On the rank of a random binary matrix

Colin Cooper* Alan Frieze† Wesley Pegden‡

June 30, 2018

Abstract

We study the rank of the random $n \times m$ 0/1 matrix $\mathbf{A}_{n,m;k}$ where each column is chosen independently from the set $\Omega_{n,k}$ of 0/1 vectors with exactly k 1's. Here 0/1 are the elements of the field GF_2. We obtain an asymptotically correct estimate for the rank in terms of c, n, k, assuming that $m = cn/k$.

In addition, we assign i.i.d. $U[0,1]$ weights $X_{\mathbf{c}}, \mathbf{c} \in \Omega_{n,k}$ and let the weight of a set of columns C be $X(C) = \sum_{\mathbf{c} \in C} X_{\mathbf{c}}$. Let a *basis* be a set of $n - 1_{k \text{ even}}$ linearly independent columns. We obtain an asymptotically correct estimate for the minimum weight of a basis. This generalises the well-known result [7] for $k = 2$ viz. that the expected length of a minimum weight spanning tree tends to $\zeta(3)$.

1 Introduction

Let $\Omega_{n,k}$ denote the set of 0/1 vectors with exactly k 1's. Here 0/1 are the elements of the field GF_2. Let $\mathbf{A}_{n,m;k}$ be the random $n \times m$ 0/1 matrix where each column is chosen independently from the set $\Omega_{n,k}$. In a recent paper [6], we studied the binary matroid $\mathcal{M}_{n,m;k}$ induced by the columns of $\mathbf{A}_{n,m;k}$. We showed that for any fixed binary matroid M, there were 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 viewed this as a contribution to the theory of random matroids, [1], [2], [9], [11], [12].

In this paper we study another related apect of $\mathbf{A}_{n,m;k}$, viz. its rank and improve on results from Cooper [4].

Our results on rank enable us to give an expression, (5), for the solution value of the following optimization problem. Suppose that we assign i.i.d. $U[0,1]$ weights $X_{\mathbf{c}}, \mathbf{c} \in \Omega_{n,k}$ and let the

*Research supported in part by EPSRC grant EP/M005038/1
†Research supported in part by NSF Grant DMS1661063
‡Research supported in part by NSF grant DMS1363136

weight of a set of columns C be $X(C)=\sum_{\mathbf{c}\in C} X_{\mathbf{c}}$. Let a *basis* be a set of $n-1_{k \text{ even}}$ linearly independent columns. What is the expected weight $W_{n,k}$ of a minimum weight basis? When $k=2$ this amounts to estimating the expected length of a minimum weight spanning tree of K_n and this has the limiting value of $\zeta(3)$, Frieze [7].

Our result on the rank of $\mathbf{A}_{n,m;k}$ takes a little setting up. Let $H=H_{n,m;k}$ denote the random k-uniform hypergraph with vertex set $[n]$ and m random edges taken from $\binom{[n]}{k}$. There is a natural bijection between $\mathbf{A}_{n,m;k}$ and $H_{n,m;k}$ in which column $\mathbf{c}$ is replaced by the set $\{i : \mathbf{c}_i = 1\}$. The ρ-core of a hypergraph H (if it is non-empty) is the maximal set of vertices that induces a sub-hypergraph of minimum degree ρ. The 2-core $C_2=C_2(H)$ plays an important role in our first theorem.

1.1 Matrix Rank

Notation: We write $X_n \approx Y_n$ for sequences $X_n, Y_n, n\geq 0$ if $X_n=(1+o(1))Y_n$ as $n\to\infty$. We will use some results on the 2-core of random hypergraphs. We remind the reader that the size of of the 2-core has been asymptotically determined, see for example Cooper [5] or Molloy [10]; we recall the basic results here. To do so, we parameterise $m=cn/k$ and consider the equation

$$x=(1-e^{-cx})^{k-1}. \tag{1}$$

For $k\geq 3$, define $\widehat{c}_k$ by

$$\widehat{c}_k=\min\left\{c : x=(1-e^{-cx})^{k-1} \text{ has a solution } x_c\in(0,1]\right\}.$$

It is known that $c<\widehat{c}_k$ implies that $C_2=\emptyset$. If $c>\widehat{c}_k$, $c=O(\log n)$ then putting x_c equal to the largest solution to (1) in $[0,1]$ then q.s.[1]

$$\left||C_2|-n(x_c^{1/(k-1)}-cx_c+cx_c^{k/(k-1)})\right|\leq n^{3/4}. \tag{2}$$

$$\left||E(C_2)|-n(cx_c^{k/(k-1)}/k)\right|\leq n^{3/4}. \tag{3}$$

The above equations (1), (2) and (3) are usually proved for $c=O(1)$, but here we need $c=O(\log\log n)$and this is not a problem.

Let c_k^* be the value of c for which the 2-core has asymptotically the same number of vertices and edges. More precisely, we use (2) and (3) to define c_k^* by

$$c_k^* := \min\left\{c\geq\widehat{c}_k : x_c^{1/(k-1)}-cx_c+cx_c^{k/(k-1)}=\frac{cx_c^{k/(k-1)}}{k}\right\}. \tag{4}$$

Then let $m_k^*=c_k^* n/k$. We will prove,

Theorem 1.1. *W.h.p.*

$$\operatorname{rank}(\mathbf{A}_{n,m;k})\approx\begin{cases}|E(H)| & m<m_k^*.\\ |E(H)|-|E(C_2)|+|C_2| & m\geq m_k^*.\end{cases}$$

[1]A sequence $\mathcal{E}_n$ of events occurs *quite surely* (q.s.) if $\mathbf{Pr}(\neg\mathcal{E}_n)=O(n^{-C})$ for any constant $C>0$.

Note that when $k = 2$ we have $c_k^* = 0$ and the theorem follows from the fact that an isolated tree with t edges induces a sub-matrix of rank t in $\mathbf{A}_{n,m;k}$. We therefore concentrate on the case $k \geq 3$.

