}^{kM}\left(\frac{e^{k/3}\sigma s}{\ell^{1/2}M^{1/2}}\right)^\ell e^{-\gamma k\sigma s/2}, \tag{28}$$

since k is large. Now if u_ℓ is the ℓth root of the summand in (28) then

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

since $\gamma k>1$.

It now follows, since $\gamma k>1$ and the largest term in the sum in (28) is at $\ell=\gamma k\sigma s$, that

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

This completes the proof of part (a) of the lemma.

Assume now that $Ne^{-k}\leq s\leq N/2$. (If $|S|>N/2$ we apply the analysis to the complement of S.) 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

$$\mathbf{Pr}(\mathcal{A}_{S,0})=\left(\sum_{i\neq1}^k\frac{\binom{s}{i}\binom{N-s}{k-i}}{\binom{N}{k}}\right)^M=\left(1-\frac{ks}{N-s-k+1}\prod_{i=0}^{k-1}\left(1-\frac{s}{N-i}\right)\right)^M. \tag{29}$$

$$\mathbf{Pr}(\mathcal{A}_{S,1})=\left(\sum_{i\neq2}^k\frac{\binom{s}{i}\binom{N-s}{k-i}}{\binom{N}{k}}\right)^M=\left(1-\frac{k(k-1)s(s-1)}{2(N-s-k+2)(N-s-k+1)}\prod_{i=0}^{k-1}\left(1-\frac{s}{N-i}\right)\right)^M. \tag{30}$$

Now we can, for some r, bound the probability of $\mathcal{E}_{s,\boldsymbol{\alpha}}$ by the product of the RHS of (29) with M replaced by r and the RHS of (30) 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

$$\mathbf{Pr}(\mathcal{E}_{s,\boldsymbol{\alpha}}) \leq \left(1 - \frac{(1+o(1))k(k-1)s^2e^{-ks/N}}{2(N-s)^2}\right)^M \leq \left(1 - \frac{k^2e^{-3k}}{3}\right)^M \leq (1-e^{-4k})^M.$$

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

$$\mathbf{Pr}(\mathcal{E}_{S,\boldsymbol{\alpha}} \mid \mathcal{B}) \leq \frac{(1-e^{-4k})^M}{\mathbf{Pr}(\mathcal{B})}. \tag{31}$$

We need a lower bound for $\mathbf{Pr}(\mathcal{B})$. By (16) 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 (31) we see that for large k, since $(1-\gamma)^2\sigma \geq e^{5k}$,

$$\mathbf{Pr}(\mathcal{E}_{S,\boldsymbol{\alpha}} \mid \mathcal{B}) \leq (1-e^{-4k})^M \leq (1-e^{-4k})^{e^{5k}N}.$$

So,

$$\mathbf{Pr}(\exists S, |S| \geq Ne^{-k} : \mathcal{E}_{S,\boldsymbol{\alpha}}) \leq \sum_{s=Ne^{-k}}^{N}\binom{N}{s}(1-e^{-4k})^{e^{5k}N} = o(1).$$

□

2.5. **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 $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{n_4}$ 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 the case (a) and we assume that $s \leq s_0 = n_2e^{-k}$. In which case we find that

$$\mathbf{Pr}(\mathcal{E}_0) \leq_b n^{1/2}\sum_{s=2}^{s_0}\left(\frac{se^{2k/3}}{n_2}\right)^{ks/30} = O(n^{-k^2/50}). \tag{32}$$

Now suppose that $\mathbf{r}_i$ has $\beta_i n_4$ ones. We can assume from (32) that $\beta_i \geq \varepsilon_0 = e^{-k}n_2/n_4 \geq e^{-k}/2$. 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. 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 failure probability in (32) is $o(n^{-1})$.

Remark 2.2. *We see that if we fix a positive integer K then 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 true 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* (32) *will be small enough to deal with all such I if $K < k^2/50$.*

2.6. **A few rows of $\mathbf{B}^{-1}$ are not enough to cover** $[n_4]$**.** Let $S_i, i \in [n_4]$ 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 $N = n_4$ and $X_i = [n_4] \setminus S_i$.

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.

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 $X_i = [n_4] \setminus S_i$ for $i \in [n_4]$ 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.

2.7. **Constructing a representative matrix.** We choose a constant K and consider an arbitrary set R of K rows of $\mathbf{B}^{-1}$. Let $\varepsilon_1 = 2^{-2K}\varepsilon_0$ and consider the $K \times 2^K$ matrix D_R with entries in $\{0, 1\}$ and where row $\mathbf{u}_i$ is associated with set S_i and column $\boldsymbol{\sigma} = (\xi_1, \xi_2, \ldots, \xi_K) \in 2^{[K]}$ corresponds to 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 $S_j^1 = S_j, S_j^0 = \bar{S}_j = [n_4] \setminus S_j$. The columns run over the 2^K sequences $\{0, 1\}^K$. For each $\ell \in [n_4]$ there is a unique $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\ell)$ such that $\ell \in \mathbf{a}_{\boldsymbol{\sigma}}$. Thus 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 D_R 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 D_R contains a zero in position $\boldsymbol{\sigma}$. We claim now that D_R has row rank K.

Fix some $\emptyset \neq I \subseteq [K]$ and let $\mathbf{r}^I = \sum_{i \in I} \mathbf{r}_i$ and $S_\oplus = \left\{j : \mathbf{r}_j^I = 1\right\}$. 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 D_R has full row rank and therefore it contains a $K \times K$ non-singular submatrix D_R^*.