Using (2), we can express Theorem 1.1 directly in terms of c by

Corollary 1.2. *Suppose that $k \geq 3$ and $m = cn/k$. Then, w.h.p.*

$$\text{rank}(\mathbf{A}_{n,m;k}) \approx \begin{cases} m & c < c_k^*. \\ m - m{x_c}^{k/(k-1)} + n(x_c^{1/(k-1)} - cx_c + cx_c^{k/(k-1)}) & c \geq c_k^*. \end{cases} \tag{5}$$

We can also ask for the size of m needed to ensure full rank up to parity; i.e., rank

$$n^* = n - 1_{k \text{ even}}.$$

Theorem 1.3. *Suppose that $k \geq 3$.*

(i) Given a constant $A > 0$, there exists $\gamma = \gamma(A)$ such that for $m \geq \gamma n \log n$,

$$\mathbf{Pr}(\text{rank}(\mathbf{A}_{n,m;k}) < n^*) = o(n^{-A}).$$

(ii) If $m = n(\log n + c_n)/k$ then

$$\lim_{n \to \infty} \mathbf{Pr}(\text{rank}(\mathbf{A}_{n,m;k} = n^*)) = \begin{cases} 0 & c_n \to -\infty \\ e^{-e^{-c}} & c_n \to c \\ 1 & c_n \to +\infty. \end{cases}$$

We can easily modify the proof of part (ii) of Theorem 1.3 to give the following hitting time version. Suppose that we randomly order the columns of $\mathbf{A}_{n,M;k}$ where $M = \binom{n}{k}$. Let $\mathbf{M}_m$ denote the matrix defined by the first m columns in this order.

$$m_1 = \min\{m : \mathbf{M}_m \text{ has } n^* \text{ non-zero rows}\} \text{ and let } m^* = \min\{m : \mathbf{M}_m \text{ has rank } n^*\}.$$

Theorem 1.4. *$m_1 = m^*$ w.h.p.*

1.2 Minimum Weight Basis

The expression (5) enables us to estimate the expected optimal value to the minimum weight basis problem defined above. Suppose that we assign i.i.d. $U[0,1]$ weights $X_{\mathbf{c}}, \mathbf{c} \in \Omega_{n,k}$ to the $|\Omega_{n,k}| = \binom{n}{k}$ distinct 0/1 vectors with exactly k unit entries. The weight of a set of columns C is $X(C) = \sum_{\mathbf{c} \in C} X_{\mathbf{c}}$. Let $W_{n,k}$ be the minimum weight of any basis of $n^* = n - 1_{k \text{ even}}$ linearly independent columns, chosen from the $\binom{n}{k}$ column vectors $\mathbf{c} \in \Omega_{n,k}$, and where $\mathbf{c}$ has weight $X_{\mathbf{c}}$. Thus $\mathbf{A}_{n,p;k}$ consists of those column vectors with weight at most p.

We show in Section 3 below that if $W_{n,k}$ denotes the weight of a minimum weight basis then

$$\mathbf{E}(W_{n,k}) = \int_{p=0}^{1} (n^* - \mathbf{E}(\mathrm{rank}(\mathbf{A}_{n,p;k})))dp. \tag{6}$$

Corollary 1.2 and Theorem 1.3 can be substituted into (6) to yield an asymptotic formula for $W_{m,k}$.

Theorem 1.5. *Let $x = x(c)$ be the largest solution of $x = (1 - e^{-cx})^{k-1}$ in $(0, 1]$, then*

$$\frac{n^{k-2}}{(k-1)!}\mathbf{E}(W_{n,k}) \approx c_k^* \left(1 - \frac{c_k^*}{2k}\right) + \int_{c_k^*}^{\infty} \left(e^{-cx}\left(1 + \frac{(k-1)cx}{k}\right) - \frac{c}{k}(1-x)\right) dc \tag{7}$$

We note the remarkable fact that by a result of Frieze [7], the expression in (7) must equal $\zeta(3)$ for $k = 2$, for $c_2^* = 0$. We have numerically estimated the first few values as a function of k:

k	2	3	4	5	6	7	8	9	10
$\frac{n^{k-2}}{(k-1)!}\mathbf{E}(W_{n,k})$	$\zeta(3) \approx 1.202$	1.563	2.021	2.507	3.003	3.501	4.000	4.500	5.000

It appears the values are getting close to $k/2$ as k grows, and this is indeed the case.

Theorem 1.6. *For $k \geq 3$, and some ε_k, $|\varepsilon_k| \leq 5$,*

$$\lim_{n\longrightarrow\infty} \frac{n^{k-2}}{(k-1)!}\mathbf{E}(W_{n,k}) = \frac{k}{2}\left(1 + \varepsilon_k e^{-k}\right). \tag{8}$$

2 Matrix Rank

We will study the process whereby we add random columns one by one to create a sequence of matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots,$ so that $\mathbf{A}_m$ is distributed as $\mathbf{A}_{n,m;k}$. The corresponding hypergraphs generate a sequence $H_1, H_2, \ldots, H_m$, where H_m is distributed as $H_{n,m;k}$. Let the columns of $\mathbf{A}_m$ be $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m$, where they have been added in this order. The edges of H_m corresponding to these columns are $e_1, e_2, \ldots, e_m$. We let $c = km/n$.

Case 1: $c < c_k^*$**.**
When $c \leq c_k^*$ we re-order the edges so that $e_1, e_2, \ldots, e_{m_1}$ are not wholly contained in C_2 and $e_{m_1+1}, e_{m_1+2}, \ldots, e_m \subseteq C_2$. Of course, if $c < \widehat{c}_k$ then $m_1 = m$. Let $v_1, v_2, \ldots, v_{i_1}$ be the set of isolated vertices in H_m. We order the remaining vertices not in C_2 as $v_{i_1+1}, v_{i_1+2}, \ldots, v_{i_1+m_1}$ so that for each $1 \leq j \leq n_1$ there is exactly one edge that contains v_{i_1+j} and is contained in $\{v_{i_1+j}, v_{i_1+j+1}, \ldots, v_n\}$. To see this, consider the standard construction of C_2 by first removing isolated vertices and then repeatedly removing vertices of degree one in the remaining hypergraph.

It follows that we can re-order the edges of H_{m_1} so that when we add column $\mathbf{c}_i, 1 \leq i \leq m_1$ there will be at least one row v such that $\mathbf{c}_i(v)$ is the only column with a one in row v. Thus, adding this column will increase the rank by one and so the rank of $\mathbf{A}_{m_1}$ is m_1. The remaining columns of $\mathbf{A}_m$ correspond to edges contained entirely within the 2-core.

The columns associated with the 2-core are distributed as uniformly random, subject to each vertex/row of the 2-core being in at least two columns. It follows from Pittel and Sorkin [13] that the rank of the columns $\mathbf{c}_{m_1+1}, \mathbf{c}_{m_1+2}, \ldots, \mathbf{c}_m$ is $\approx m_2 = m - m_1$. Indeed, [13] shows that if we choose a random $m_2 \times \nu$, 0/1 Boolean matrix $\mathbf{B}$ such that (i) $m_2 \leq n$, (ii) each row has exactly k 1's and (iii) each column has at least two 1's and (iv) $\mathbf{B}$ is sampled uniformly subject to (i), (ii), (ii) then w.h.p. $\mathbf{B}$ has rank $\approx m_2$. In our case, the m_2 columns of $\mathbf{A}_m$ associated with C_2 are distributed as the transpose of the problem studied in [13].

From the above we see that the rows and columns of the matrix $\mathbf{A}_m$ can be arranged as in (9) below, and where A_1 is lower triangular with 1's on the diagonal. It follows from the above discussion that $\text{rank}(\mathbf{A}_m) = m_1 + \text{rank}(\mathbf{C}_2)$. For $c < c_k^*$, $\mathbf{C}_2$ has asymptotic full column rank m_2. For this case the first claim of (5), and Theorem 1.1, have been verified.

$$\mathbf{A}_m = \begin{array}{c c} & \begin{array}{cc} m_1 & m_2 \end{array} \\ \begin{array}{c} m_1 \\ n_1 - m_1 \\ |C_2| \\ n_0 \end{array} & \begin{pmatrix} A_1 & 0 \\ A_2 & 0 \\ \star & \mathbf{C}_2 \\ 0 & 0 \end{pmatrix} \end{array} \tag{9}$$

Case 2: $c \geq c_k^*$.
To prove Theorem 1.1 for $c \geq c_k^*$ we only need to verify that w.h.p.

$$\text{rank}(\mathbf{C}_2) \approx |C_2|. \tag{10}$$

In this case we need some basic facts about hypergraphs. We say a hypergraph H is *linear* if edges only intersect in at most one vertex. We define a k-uniform *cactus* as follows. A single edge is a cactus. An $(\ell+1)$-edge cactus C' is the structure obtained from an ℓ-edge cactus C with vertex set $V(C), |V(C)| = (k-1)\ell + 1$ as follows. Choose $x \in V(C)$ and let $V(C') = V(C) \cup \{v_1, ...v_{k-1}\}$ where $\{v_1, ...v_{k-1}\}$ is disjoint from $V(C)$. The edge set $E(C')$ of C' is $E(C) \cup \{e'\}$ where $e' = \{x, v_1, ...v_{k-1}\}$. We need the following simple lemma.

Lemma 2.1. *A connected k-uniform simple hypergraph C with no cycles is a cactus.*

Proof. This can easily be verified by induction. We simply remove one terminal edge $e = \{v_1, v_2, \ldots, v_k\}$ of a longest path P. We can assume here that $v_2, \ldots, v_k$ are all of degree one, else P can be extended. Deleting e gives a new connected hypergraph C' which is a cactus by induction. □

For a k-uniform linear hypergraph H let $L(H) = (k-1)|E(H)| + 1$.

Lemma 2.2. *Let H be a connected k-uniform linear hypergraph.*

(a) $|V(H)| \leq L(H)$.

(b) $|V(H)| = L(H)$ *if and only if H does not contain any cycles.*

(c) By deleting at most $L(H) - |V(H)|$ *edges we can create a subgraph* H' *with* $V(H') = V(H)$ *and no cycles.*

Proof. (a) Because H is connected, we can order the edges $e_1, e_2, \ldots, e_m$ of H so that $e_i \cap \bigcup_{j=1}^{i-1} e_j \neq \emptyset$. The addition of edge e_i adds at most $|e_i| - 1$ new vertices and $|V(H)| \leq L(H)$ follows by induction.

(b) Suppose first that H contains a cycle C with edges $e_1, e_2, \ldots, e_r$, Now consider H' which is obtained from H by deleting e_1 and the $t \leq k-2$ vertices that become isolated. Then $|V(H')| = |V(H)| - t$ and $L(H') = L(H) - (k-1) = |V(H)| - (k-1) = |V(H')| + t - (k-1) < |V(H')|$ and this contradicts the condition that $|V(H')| \leq L(H')$.

Suppose next that H does not contain any cycles. Then there must be a vertex of degree one. We simply delete this vertex and use induction on the number of vertices. Putting back the vertex adds one to the number of vertices and one to L.

(c) If $V(H) < L(H)$ then H contains a cycle C. Therfore we can remove an edge of C without changing the vertex set and this will strictly decrease L. □

In the following lemma we prove a property of $H_{n,m;k}$. It will be more convenient to work with $H_{n,p;k}$ where $m = \binom{n}{k}p$. We use the fact that for any hypergraph property $\mathcal{H}$ that is monotone increasing or decreasing with respect to adding edges,

$$\mathbf{Pr}(H_{n,m;k} \in \mathcal{H}) \leq O(1)\, \mathbf{Pr}(H_{n,p;k} \in \mathcal{H}). \tag{11}$$

This is well-known for graphs and is essentially a property of the binomial random variable, $E(H_{n,p;k})$, the number of edges of $H_{n,p;k}$.

Similarly, if $\mathcal{A}$ is a matrix property that is monotone increasing or decreasing with respect to adding columns, then

$$\mathbf{Pr}(\mathbf{A}_{n,m;k} \in \mathcal{A}) \leq O(1)\, \mathbf{Pr}(\mathbf{A}_{n,p;k} \in \mathcal{A}). \tag{12}$$

For a set of vertices S, let $\ell_i(S)$ denote the number of edges e such that $|e_i \cap S| = i$.

Lemma 2.3. *Suppose that $m = O(n \log n)$. Then w.h.p.,*

(a) For every set of vertices S of size $s \leq s_0 = n^{1-\alpha}$ *we have* $\sum_{i=2}^{k}(i-1)\ell_i(S) \leq s + \lfloor \theta s \rfloor$, *where* $\theta = \frac{2 \log\log n}{\alpha \log n}$.

(b) There are at most $n^{o(1)}$ vertices in cycles of size at most $\log^{1/2} n$.