2.8. **Finishing the proof of Theorem 1.1.** Recall that the minor M can be represented by a $\nu \times \mu$ matrix $\mathbf{R}_M$. Let R be a set of row indices where (i) $|R| = \nu$ and (ii) $\left|[n] \setminus \bigcup_{i \in R} S_i\right| \geq \delta_s n, s = \lceil \log_2 \nu \rceil$ (see Lemma 2.3). Suppose that $\mathbf{c}$ is a column of $\mathbf{A}_m$. 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 $c_{\boldsymbol{\sigma}}, \boldsymbol{\sigma} \in 2^{[K]}$. (Remember that $\mathbf{r}_i$ is row i of $\mathbf{B}^{-1}$ and that $\mathbf{u}_i$ is row i of D_R.) For a column $\mathbf{x}$ of $\mathbf{A}_m$, let $\phi_R(\mathbf{x})$ be the column $\mathbf{x}$ restricted to the ν rows of R. Let $\mathbf{c}_1$ be the first column of the target matrix M and let $\mathbf{c}$ be a candidate column. Then,

$$\mathbf{Pr}(\phi_R(\mathbf{B}^{-1}\mathbf{c}) = \mathbf{c}_1) = \mathbf{Pr}(D_R \mathbf{c}_R = \mathbf{c}_1) \geq \mathbf{Pr}(\mathbf{c}_R = (D_R^*)^{-1}\mathbf{c}_1) = \Omega(1). \tag{33}$$

Explanation of inequality in (33)**:** This merely compares the event that some equations are solvable, to they are solvable in some specific way.

Explanation of $\Omega(1)$ **in** (33)**:** Let $\mathbf{v} = (D_R^*)^{-1}\mathbf{c}_1$ and let $J = \{\boldsymbol{\sigma} : v_{\boldsymbol{\sigma}} = 1\}$. Each index $\boldsymbol{\sigma}$ corresonds 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 $\mathbf{c}$ chooses an odd number of ones in each $\mathbf{a}_{\boldsymbol{\sigma}}, \boldsymbol{\sigma} \in J$ and any other ones $[n_4] \setminus \bigcup_{i \in I} S_i$. All of the sets where we need to place ones are of size $\Omega(n)$ and the $\Omega(1)$ in (33) follows.

It follows from this that we can find a copy 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. 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.

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] O. Dubois and J. Mandler, The 3-XORSAT Threshold, *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science* (2002) 769-778.

[5] C. Cooper, On the rank of random matrices, *Random Structures and Algorithms* 16 (2000) 209-232.

[6] C. Cooper, The cores of random hypergraphs with a given degree sequence, *Random Structures and Algorithms* 25 (2004) 353-375.

[7] A.M. Frieze and M. Karoński, Introduction to Random Graphs, Cambridge University Press 2015.

[8] S. Janson, T. Łuczak and A. Ruciński, Random Graphs, John Wiley and Sons, New York, 2000.

[9] M. Kahle, Topology of random simplicial complexes: a survey, *AMS Contemporary Volumes in Mathematics*,

[10] M. Molloy, Cores in random hypergraphs and boolean formulas, *Random Structures and Algorithms* 27 124-135 (2005).

[11] J. Oxley and D. Kelly, On random representable matroids, *Studies in Applied Mathematics* 71 (1984) 181-205.

[12] N. Linial and R. Meshulam, Homological connectivity of random 2-complexes, *Combinatorica* 26 (2006) 475-487.

[13] J. Oxley, Matroid Theory, Second Edition, Oxford University Press, New York, 2011.

[14] J. Oxley, L. Lowrance, C. Semple and D. Welsh, On properties of almost all matroids, *Advances in Applied Mathematics* 50 (2013) 115-124.

[15] M. Molloy, Cores in random hypergraphs and random formulas, *Random Structures and Algorithms* 27 (2005) 124-135.

[16] W.T. Tutte, A homotopy theorem for matroids. I, II, *Transactions of the American Mathematical Society* 88 (1958) 144174.

Appendix A. Proof of (16)

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

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

Fix $\vec{\xi} \in S$. Then,

$$\mathbf{Pr}(\mathbf{x} = \vec{\xi}) = \left(\frac{M!}{\xi_1! \xi_2! \dots \xi_N!} \right) \Big/ \left(\sum_{\mathbf{x} \in S} \frac{M!}{x_1! x_2! \dots x_N!} \right).$$

On the other hand,

$$\begin{aligned} \mathbf{Pr}\left((\rho_1, \rho_2, \dots, \rho_N) = \vec{\xi} \,\middle|\, \sum_{1 \le j \le N} \rho_j = A \right) &= \left(\frac{\prod_{1 \le j \le N} \lambda^{\xi_j}}{f_B(\lambda) \xi_j!} \right) \Big/ \left(\sum_{\mathbf{x} \in S} \prod_{1 \le j \le N} \frac{\lambda^{x_j}}{f_B(\lambda) x_j!} \right) \\ &= \left(\frac{f_B(\lambda)^{-N} \lambda^s}{\xi_1! \xi_2! \dots \xi_N!} \right) \Big/ \left(\sum_{\mathbf{x} \in S} \frac{f_B(\lambda)^{-N} \lambda^s}{x_1! x_2! \dots x_N!} \right) \\ &= \mathbf{Pr}(\mathbf{x} = \vec{\xi}). \end{aligned}$$

□