Proof. (a) We can use (11) here with $p = \frac{C \log n}{n^{k-1}}$ for some constant $C > 0$. The expected number of sets failing this property can be bounded by

$$\sum_{s=2}^{s_0} \binom{n}{s} \sum_{L > s + \lfloor \theta s \rfloor} \sum_{\sum_{i=2}^{k}(i-1)\ell_i = L} \prod_{i=2}^{k} \left(\binom{\binom{s}{i}}{\ell_i} \left(\binom{n-s}{k-i} \frac{C \log n}{n^{k-1}} \right)^{\ell_i} \right) \tag{13}$$

$$\leq \sum_{s=2}^{s_0} \left(\frac{ne}{s}\right)^s \sum_{L > s + \lfloor \theta s \rfloor} \sum_{\sum_{i=2}^{k}(i-1)\ell_i = L} \prod_{i=2}^{k} \left(\frac{Ce^2 s^i \log n}{i!(k-i)!\ell_i n^{i-1}} \right)^{\ell_i}$$

$$\leq \sum_{s=2}^{s_0} (Ce^3 \log n)^L \sum_{L > s + \lfloor \theta s \rfloor} \sum_{\sum_{i=2}^{k}(i-1)\ell_i = L} \left(\frac{s}{n}\right)^{\sum_{i=2}^{k}(i-1)\ell_i - s} \prod_{i=2}^{k} \left(\frac{s}{\ell_i}\right)^{\ell_i}$$

$$\leq \sum_{s=2}^{s_0} \sum_{L > s + \lfloor \theta s \rfloor} (Ce \log n)^L \left(\frac{s}{n}\right)^{L-s} \sum_{\sum_{i=2}^{k}(i-1)\ell_i = L} e^{ks},$$

$$\leq \sum_{s=2}^{s_0} \sum_{L > s + \lfloor \theta s \rfloor} \left(2Ce^k \log n \left(\frac{s}{n}\right)^{1-s/L} \right)^L,$$

$$= o(n^{-1}). \tag{14}$$

The fourth line follows because $(s/\ell)^\ell \leq s^\ell/\ell! \leq e^s$, and the last line because

$$\left(\frac{s_0}{n}\right)^{1-s/L} = O\left(\frac{1}{\log^{2-o(1)} n}\right).$$

(b) The expected number of vertices in small cycles can be bounded by

$$\sum_{\ell=2}^{\log^{1/2} n} \binom{n}{(k-1)\ell} ((k-1)\ell)! p^\ell \leq \sum_{\ell=2}^{\log^{1/2} n} (n^{k-1}p)^\ell \leq \sum_{\ell=2}^{\log^{1/2} n} (C \log n)^\ell = n^{o(1)}.$$

Part (b) now follows from the Markov inequality. □

2.1 Growth of the mantle

We now consider the change in the rank of the sub-matrix $\mathbf{C}_2$ of the edge-vertex incidence matrix $\mathbf{A}_m$ (see (9)), caused by adding an edge e. Suppose that the addition of e increases the size of the 2-core. Our calculations of rank are not exact, and the difference caused by a few edges of co-degree two, or small isolated components of the 2-core can be subsumed into our error terms.

Let A denote the set of additional vertices and F denote the set of additional edges added to C_2 by the addition of e, where $A \subset V(F)$. We include e in F. Let $H = (A, F)$.

We remark first that with c, x as in (1), that q.s.

$$|C_2| - n(1 - e^{-cx}(1 + cx)) + | \leq n^{3/4}, \qquad \text{and} \qquad |E(C_2)| - mx^{k/(k-1)}| \leq n^{3/4}. \tag{15}$$

Therefore we can assume that adding an edge to $\mathbf{A}_m$ can only increase $C_2,\ E(C_2)$ by at most $n^{3/4}$. We use Lemma 2.3 with $\alpha = 3/4$ in our discussion of the hypergraph H.

The increase in the rank of $\mathbf{C}_2$ obtained by adding $|F|$ extra edges is at most $|F|$, as each additional edge increases the column rank by at most one. We relate $|F|$ and $|A|$ as follows.

Case 1: First consider the case where there are no cycles in H. We prove a bijection between A and $F - e$. There must exist $v \in A$ and $v \in f \in F$, $f \neq e$, such that $f \setminus \{v\} \subseteq C_2$. For this consider a longest path from e to C_2 through H, and let f be the last edge on this path. Because every vertex of A has degree at least two in F, vertex v must be in some other edge $g \in F$. But we cannot have $g \setminus \{v\} \subseteq C_2$, else $v \in C_2$ (a contradiction) or $A = \{v\}$, $g = e$, $F = \{e, f\}$ and the induction terminates. Map v to f in the bijection. Remove f from F, v from A, replace v everywhere in edges of F by a distinct vertex of C_2 and apply induction.

The increase in rank in this case is at least $|A| = |F| - 1$, and at most $|F| = |A| + 1$.

Case 2: The total contribution to the rank of the 2-core in $m = O(n \log n)$ steps from the case where H_0 contains a cycle of length at most $\log^{1/2} n$ can be bounded by $n^{3/4+o(1)}$. This follows from Lemma 2.3(b) and (15). This is negligible, since the core has size $\Omega(n)$ in the regime we are discussing.

Case 3: Suppose that H contains cycles which we remove by deleting s edges. When we do this we may lose up to ks vertices from A. Let the resulting vertex set be A' and edge set be F'. Up to ks vertices of A' may have degree 1. Attach these vertices to C_2 using disjoint edges to give edge set F''. All vertices of A' now have degree at least 2 in F'' and F'' has no cycles. According to the argument in Case 1, the increase in rank due to adding F'' is $|A'| \geq |A| - ks$ and this is at most ks larger than the increase in rank due to adding F'. Thus the increase in rank due to adding $F \supseteq F'$ is at least $|A| - 2ks$ and at most $|F| \leq |A| + s + 1$. It follows from Lemma 2.2(c) that $s = o(|A|)$.

In summary we find that if $m = O(n \log n)$ and $m \geq c^* n/k$ then, with probability $1 - o(n^{-1})$, the rank of $\mathbf{C}_2$ satisfies

$$(1 - o(1))\, |C_2| \leq \text{rank}(\mathbf{C}_2) \leq |C_2|. \tag{16}$$

The upper bound follows because the rank of $\mathbf{C}_2$ is at most the number of rows in $\mathbf{C}_2$. This proves (10). To finish the proof of Theorem 1.1 we require that (16) remains true if we take expectations. For this we use the error probability of $o(n^{-1})$ in (14).

2.2 Proof of Theorem 1.3

Proof of part (i):
Given a set of rows S, the number of choices of column (distinct edges) that have an odd

number of non-zero entries in S is

$$T_{s,k} = \binom{s}{1}\binom{n-s}{k-1} + \binom{s}{3}\binom{n-s}{k-3} + \cdots + \binom{s}{k}.$$

If $\mathrm{rank}(\mathbf{A}_{n,p;k}) < n^*$ then there exists a set S of rows such that (i) each column of $\mathbf{A}_{n,p;k}$ has an even number of non-zero entries j in S and (ii) $|S| \leq n^*$. For a fixed S, denote this event by $\mathcal{B}_S$ and note that it is monotone decreasing. Then

$$\mathbf{Pr}(\mathcal{B}_S) = (1-p)^{T_{s,k}}. \tag{17}$$

For $s \geq k$,

$$T_{s,k} \geq \binom{s}{1}\binom{n-s}{k-1} + \binom{s}{k} = \frac{s}{(k-1)!}\left(\frac{s^{k-1}}{k} + (n-s)^{k-1}\right)(1+o(1))$$

The bracketed term on the right hand side is minimized when $s = \alpha n$ where $\alpha = k^{1/(k-2)}/(1+k^{1/(k-2)})$. Let $\beta_k = (\alpha^{k-1}/k + (1-\alpha)^{k-1})$ then

$$T_{s,k} \geq \beta_k s \frac{n^{k-1}}{(k-1)!}(1+o(1)).$$

We can choose $p = \frac{(A+2)\log n}{\beta_k \binom{n-1}{k-1}}$ and then use monotonicity of rank as a function of p to claim the result for larger p.

$$\begin{aligned}
\mathbf{Pr}(\exists S : \mathcal{B}_S \text{ occurs}) &\leq \sum_{s=1}^{n^*} \binom{n}{s}(1-p)^{T_{s,k}} \\
&\leq \sum_{s=1}^{n^*} \left(\frac{ne}{s} \cdot \exp\left\{-p\beta_k \frac{n^{k-1}}{(k-1)!}(1+o(1))\right\}\right)^s \qquad (18)\\
&\leq \sum_{s=1}^{n^*} n^{-(A+1)s} = O\left(\frac{1}{n^{A+1}}\right).
\end{aligned}$$

We now use (12) to transfer this bound to $\mathbf{A}_{n,m;k}$.

Proof of part (ii).
Let $m = n(\log n + c_n)/k$. Assume that $c_n \to c$. We first observe that if Z_s denotes the number of sets of $s = O(1)$ empty rows then

$$\begin{aligned}
\mathbf{E}(Z_s) &= \binom{n}{s}\frac{\binom{\binom{n-s}{k}}{m}}{\binom{\binom{n}{k}}{m}} = \binom{n}{s}\prod_{i=0}^{m-1}\frac{\binom{n-s}{k}-i}{\binom{n}{k}-i} = \binom{n}{s}\left(\frac{\binom{n-s}{k}}{\binom{n}{k}}\right)^m\left(1+O\left(\frac{m^2}{n^k}\right)\right) \\
&\approx \frac{n^s}{s!}\cdot\prod_{i=0}^{k-1}\left(1-\frac{s}{n-i}\right)^m = \frac{n^s}{s!}\cdot\prod_{i=0}^{k-1}\exp\left\{-\frac{ms}{n}+O\left(\frac{m}{n^2}\right)\right\} \approx \frac{n^s}{s!}e^{-skm/n} \approx \frac{e^{-cs}}{s!}. \qquad (19)
\end{aligned}$$

The method of moments implies that Z_1 is asymptotically Poisson with mean e^{-c} and so

$$\mathbf{Pr}(Z_1 = 0) \approx e^{-e^{-c}}. \tag{20}$$

Going back to (18) with $p = (\log n + c_n)/\binom{n-1}{k-1}$ we see that we only need to consider $2 \le s \le 4n^{1-\beta_k}$. For these values of s, $T_{s,k}$ is bounded below by $s\binom{n-s}{k-1} \approx s\binom{n-1}{k-1}$. Thus we can bound the RHS of (18) from above by

$$\sum_{s=1}^{4n^{1-\beta_k}} \left(\frac{3n}{s} \cdot \exp\left\{ -p\binom{n-1}{k-1} \right\} \right)^s = \sum_{s=1}^{4n^{1-\beta_k}} \left(\frac{O(1)}{s} \right)^s.$$

Thus,

$$\mathbf{Pr}(\exists S, \log\log n \le |S| \le 4n^{1-\beta_k} : \mathcal{B}_S \text{ occurs}) \le \sum_{s=\log\log n}^{n^{1-\beta_k}} \left(\frac{O(1)}{s} \right)^s = o(1). \tag{21}$$

Finally we consider $2 \le s \le L = \log\log n$. The final step is to prove (w.h.p) that when $p = (\log n + c_n)/\binom{n-1}{k-1}$, $c_n \longrightarrow c$ constant the only obstruction to $\text{rank}(\mathbf{A}_{n,p;k}) = n^*$ is the existence of empty rows ($Z_1 > 0$).

Given a set S, the number of choices of column that have an odd number of non-zero entries in S (Type A columns) is given by $T_{s,k}$ above, and the number of choices of columns that have an even number of non-zero entries in S (Type B columns) is

$$R_{s,k} = \binom{s}{2}\binom{n-s}{k-2} + \cdots + \binom{s}{k-1}(n-s).$$

For $s \le L$, $R_{s,k} \le s^2 n^{k-2}$. The expected number μ_s of sets S with no Type A columns and at least one Type B column is

$$\mu_s = \binom{n}{s} \left(1 - (1-p)^{R_{s,k}}\right) (1-p)^{T_{s,k}} \le \frac{n^s}{s!} (pR_{s,k})\, e^{-ps\binom{n-1}{k-1}(1+o(1))} = O\left(\frac{\log n}{n} \right) e^{-cs}.$$

Thus, for constant c,

$$\sum_{s=2}^{L} \mu_s = o(1). \tag{22}$$

Thus w.h.p. there is no set of $2 \le s \le \log\log n$ rows where the dependency does not come from the rows all being zero. $\square$

2.3 Proof of Theorem 1.4

Because c in (20) is arbitrary and having a zero row is a monotone decreasing event, we can see that if $m_0 = n(\log n - \log\log n)/k$ then $Z_1 = Z_1(m_0) > 0$ w.h.p. The reader can easily check that equations (21) and (22) continue to hold. It follows that w.h.p. the rank of $\mathbf{M}_{m_0}$ is $n^* - Z_1$. It then follows that $m_1 = m^*$ if we never add a column that reduces the number of non-zero rows by more than one. Now (20) implies that the expected number of zero rows in $\mathbf{M}_{m_0}$ is $O(\log n)$ and so $Z_1 \leq \log^2 n$ w.h.p. So given this, the probability we add add a column that reduces the number of non-zero rows by more than one in the next $O(n \log n)$ column additions, is $O(n \log n \times ((\log^2 n)/n)^2 = o(1)$.

3 Minimum Weight Basis

The first task here is to prove (6). Let $B_{n,k}$ denote a minimum weight basis and let $W_{n,k}$ denote its weight. For a given a real number X we can write

$$X = \int_{p=0}^{X} dp = \int_{p=0}^{1} 1_{p \leq X} dp.$$

Thus

$$\begin{aligned}
W_{n,k} &= \sum_{\mathbf{c} \in B_{n,k}} X_{\mathbf{c}} \\
&= \sum_{\mathbf{c} \in B_{n,k}} \int_{p=0}^{1} 1_{p \leq X_{\mathbf{c}}} dp && (23) \\
&= \int_{p=0}^{1} \sum_{\mathbf{c} \in B_{n,k}} 1_{p \leq X_{\mathbf{c}}} dp \\
&= \int_{p=0}^{1} |\{\mathbf{c} \in B_{n,k} : p \leq X_{\mathbf{c}}\}| dp \\
&= \int_{p=0}^{1} (n^* - \operatorname{rank}(\mathbf{A}_p)) dp. && (24)
\end{aligned}$$

Here $\mathbf{A}_p$ is any matrix made up of those columns $\mathbf{c} \in \Omega_{n,k}$ with $X_{\mathbf{c}} \leq p$. And let A_p denote the corresponding hypergraph.

Explanation for (24)**:** Finding a minimum cost basis B can be achieved via a *greedy algorithm*. We first order the columns of $\Omega_{n,k}$ as $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N, N = \binom{n}{k}$ in increasing order of weight $X_{\mathbf{c}}$. Treating B as a set of columns, we initialise $B = \emptyset$, and for $i = 1, 2, \ldots, N$ add $\mathbf{c}_i$ to B if it is linearly independent of the columns of B selected so far. This means that for any $0 \leq p \leq 1$, the number of columns in B with $X_{\mathbf{c}} > p$ must be equal to the co-rank of the set of columns selected before them i.e $B_p = \{\mathbf{c} \in B : X_{\mathbf{c}} \leq p\}$. We claim that B_p is a maximal linear independent subset of the columns of $\mathbf{A}_p$. If it were not maximal, then another column of $\mathbf{A}_p$ would have been added to B_p by the greedy algorithm.

We obtain $\mathbf{E}W_{n,k}$ in (6) by taking the expectation of (24), using Fubini's theorem to take the expectation inside the integral.

We first argue that

$$\mathbf{E}(W_{n,k}) = \Omega(n^{-(k-2)}). \tag{25}$$

Let $\mathbf{c} = (c_1, ..., c_n)$, where $c_i \in \{0,1\}$ denotes the i-th coordinate of $\mathbf{c}$. We can bound $W_{n,k}$ from below by $\sum_{i=1}^{n} \min\{X_{\mathbf{c}} : c_i = 1\}$. Let $N = \binom{n}{k}$. The number of ones in a fixed row of $\mathbf{A}_{n,N;k}$ is $L = Nk/n$. The expected minimum of L independent uniform $[0,1]$ random variables is $1/(L+1)$. Hence

$$\mathbf{E}(W_{n,k}) \geq \frac{n^2}{k\binom{n}{k} + n}$$

and (25) follows.

It follows from Theorem 1.3(i) with $A = k$, $p = km/(n\binom{n-1}{k-1})$. and (6) that

$$\begin{aligned}
\mathbf{E}(W_{n,k}) &\approx \int_{p=0}^{k!\gamma n^{1-k}\log n} (n^* - \mathrm{rank}(\mathbf{A}_p))dp \\
&= \frac{(k-1)!}{n^{k-1}} \int_{c=0}^{k\gamma \log n} (n^* - \mathrm{rank}(\mathbf{A}_{c(k-1)!/n^{k-1}}))dc \\
&= (I_1 + I_2)\frac{(k-1)!}{n^{k-1}}, \qquad (26)
\end{aligned}$$

where $I_1 = \int_{c=0}^{c_k^*} \cdots dc$ and $I_2 = \int_{c_k^*}^{k\gamma\log n} \cdots dc$.

Since $H_{c/n^{k-1}}$ q.s. has $m \approx cn/k$ edges, it follows from Theorem 1.1 that

$$I_1 \approx \int_{c=0}^{c_k^*} \left(n^* - \frac{cn}{k}\right) dc \approx c_k^* n \left(1 - \frac{c_k^*}{2k}\right). \qquad (27)$$

We claim, and this is proved below, that we can write I_2 in the form given in (28) and (29). The integrand in (28) is the expression for rank from Corollary 1.2, with $x^{1/(k-1)} = (1-e^{-cx})$ substituted from (1).

$$\begin{aligned}
I_2 &\approx n \int_{c_k^*}^{k\gamma \log n} \left(1 - \left(\frac{c}{k} - \frac{cx^{k/(k-1)}}{k} + (1 - e^{-cx}(1+cx))\right)\right) dc \qquad (28) \\
&\approx n \int_{c_k^*}^{\infty} \left(e^{-cx}(1 + cx(k-1)/k) - \frac{c}{k}(1-x)\right) dc \qquad (29)
\end{aligned}$$

Expression (7) of Theorem 1.5 now follows from (27) and (29).

Verification of (28)**:** We will express $I_2 = n(I_a + I_b)$ where $I_a = \int_{c=c_k^*}^{K\log\log n} \cdots dc$ and $I_b = \int_{K\log\log n}^{k\gamma\log n} \cdots dc$ for some large constant K. For I_a we use (15) and (16) to argue that with probability $1 - o(n^{-1})$, $\mathrm{rank}(\mathbf{C}_2)$ is within $\left(1 - O\left(\frac{\log\log n}{\log n}\right)\right)$ of n times the expression used for it in the integrand. Because the range of integration is $O(\log\log n)$, the errors that mount up are of order $o(1)$.

For I_b we use the monotonicity of rank to replace the integrand at $c \in [K\log\log n, k\gamma\log n]$ by the integrand at $c_1 = K\log\log n$. We observe that as $c \to \infty$ we can easily check that

$$1 - 2ke^{-c} \leq x \leq 1. \qquad (30)$$

Indeed, putting $x = 1 - y$ we have $(1-y)^{1/(k-1)} = 1 - e^{-c(1-y)}$. We see that if $f(y) = (1-y)^{1/(k-1)} - (1 - e^{-c(1-y)})$ then $f(0) > 0$ and $f(2ke^{-c}) < 0$ for large c.

Thus

$$\frac{c_1}{k} - \frac{c_1 x^{k/(k-1)}}{k} + (1 - e^{-c_1 x}(1 + c_1 x)) \geq 1 - e^{-c_1 x}(1 + c_1 x) \geq 1 - e^{-99c_1/100} \geq 1 - \frac{1}{\log^{K/2} n}.$$

Consequently, $I_b = O(1/\log^{K/2-1} n)$.

Verification of (29)**:** For this we use (30) to see that

$$I_b \leq \int_{c=c_1}^{\infty} \left(\frac{c}{k} \left(1 - (1 - e^{-c/2})^{k/(k-1)}\right) + e^{-c/2} \right) dc \leq \int_{c=c_1}^{\infty} \left(\frac{ce^{-c/2}}{k-1} + e^{-c/2} \right) dc = o(1).$$

3.1 Bounds for finite k

We begin by estimating c_k^*. Let x be as in (1), then going back to the definition (4), we can determine the value of $c_k^* = c(x)$ from

$$c\left(\frac{k-1}{k}\right) x^{\frac{k}{k-1}} - cx + x^{\frac{1}{k-1}} = 0. \tag{31}$$

Solve for c, and put $y = x^{1/(k-1)}$ to give

$$c \;=\; \frac{1}{y^{k-2} - ((k-1)/k)y^{k-1}}. \tag{32}$$

Substituting for c via (1) gives

$$y = 1 - \exp\left\{ -\frac{ky}{k-(k-1)y} \right\}. \tag{33}$$

If $x \in (0,1)$ then $y \in (0,1)$, and $y \geq x$. We look for solutions of the form $y = 1 - z$. Making this substitution (33) becomes $z = q(z)$ where

$$q(z) = \exp\left\{ -\frac{k(1-z)}{1+(k-1)z} \right\}.$$

Let

$$z = z(\delta) = \frac{\delta}{k-(k-1)\delta}, \tag{34}$$

then (stretching notation somewhat) $q(\delta) = e^{-k(1-\delta)}$. Consider $f(\delta) = z(\delta) - q(\delta)$, then

$$f(\delta) \geq \frac{\delta}{k}\left(1 + \frac{k-1}{k}\delta\right) - e^{-k}e^{k\delta}.$$

Substitute $\delta = \theta k e^{-k}$ to give

$$f(\theta) \geq e^{-k}\left(\theta(1+\theta(k-1)e^{-k}) - e^{\theta k^2 e^{-k}}\right).$$

The function $k^2 e^{-k}$ in the exponent of the last term is monotone deceasing for $k \geq 2$. Let $\theta = 3/2$, then for $k \geq 4$, it can be checked that $f(\theta, k) > 0$. Now $f(0) < 0$ and so there is a solution to $f(\delta) = 0$ in the interval $(0, \theta k e^{-k})$.

Substitute $y = 1 - z$ into (32) to obtain

$$\frac{c}{k} = \frac{1}{(1-z)^{k-2}(1+(k-1)z)} \tag{35}$$

Lemma 3.1. *(i) Let $\theta = 3/2$, then for $k \geq 4$,*

$$k(1-\theta e^{-k}) \leq c_k^* \leq k. \tag{36}$$

(ii) For $k = 3$, $c_3^* = 2.753699....$

(iii) If $k \geq 4$ and $c \geq c_k^*$ then the solution x to (1) satisfies $x \geq 1 - 3ke^{-c}/2$.

Proof. (i) For the upper bound we note that for $k \geq 3$ the denominator of c in (35) is monotone increasing for $z \leq 1/(k-1)^2$ from a value of one when $z = 0$. For the lower bound, as $1/(1-z)^{k-2} > 1 + (k-2)z$, it follows from (35), the definition of z in (34), and $\delta < \theta k e^{-k}$ that

$$\frac{c}{k} > \frac{1+(k-2)z}{1+(k-1)z} = 1 - \frac{\delta}{k} > 1 - \theta e^{-k}.$$

(ii) Set $y = \sqrt{x}$ and invert (1) to obtain

$$c = \frac{1}{y^2}\log\frac{1}{1-y}.$$

Inserting this into (32) gives

$$y + \left(\frac{2}{3}y - 1\right)\log\frac{1}{1-y} = 0.$$

This was solved numerically to give the following results for y, x, c_3^*

$$y = 0.8833916, \quad x = 0.9398891, \quad c_3^* = 2.753699. \tag{37}$$

(iii)

Let $x = 1 - \varepsilon$. We first verify that $\varepsilon \leq 1/c$. Putting $f(\varepsilon) = 1 - \varepsilon - (1 - e^{-c+c\varepsilon})^{k-1}$ we see that $f(0) > 0$ and $f(1/c) < 0$ for $c \geq c_k^*$ as given in (i). If $ay < 1$, then $1 - (1-y)^a < ay$. As $(k-1)e^{-c+c\varepsilon} < 1$ for any $\varepsilon < 1 - (\log(k-1))/c$,

$$f(c^{-1}) = 1 - c^{-1} - (1 - e^{-c+1})^{k-1} \leq 1 - c^{-1} - 1 + (k-1)e^{-c+1}.$$

Now $c(k-1)e^{-c+1}$ is decreasing as a function of c. And for $k \geq 4$, $k(k-1)e^{-c+1}$ and $e^{(3k/2)e^{-k}}$ are decreasing as functions in of k. Therefore, for c satisfying (36),

$$c(k-1)e^{-c+1} < k(k-1)e^{-(k-1)}e^{(3k/2)e^{-k}} < 1.$$

Let $x = 1 - \varepsilon$, and $\delta = e^{-c+c\varepsilon}$. Rewrite (1) as

$$-\log(1-\varepsilon) = \varepsilon + \frac{\varepsilon^2}{2} + \cdots = (k-1)\left(\delta + \frac{\delta^2}{2} + \cdots\right). \tag{38}$$

It must hold that $\varepsilon \leq (k-1)\delta$ otherwise the left hand side is greater than the right hand side. Thus, as $\varepsilon < 1/c$,

$$\varepsilon \leq (k-1)e^{-c+c\varepsilon} \leq (k-1)e^{-c+1}.$$

A repeated application of this bound, (36) and direct calculation gives

$$\varepsilon \leq (k-1)\exp\left\{-c + (k-1)ce^{-c+1}\right\} \leq (k-1)\exp\left\{-c + (k-1)ke^{1-(1-\theta e^{-k})k}\right\} \leq 3ke^{-c}/2.$$

□

Going back to (27) and using Lemma 3.1(i), we see that for $k \geq 4$,

$$\frac{kn}{2}\left(1 - \frac{9}{4}e^{-2k}\right) \leq I_1 \leq \frac{kn}{2}. \tag{39}$$

We evaluate I_2 from (28)–(29) in two parts. Firstly, using Lemma 3.1(iii) for $c \geq c_k^*$,

$$-\frac{3}{2}ce^{-c} \leq -\frac{c}{k}(1-x) \leq 0. \tag{40}$$

Note also that $1 - 3ke^{-c}/2 \geq 1 - 1/2k$ for $k \geq 4$ and $c \geq c_k^*$. Thus

$$e^{-c}\left(1 + c\frac{(k-1)(2k-1)}{2k^2}\right) \leq e^{-cx}\left(1 + cx\frac{k-1}{k}\right) \leq e^{1/2}e^{-c}\left(1 + \frac{c(k-1)}{k}\right).$$

For the LHS we replace e^{-cx} by e^{-c} (since $x \leq 1$) and x by $1 - 1/2k$. For the RHS we replace $cx(k-1)$ by $c(k-1)$, and $e^{-cx} = e^{-c+c\varepsilon}$. Using Lemma 3.1(i) and (iii), as $c^* > 1$, it follows that

$$e^{c\varepsilon} \leq e^{(3k/2)ce^{-c}} \leq e^{(3k/2)c^* e^{-c^*}} \leq e^{1/2}. \tag{41}$$

Adding the contributions from (40) and (41) we find that

$$n\int_{c^*}^{\infty} e^{-c}\left(1 - c\frac{k^2+3k-1}{2k^2}\right)\, dc \leq\ I_2\ \leq ne^{1/2}\int_{c^*}^{\infty} e^{-c}\left(1 + \frac{c(k-1)}{k}\right)\, dc.$$

Thus, with the *indefinite integral* $\int e^{-c}(1+Ac) = -e^{-c}(1+A+Ac)$, we get

$$ne^{-c_k^*}\left(\frac{k^2-3k+1}{2k^2} - c_k^*\frac{k^2+3k-1}{2k^2}\right) \leq\ I_2\ \leq ne^{1/2}e^{-c_k^*}\left(\frac{2k-1}{k} + c_k^*\frac{k-1}{k}\right),$$

or more simply

$$-n\frac{k}{2}e^{-c_k^*}\left(1+\frac{3}{k}\right) \leq \; I_2 \; \leq n\frac{k}{2}e^{-c_k^*}\,2e^{1/2}\left(1+\frac{1}{k}\right).$$

Noting that $e^{-c_k^*} \leq 6e^{-k}/5$ for $k \geq 4$, we have

$$n\frac{k}{2}\left(1-\frac{9}{4}e^{-2k}-\frac{21}{10}e^{-k}\right) \leq I_1+I_2 \leq n\frac{k}{2}\left(1+3e^{1/2}e^{-k}\right).$$

Thus, for some ε_k, $|\varepsilon_k| \leq 5$,

$$I_1+I_2 = n\frac{k}{2}\left(1+\varepsilon_k e^{-k}\right).$$

4 Open questions

Q1 The formula for the cost of a minimum weight basis when $k \geq 3$ given by Theorem 1.5 is asymptotically accurate, but lacks the elegance of the case where $k = 2$. Can the expression be simplified for say, $k = 3$?

Q2 The $\zeta(3)$ result of [7] was generalised quite substantially to consider minimum weight spanning trees of d-regular graphs, when d is large, see [3]. In the context of $\mathbf{A}_{n,m;k}$, this suggests that we consider the case where each row has exactly d ones. Here we can study the rank as well as $W_{n,k}$.

References

[1] J. Altschuler and E. Yang, Inclusion of Forbidden Minors in Random Representable Matroids, arXiv:1507.05332 [math.CO].

[2] N. Bansal, R.A. Pendavingh and J.G. van der Pol, On the number of matroids, *Combinatorica* 35 (215) 253-277.

[3] A. Beveridge, A. M. Frieze and C. J. H. McDiarmid, Minimum length spanning trees in regular graphs, *Combinatorica* 18 (1998) 311-333.

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

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

[6] C. Cooper, A.M. Frieze and W. Pegden, Minors of a random binary matroid.

[7] A.M. Frieze, On the value of a random minimum spanning tree problem, *Discrete Applied Mathematics* 10 (1985) 47-56.

[8] S. Janson, The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph, *Random Structures and Algorithms* 7 (1995) 337-355.

[9] W. Kordecki and A. Lyczkowska-Hanćkowiak, Exact Expectation and Variance of Minimal Basis of Random Matroids, *Discussiones Mathematicae Graph Theory* 33 (2013) 277-288.

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

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

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

[13] B. Pittel and G. Sorkin, The Satisfiability Threshold for k-XORSAT, *Combinatorics, Probability and Computing* 25 (2016) 238-268.